AUDIOVISUAL SPEECH PERCEPTION IN 4-MONTH-OLD INFANTS

by

RENÉE NICOLE DESJARDINS

B.A. (Honours), Queen's University, 1985
M.A., The University of British Columbia, 1993

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

August 1997

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 August 15, 1997

DE-6 (2/88)

ABSTRACT

Previous research indicates that for adults and children the perception of speech can be significantly influenced by watching a speaker's mouth movements. For example, hearing the syllable /bi/ while watching a speaker mouth the syllable /vi/ results in reports of a 'heard' /vi/. Some evidence suggests young infants also may be able to integrate heard and seen speech. One theory suggests that an innate link between perception and production (Liberman & Mattingly, 1985) accounts for this phenomenon while another theory suggests that experience (e.g., producing speech sounds) may be necessary into order to develop fully the underlying representation of visible speech (Desjardins, Rogers & Werker, in press; Meltzoff & Kuhl, 1994).

My dissertation addresses the above controversy by examining whether the integration of heard and seen speech is obligatory for young infants as it is for adults. In Experiment 1, 4-month-old female infants habituated to audiovisual /bi/ showed renewed visual interest to an auditory /bi/-visual /vi/ suggesting that they may have perceived the auditory /bi/-visual /vi/ as /vi/, as do adults. In Experiment 2, neither male nor female infants showed renewed visual interest to a dishabituation stimulus which represents only a change in mouth movements. In Experiment 3, male infants looked longer to an audiovisual /bi/ than to an audiovisual /vi/ following habituation to an audio /bi/-visual /vi/, while female infants tended to look only slightly longer to an audiovisual /vi/ than to an audiovisual /bi/.

Taken together these experiments suggest that at least some infants are able to integrate heard and seen speech, but that they do not do so consistently. Although an innate mechanism may be responsible for integration, a role for experience is suggested as integration does not appear to be obligatory for young infants as it is for adults.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to thank all the parents who brought their infants in for these studies; without their assistance, this research would not have been possible. I also would like to thank Dawn Brandlmayr, who assisted in recruitment and data collection and who handled with graciousness all my requests for this or that to be sent to me. I am also grateful to Erina Sim for help with data collection and to Sonya Bird and Aleta Cooney for coding. I would like to thank Judi Pegg and Christine Stager for all their helpful suggestions and overall supportiveness. Thanks also to Marie Habke who has supported me with patience and humour throughout graduate school.

I would like to thank my thesis committee: Richard Tees, for re-awakening my love of perceptual development; Geoff Hall, for his constant encouragement and help with statistical issues; and Janet Werker, my advisor, for her enthusiasm, for her enduring support in all forms, and also for allowing me to discover for myself just how special speech is.

Lastly, I cannot say a big enough thank-you to my husband, Ross McKitrick, and our daughter, Madeleine. Their support, patience and encouragement has made all the difference in the world.

This dissertation is dedicated to Ruth Henderson Desjardins (1921-1996), who always wanted to be a scientist.

INTRODUCTION

The perception of speech occurs through more than just the auditory channel. Anyone familiar with the hearing-impaired understands that information provided by a speaker's mouth movements can facilitate comprehension. However, what is less well-known is that seen speech can be used to augment speech perception even in those with normal hearing. For example, in a noisy environment, watching a speaker's mouth movements significantly reduces the errors made by listeners (Binnie, Montgomery & Jackson, 1974).

Typically, in a noisy environment certain patterns of auditory confusions among English consonants are evident. In particular, identification of one articulatory feature, place of articulation,[1] suffers more than others at poor signal-to-noise ratios; people tend to readily confuse the different places (Miller & Nicely, 1955). Conversely, place of articulation is thought to be the feature most readily recoverable from watching a speaker's mouth movements. Adults are able to categorize consonants into five place-of-articulation groupings when only the speaker's mouth movements are seen but no speech is heard (see Table 1) (Binnie, Montgomery & Jackson, 1974). Thus, vision complements audition: what is most easily confused by the ear is most easily recovered by the eye. Access to a speaker's articulating mouth enhances both adults' (Binnie et al.) and children's (Dodd, 1977) perception of speech.

---

[1] In order to form consonants, the lips, tongue tip or blade, or the back of the tongue are used to restrict the air flow through the vocal tract; this feature is known as place of articulation (Ladefoged, 1993).

Table 1

Five Discrete Categories of Visual Syllables†

Visual Category	Consonants
bilabial	/p, b, m/
labiodental	/f, v/
interdental	/θ, ð/
rounded labials	/ʃ, ʒ/
linguals	/s, z, t, d, n, k, g/

†Adapted from Binnie et al. (1974).

The Nature of Auditory-Visual Integration of Speech

A compelling demonstration that the perception of speech can be strongly influenced by watching a speaker's mouth movements as well as by listening to the auditory signal is seen in the "McGurk effect." Pairing a video of a face saying /ga/ in synchrony with a sound tract of /ba/ results in reports of a 'heard' /da/—a syllable intermediate in place of articulation between /ba/ and /ga/.[2] In some cases, pairing of mismatched or incongruent auditory and visual stimuli may result in a percept which is entirely 'captured' by the visual signal, referred to as visual capture. For example, an auditory /ba/ when paired with a face saying /va/ results in reports of a heard /va/—the resulting percept is captured by the visual /va/. In both cases, adults typically report a unified percept—they are unaware that the auditory and visual syllables are mismatched.

[2]In some cases, such as a visual /ba/ paired with an auditory /ga/, adults report hearing /bga/—both the auditory and visual consonants are maintained.

One notable characteristic of audiovisual speech perception is that integration is not under conscious control but rather appears to be mandatory. Adults instructed to watch a speaker's articulation but to ignore what they see and only say what they hear report percepts influenced by both visual and auditory channels (Summerfield & McGrath, 1984). Moreover, other cues that the auditory and visual channels might be providing discrepant information do not decrease the strength of the effect. For example, Green, Kuhl, Meltzoff and Stevens (1991) found that the McGurk effect was just as strong when a male voice was dubbed onto a video of a female face (and vice versa) as when the voice and face represented talkers of the same gender. This effect held even though participants clearly noticed the discrepancy between face and voice in cross-gender conditions. More recently, Walker, Bruce and O'Malley (1995) replicated the finding that cross-gender pairings does not impact the strength of the McGurk effect, but only when unfamiliar faces are used.

The nature of audiovisual integration of speech is such that participants need not consciously recognize the visual display as a face for integration to occur. A dynamic point-light display composed of 28 moving "dots" (attached to a speaker's cheeks, chin, nose, teeth and tongue tip) showing the articulatory gesture "va" when accompanied by an auditory /ba/ was sufficient to induce visual capture—a percept of /va/ (Saldaña, Rosenblum & Osinga, 1992). Some participants apparently did not even recognize that the visual display was that of a face in motion.

Audiovisual speech perception is one of several lines of evidence used to support the existence of a specialized speech processing module.[3] In the revised version of the motor theory of speech, Liberman and Mattingly (1985) propose that the listener recovers the speaker's *intended* phonetic gesture which is specified by *invariant* neural commands signaling the appropriate movements of the articulators. The link between production and perception is innately specified; and a specialized processing module[4] of the sort proposed by Fodor (1983) allows us to recover the intended phonetic gesture. As a module, it has certain characteristics: domain-specific, mandatory, informationally encapsulated, largely innate, and a fixed neural architecture. Audiovisual speech perception seems to share some of the characteristics typically associated with a module,[5] and

---

[3]Two aspects of the speech signal in particular—lack of invariance and coarticulation—have lead theorists (e.g., Liberman, Cooper, Shankweiler & Studdert-Kennedy, 1967; Liberman & Mattingly, 1985) to conclude that the processing of speech must be special. Lack of invariance is apparent when one examines the acoustic patterns (as seen in spectrograms) of the same phoneme in different vowel contexts. For example, the second formant transition of a /d/ in a /i/ vowel context is rising while in a /u/ context is falling, yet in both cases the hearer perceives the consonant /d/. Coarticulation of phonetic segments means that in the acoustic signal the information from one phoneme overlaps with information from another phoneme. Furthermore, the degree of overlap varies depending on the context. Yet, we perceive discrete phonetic units. The special problems of processing speech suggest to some a specialized processing module.

[4]The idea of a specialized sensorimotor module is not unique to speech perception. Goodale (1988, p. 264) argues that the visual system is characterized by "a network of relatively independent sensorimotor channels, each of which supports a particular kind of visually guided behavior." He suggests that the sensori*motor* nature of the visual system has been underrepresented; researchers have tended to examine instead the representation of the external world as provided by the visual system.

[5]Summerfield (1991) suggests that audiovisual speech perception may be subserved by a sub-module—a module within a module—of the speech perception module. If lip-reading is a submodule distinct from the (auditory) speech perception module, this would make it easier to account for the fact that adults vary much more in their ability to lip read than they do in their ability to perceive speech aurally.

Liberman and Mattingly argue that findings from audiovisual speech perception are consistent with their motor theory of speech perception.[6]

That the integration of audiovisual speech is likely the result of a speech processing system rather than a more general integration process is revealed by studies on the effects of asynchrony on integration. The McGurk effect holds under conditions of audiovisual asynchronies of up to 225ms (Gerdeman, 1994; Munhall, Gribble, Sacco & Ward, 1996) while the perception of an auditory signal and a visual signal as emanating from the same location—the "ventriloquism effect"—is disrupted if the auditory and visual events are asynchronous (e.g., Jack & Thurlow, 1973). Furthermore, responses to incongruent audiovisual speech stimuli are affected by inverting a speaker's face (Bertelson, Vroomen, Wiegeraad & de Gelder, 1994; Green, 1995) while the degree of displacement of the sound source in the ventriloquism effect is not (Bertelson et al.). Such findings suggest that integration of auditory and visual speech information is a distinct process from the integration of nonspeech auditory and visual information.[7]

Moreover, the processing of audiovisual speech perception appears to be different from the processing of faces. Partial but not complete independence of

---

[6]Audiovisual speech perception data have also been used to support a Gibsonian approach to perception. The Gibsonian and Motor Theory approaches are similar in that both propose that speech perception involves recovering the articulatory gesture, however, the Gibsonians do not think that speech perception is "eccentric" nor that it requires a special processing module (e.g., Fowler & Rosenblum, 1991; Rosenblum, 1994). For the Gibsonians, articulatory gestures are distal events which through lawful relationships give rise to a proximal stimulus—the acoustic signal and the visual signal. Thus, "the impressive evidence that audiovisual integration is a mandatory property of perception can be appreciated without reference to modularity" according to Rosenblum (1994, p.115). Speech perception is like the perception of any other distal event.

Audiovisual speech research has also been used to support a Fuzzy Logical Model of Perception (see Appendix A.)

[7]According to the Gibsonian theory, (e.g., Rosenblum, 1994), the apparent dissociation of audiovisual speech and the ventriloquism effect—when an auditory stimulus and a visual stimulus which are displaced spatially but co-occur temporally are perceived as emanating from the same source—can be explained without reference to modularity: the former involves the identification of an *event* whereas the latter involves the identification of a *location*. Thus, one would expect temporal asynchronies of audio and visual information to disrupt the identification of an event to a lesser degree than the determination of a location. Note that although the dissociation of these phenomena does not require reference to distinct modules, the data are not inconsistent with the modularity proposal either. Moreover, the Gibsonians have not provided new empirical evidence in the area of audiovisual speech perception supporting their proposal.

facial recognition and processing of visible speech was demonstrated in a typical adult population. Walker, Bruce and O'Malley (1995) used faces which were familiar to one group of adults and unfamiliar to another group. Stimuli were produced in which (1) the face and voice from the same person were paired, (2) the same face was paired with a different voice of the same gender and (3) the same face was paired with a different voice of the other gender. When an audio /ba/-visual /ga/ or audio /bi/-visual /gi/ was used, adults unfamiliar with the faces did not differ in the proportion of responses showing visual influence reported across the three types of face-voice pairings. Adults familiar with the faces reported the same proportion of visual influence as adults unfamiliar with the faces only when the face and voice from the same person were paired. Those participants familiar with the faces reported significantly less visual influence when the face and voice came from different persons either of the same or different genders. This pattern of results suggests that perception of visible speech and identification of faces are not completely independent processes in adults.

In an experiment comparing children with autism to normal controls matched for verbal ability, de Gelder, Vroomen and van der Heide (1991) found a dissociation between lip-reading ability—identifying a syllable or a word by means of watching a speaker's visible articulation without the accompanying sound—and face recognition. For the children with autism, there was no relationship between performance on a lip-reading task and recognition of faces whereas for controls, significant correlations were noted. The fact that for the control group of children lip-reading and face recognition were correlated is consistent with Walker et al.'s (1995) findings that in typical adults, these two processes are not completely independent, as described above. Interestingly, with respect to the integration of auditory and visual signals, unlike controls, children with autism showed very little

influence of the visible information in the audiovisual speech perception condition although they were good at lip-reading.

The perception of visible speech also appears to dissociate from perception of emotional expression. A patient with right-hemisphere (RH) damage and prosopagnosia was impaired on the identification of emotional expression from dynamic point-light displays but normal on all tests of lip-reading; a patient with left-hemisphere (LH) damage who performed above chance on the emotional display identification showed impairment on some aspects of lip-reading and was not susceptible to McGurk effects (Campbell, 1992; Perrett (discussant) cited in Campbell, 1992).

Although lip-reading does not appear to be tied to perception of emotional expressions, lip-reading does appear to be linked to speech processing as indicated by neuropsychological evidence from four patients with presumed posterior lesions (Campbell et al., 1990). Two patients with LH lesions showed impaired lip-reading ability while one of two RH lesioned patients showed normal lip-reading skills and responses to McGurk-type stimuli. Campbell et al. suggest that there is an amodal phonological processor in the left-hemisphere which incorporates input from heard and seen speech. They speculate, as well, that "the phonological processor is more effectively driven from RH than LH vision reception areas" (Campbell et al., 1990, p.800).

There is some evidence that seen speech is processed in the primary auditory cortex, rather than in the occipital cortex. Magnetoencephalographic (MEG) recordings were made while participants heard either congruent or incongruent audiovisual syllables (Sams et al., 1991). No coherent activity was noted in either the occipital cortex or the occipito-temporal cortex. However, clear evidence of activity in the auditory cortex and surrounding belt areas was observed. The authors are not able to be more specific about location of

processing other than in identifying the primary auditory cortex. These findings provide some support for Campbell's (1992) notion of an amodal phonological processor in the left hemisphere.

Indirect support for the motor theory supposition that perception and production are innately linked comes from studies which use intraoperative electrical stimulation of cortical regions (see Ojemann, 1991). Ojemann (1988, 1991) found cortical sites (left superior temporal gyrus) which when stimulated, disrupted both speech perception and the production of sequential orofacial movements used in speech; however, many more sites were located which respond only to speech perception or production but not both.

The aforementioned studies suggest that audiovisual speech perception does have some of the characteristics which are typically associated with a module. For example, the integration of auditory and visual speech appears to be mandatory, informationally encapsulated, not under conscious control, relatively domain specific, and possibly subserved by dedicated neural architecture in the left hemisphere. Additionally, some cross-language studies suggest that audiovisual speech perception—although perhaps largely innate—is influenced by experience within a particular linguistic community; a phenomenon which is not inconsistent with a modular view.[8]

Native Language Influences on Audiovisual Speech Perception

Language-specific audiovisual speech perception has been demonstrated in several studies. While Massaro and colleagues (Massaro, Cohen, Gesi, Heredia & Tsuzaki, 1993; Massaro, Cohen & Smeele, 1995) find that Japanese, Spanish, Dutch and English speakers are all influenced by visible speech, Sekiyama and

---

[8]See Fodor (1985) and commentaries.

Tohkura (1993) find that Japanese speakers are much less influenced by the visible component. Using incongruent audiovisual Japanese syllables, they found that Japanese speakers were not as influenced by the visible articulation as were English speakers for the same syllables—even though the Japanese are just as good at lip-reading as English speakers. And, there were no significant differences between the two groups on English syllables. One explanation offered by Sekiyama and Tohkura for their findings is cultural: Japanese listeners tend not to look at the face of a speaker as to do so is impolite—the face-avoidance hypothesis. Thus, Japanese speakers may have had less experience integrating auditory and visible speech in Japanese than English speakers have had integrating auditory and visible speech in English.

Influence of the visible syllable has been found to be weaker in Chinese speakers than in American speakers, providing additional support for the face-avoidance hypothesis. Sekiyama (1997) tested native speakers of Chinese who were living in Japan on the same syllables as were used in the Sekiyama and Tohkura (1993) study. The Chinese in the current study were more like the Japanese in the previous study than like the Americans in the proportion of visual influence. Culturally, the Chinese people are more similar to the Japanese people in terms of the extent to which it is impolite to stare at a speaker's face. Thus, the data provide support for the face-avoidance hypothesis. Sekiyama points out, however, that in the Chinese language, and to a lesser extent in Japanese, tonal differences distinguish meaning. Thus, a stronger reliance on the auditory component is not surprising. Although further research is necessary to distinguish between these two explanations, it is clear that experience, perhaps both linguistic and cultural, may influence the extent to which heard and seen speech are integrated.

Rather than investigating cross-linguistic differences in the strength of the McGurk effect, Werker, Frost and McGurk (1992) examined how experience within a particular linguistic community influences the nature of the percept. They found that speakers of Canadian French with some knowledge of English tended to report that an audio /ba/-visual /ða/[9] as either a /da/ or /ta/ unlike native English speakers who tended to report hearing /ða/. Werker et al. suggest that the French speakers—for whom /ða/ does not have phonemic status—were assimilating the visible interdental to the closest phonemic place of articulation, thus reporting an alveolar/dental sound. These findings suggest that the perception of audiovisual speech is modified by experience within a particular language environment.

Massaro and colleagues (Massaro et al., 1993, 1995) similarly find that adults' responses to synthetic incongruent auditory and visual tokens of /ba/ and /da/ varied according to the phonological inventory of the adults' native language. For example, among the most common responses of Dutch speakers were /*v*a/ and /*v*ha/—percepts not reported by English speakers for the same stimuli. Massaro and colleagues do not offer explanations as to why certain percepts arose except to say that the percepts were influenced by the phonological inventory of the perceiver.

That experience within a linguistic community influences audiovisual speech perception is supported by research with children—who have less experience with their native language than have adults. Several studies including the original McGurk and MacDonald (1976) paper report on findings when children have been tested with auditory-visual mismatches. McGurk and MacDonald (1976) note that children (3- to 5-year-olds and 7- to 8-year-olds) tend to be less influenced by the visual signal than are adults; however, children do respond in a similar manner to adults. Using a /ba-da/ continuum of sound and a face articulating either /ba/ or

---

[9] 'ð' is the phonetic symbol for the 'th' sound in 'the.'

/da/, Massaro (1984; Massaro, Thompson, Barron & Laren, 1986) also demonstrated that preschool children are less influenced by the visual signal than are adults despite the fact that the children are able to lip-read /ba/ and /da/ with accuracy in a visual only condition. Hockley and Polka (Hockley, 1994; Hockley & Polka, 1994), using a cross-sectional design, found increasing influence of the visual information as children increased in age (from 5 to 11 years). Taken together, these studies all support an interpretation that a relative lack of experience with language causes children to be less influenced by visible speech than are adults.

What remains unclear, however, is what kind of experience is critical. Children have experience *listening* to and *observing* others produce speech as well as *producing* speech themselves. Some recent research with preschoolers (Desjardins, Rogers & Werker, in press) suggests that experience producing speech may be an important kind of experience for the development of audiovisual speech perception. Preschoolers were divided into two groups on the basis of whether they made articulation errors on consonants. While both groups of preschoolers showed consistently less influence of visible speech on their perception of audiovisual tokens than did adults, preschoolers who made articulation errors showed considerably less visible influence than did preschoolers who did not make errors. These findings suggest that experience correctly producing consonants enhances the underlying representation of visible speech.

Additional support for the idea that experience producing speech develops the underlying representation comes from an experiment comparing adults who are unable to produce speech due to severe cerebral palsy with typical adults. Siva, Stevens, Kuhl and Meltzoff (1995) found that both groups showed comparable visible influence when tested on an auditory /aba/-visual /aga/, i.e.,

they reported perceiving /ada/ or /aða/.[10] However, when tested on an auditory /aga/-visual /aba/, the two groups differed in their responses. As expected, the adult controls reported a 'combination' percept of /abga/. The cerebral palsy adults, on the other hand, reported /aga/—a percept influenced by only the auditory stimulus. These results suggest that experience producing speech influences the ability to perceive at least some instances of audiovisual mismatches.

Matching of Auditory and Visual Signals in Infancy

What remains unclear is whether experience producing speech is a necessary prerequisite for the integration of heard and seen speech. Perhaps, only the experiences of listening to and observing speech produced are required for the integration of heard and seen speech to occur. One population which has had some experience listening to and observing speech produced, although no experience producing consonants, is young infants.

There are many studies indicating that infants, by 4 months of age, can *match* auditory and visual nonspeech stimuli; they know which of two visual stimuli matches the heard auditory stimulus. Using a preferential looking procedure, Spelke (1976) found that 4-month-olds looked longer at a visual event, a game of peekaboo or a percussion musical sequence, which was specified by a soundtrack. Infants are able to identify the visual event which matches the auditory sequences when the soundtrack is presented simultaneously with the impact of an object, or with the moment of direction change of an object or with the moment of passing through a certain spatial location (Spelke, 1979; Spelke & Born, 1983). Moreover, temporal synchrony is not the basis for matching: when both visual

---

[10]The cerebral palsied adults reported their percepts using their computerized communicators while the control group reported their percepts orally.

events occur simultaneously—wet sponges squishing and blocks banging—infants correctly look at the event which corresponds to the sound track—squishing or clacking (Bahrick, 1983).

There are a few studies using faces and voices which indicate that infants are sensitive to amodal properties such as temporal synchrony.[11] By 4 months of age, infants are able to detect when an articulating face and the accompanying speech are out of synchrony; and they prefer to listen to speech which is synchronous with the articulating face (Dodd, 1979; Spelke & Cotelyou, 1981). These findings are consistent with Lewkowicz's (1994) report on the developmental sequence of responsiveness to temporal attributes in the auditory and visual modalities. He notes that sensitivity to the synchrony of nonspeech stimuli across modalities appears around 4 months of age.[12]

Only a handful of studies involving speech suggest that infants are sensitive to other features of the relationship between faces and speech. Walker (1982) found that 5-month-olds are sensitive to the relationship between emotional expression in the face and in the voice. In a preferential looking task, infants

---

[11] It is difficult to know whether the properties of the speech signal should be classified as amodal. E.J. Gibson defines amodal information as that which is "not tied to specific sensations but is rather invariant over them" (E.J. Gibson, 1969, p. 219). Temporal events, location, intensity, rhythm/patterns, form and movement are considered amodal properties. From a Gibsonian perspective, Walker-Andrews (1994) suggests a new classification scheme for intersensory relations: amodal, artificial/arbitrary, arbitrary/natural and typical. These categories differ in the degree to which the relations between stimulus arrays and the distal object are perceptible or detectable. The category artificial/arbitrary includes those mappings which are usually referred to as polymodal associations, for example, a siren and a fire engine. The typical category is an attempt to capture other kinds of lawful relations: for example, heavy objects (a large rock) when dropped tend to make a loud sound on impact whereas small, lightweight objects (a feather) make a soft sound on impact. Walker-Andrews places audiovisual speech perception in the category arbitrary/natural because "vocalizations result from a combination of structural properties that are not always visible to the observer" (Walker-Andrews, 1994, p. 47). She also refers to this category as idiosyncratic but natural. One might expect that infants may initially find arbitrary/natural relations more difficult to detect than amodal ones. The maturity of the different sensory modalities, the type of task and the category of relation all determine whether and when infants will be able to detect the invariant relations involved.

[12] More recently, however, contrary to previous findings, Lewkowicz (in press) reported that auditory-visual asynchrony of voice and articulating face did not appear to aid 4-month-old infants' detection of a change in the auditory track in an habituation study. One possible explanation for this failure may be due to the fact that in 2 of the 3 studies, the auditory component involved a speaker reading a scientific text in an adult-directed tone of voice.

correctly matched the emotional expression conveyed by the voice to the appropriate facial display—even when the soundtrack was out-of-synchrony with the articulating face.

A few studies have demonstrated that under some conditions, young infants can match a speech sound presented auditorily with the correct visible articulation. Four-month-old infants were found to look more at a face mouthing the heard vowel (/a/ vs. /i/) rather than at a face mouthing a different vowel (Kuhl & Meltzoff, 1982, 1984). Meltzoff and Kuhl (1994) suggest that young infants' babbling of vowels functions to develop the auditory-articulatory intermodal map of speech; this map is what allows infants to match correctly a seen articulation with a heard vowel sound (Kuhl & Meltzoff, 1982; 1984).

Walton and Bower (1993) addressed the question of whether infants require experience even *observing* the correspondences between different vowels and their visible articulation in order to be able to match a heard vowel with the correct articulation. In an operant choice sucking procedure, 6- to 8-month-old infants were presented with three audiovisual pairs composed of a face articulating /u/ in combination with a voice saying either an English /u/, an English /i/ or a French /y/. Each stimulus was presented alone and the three audiovisual combinations were sequenced in a fixed order. If infants do not require experience in order to match heard and seen vowels, they should find the heard French /y/ in combination with the seen English /u/ articulation to be a 'possible' match, according to Walton and Bower, because shaping the mouth to say an English /u/ while trying to produce an English /i/ gives the approximation of a French /y/. Walton and Bower found that English-learning infants produced more sucks in order to look longer at the face articulating an English /u/ when it was accompanied by an unfamiliar French /y/—a possible combination—than at the same face articulating the English /u/ when it was accompanied by an English /i/—an impossible combination. Infants did not

suck longer in order to look at the matching audiovisual pair—visible /u/ accompanied by an auditory /u/—than at the face articulating an English /u/ accompanied by a French /y/. These results suggest to Walton and Bower that infants readily determine which audiovisual pairings are articulatory possibilities. This indeed is an impressive ability; however, as Walton and Bower failed to test whether infants could discriminate a French /y/ from an English /u/, it is unclear on what basis infants performed in this task. Perhaps infants could not distinguish between the auditory tokens and thus mistook the French /y/ for an English /u/; in which case, this study provides a replication of the basic findings of Kuhl and Meltzoff (1982, 1984), but with different tokens.

Although most of the matching experiments with young infants involve vowels, there is some evidence that infants of 5-6 months of age can match heard and seen speech composed of consonant-vowel disyllables (e.g., /vava/) (McKain, Studdert-Kennedy, Spieker & Stern, 1983). When presented with the visible articulation of /vava/ on the left of midline and /zuzu/ on the right, infants showed a preference to look at the matching articulation of a heard /zuzu/. For three of the six disyllables there was a significant preference for the matching articulation, but only when it was presented on the right side. McKain et al. interpret this finding as indicating a left-hemisphere specialization for audiovisual speech perception. However, the possibility also exists that the effect for consonants is ephemeral and only shows up under some testing conditions.

Indirect evidence for infants' sensitivity to the mouth movements of speakers comes from infants' own productions of speech sounds. Dodd (1987) found that the babbling patterns of 9- to 12-month-old infants were influenced by the type of stimulation they received. Infants babbled more consonants after interacting with an adult who babbled consonant-vowel streams, but not after interacting with a silent adult or after hearing a recording of the same adult babbling.

More recently, Legerstee (1990) reported that 3- to 4-month-old infants produced more 'imitations' of an auditorily presented vowel sound (/a/ or /u/) when a nearby female adult mouthed the same vowel than when she mouthed a different one. This finding confirms what Kuhl and Meltzoff (1982; 1984) reported observing in their vowel matching task: infants tended to produce the same vowel (/a/ vs. /i/) that they heard over the speaker.

Production errors from a small number of blind children suggest that sighted children's speech is influenced by observing the articulations of others. Mills (1987) found that sighted 1-year-old children showed fewer errors on consonants with a visible articulation than did blind children. Moreover, the blind children showed more *between* visual category substitutions of consonants than did sighted children. Mills concludes that the combined cues from both vision and audition give the sighted child an initial advantage in the acquisition of correct production.

Integration of Heard and Seen Speech in Infancy

The studies reviewed above indicate that infants as young as 4 months of age *match* auditory and visual information about the same event.[13] Although matching an auditory and a visual event is an impressive ability, it does not tell us

---

[13]A rate limiting factor in this process is the development of the visual and auditory systems. In human infants as well as in the young of other species (e.g., chick, opossum, rat, cat), the different sensory systems become functional in a particular sequence: tactile, vestibular, auditory, visual (Gottlieb, 1971). The auditory system in humans is known to be functional at a gestational age of approximately twenty-five to twenty-seven weeks (Birnholz & Benacerraf, 1983). However, the earliest onset of visual function is not known. Physiological studies indicate that the organ of Corti (the auditory receptor) is completely differentiated several months before birth whereas the retina (the visual receptor) is not fully developed even at birth (Gottlieb, 1971). Furthermore, neocortical development differs for the two sensory systems with the auditory system being more advanced at birth (Bronson, 1982) .

Even though the auditory system is initially more advanced, rapid development of the visual system allows it to "catch-up" early in infancy. The myelination of neurons to the visual cortex occurs rapidly after birth such that by 3 months of age the visual and auditory areas are equally developed. And by the end of the first year, the visual system becomes more advanced than the auditory system (Bronson, 1982). It is not entirely clear how histology maps onto function; however, by six months of age, infants' visual sensitivity is almost adult-like and controlled switches in attention are possible (Atkinson, 1984).

whether infants' perception of the auditory event is influenced by attending to the visual event. In other words, if an infant hears a syllable while watching the speaker produce that syllable, is the infant's perception of the syllable influenced by watching the speaker's visible articulation? To know that the sound /ba/ goes with a particular articulatory pattern does not imply that the syllable perceived by the infant is influenced by both the speech she hears and the articulation she sees. Thus studies of infants' ability to match auditory and visible speech do not directly test the question of whether infants' perception of speech is influenced by the visual channel such that they are subject to McGurk effects. The matching studies, however, might lead us to predict that with limited experience infants' perception of speech is influenced by a speaker's mouth movements, although perhaps somewhat less so than is adults'.[14] Evidence indicating that young infants' percepts are influenced by visible speech would suggest that some innate underlying representation of visible speech, consistent with the motor theory of speech perception (Liberman & Mattingly, 1985) is responsible for the integration of heard and seen speech.

There are currently two reports in the literature of attempts to show the integration of auditory and visible speech in 4-month-old infants. Burnham and Dodd (in press) used a live model to mime one syllable in synchrony with a prerecorded auditory syllable. An experimental group was habituated to an incongruent audio /ba/-visual /ga/ while a control group was habituated to a congruent audiovisual /ba/. Three test trials with only auditory syllables—the face remained stationary—were presented once habituation had occurred: /ba/, /da/ and /ða/. Pretesting with adults in both the live model condition and a videotaped stimulus condition revealed that, as expected, adults perceived the incongruent audio /ba/-visual /ga/ as either /da/ or /ða/ on the majority of trials. However, the

---

[14]See Appendix B for a description of different developmental models of intersensory processing.

live model was not more effective than the videotaped stimulus; thus it is curious that the authors were willing to sacrifice control by using a live model for testing infants. Moreover, as they did not videotape the model, it is not possible to determine whether she mimed in synchrony with the auditory syllable on each trial.

Two derived measures were used as the dependent variables: an auditory percept score and a 'fused' percept score (representing integration of auditory and visual). Formulae for the derived scores are as follows: Auditory Percept Score = 100 |[ba] - 1/2 [da] - 1/2 [ða] |/ Range; and, Fused Percept Score = 100 |[da] - 1/2 [ba] - 1/2 [ða] |/ Range, or =100 |[ða] - 1/2 [ba] - 1/2 [da] |/ Range—which ever yields the greater value. Burnham and Dodd do not present how they arrived at these particular formulae. The authors' rationale for using derived measures rather than simply testing for renewed visual interest is the following: (1) the test trials were presented in a single modality while habituation stimuli were bimodal, and a modality change alone can elicit renewed visual interest; and, (2) the incongruent audio /ba/-visual /ga/ is perceived by some adults as /da/ and by other as /ða/, thus for some of the infants /da/ will be familiar and /ða/ novel and for others the reverse will be true.

Using the derived measures, the authors found a significant interaction between group and score: the control group showed relatively higher auditory scores and lower fusion scores than did the experimental group. Burnham and Dodd interpret this finding as indicating that infants in the experimental condition heard the habituation stimulus, audio /ba/-visual /ga/, as either /da/ or /ða/ and not as /ba/ whereas infants in the control group, as expected, heard the audiovisual /ba/ as /ba/. Visual inspection of a figure suggests that this interaction may be entirely due to a relatively large difference in auditory fusion scores (approximately 45% and 65% for the experimental and control groups respectively) and a

relatively small difference in fusion (integration) scores (approximately 85% and 75% for the experimental and control groups respectively) between the two groups.

Using the derived scores to infer how infants perceived the habituation stimuli seems problematic. The following is an example of the auditory and fusion scores resulting from the data of an hypothetical infant. An infant in the control group might look on the test trials for 5 s, 10 s, and 12 s, for /ba/, /da/ and /ða/ respectively—this pattern of looking suggests that for this infant /ba/ is more familiar and /da/ and /ða/ are more novel hence the infant looks longer on the latter two trials. Plugging these numbers in the formulae would yield an auditory percept score of 86%, and fused percept (integration) scores of 21% and 64%. Note the substantially different fused percept scores; the authors do not say why they choose the larger value for their analyses. If this hypothetical infant had been in the experimental group who was habituated to audio /ba/-visual /ga/, the relatively large fusion score (64%) is surprising given that the actual looking time data indicates that the infant looked longer to /da/ and /ða/ than to /ba/—a pattern which would suggest that this infant did not integrate the audio /ba/-visual /ga/ but rather perceived it as /ba/. Intuitively, it would seem that this infant should have a much lower fused percept score to reflect that fact that she did not seem to integrate the audio /ba/-visual /ga/.

An alternative to the use of derived scores would be to examine relative differences in looking time to the three test stimuli. If infants' perceived the habituation stimulus, audio /ba/-visual /ga/, as /da/ (the syllable most frequently reported by adults), then on the test trials, infants should look relatively longer to the stationary face on the auditory /ba/ trial than on the auditory /da/ trial. If, on the other hand, infants perceived the habituation stimulus as /ba/, they should look relatively longer to the stationary face on the auditory /da/ or /ða/ trials than on the auditory /ba/ trial. This type of analysis in addition to the derived scores would

have made the report more convincing. As it stands, this study provides equivocal evidence at best for the integration of heard and seen consonants.

It is difficult to demonstrate integration of an auditory /ba/ and a visual /ga/ in infants for the reason that adults show two distinct responses: they hear either /da/ or /ða/. Thus, Rosenblum, Schmuckler and Johnson (1997) chose instead to use as their incongruent audiovisual stimulus, audio /ba/-visual /va/; their pretesting indicated that adults hear this stimulus as /va/ 98% of the time. Rosenblum et al. habituated twenty 4-month-old infants to an audiovisual /va/—the lower half of a man's articulating face accompanied by synthetic auditory tokens—and then after habituation criterion was met, infants were given two trials of each of two stimuli: audio /ba/-visual /va/ and audio /da/-visual /va/ in alternating fashion with starting stimulus counterbalanced. The habituation stimuli varied in pitch (same contour, but four different average fundamental frequencies); whether the frequency contours of the test stimuli were identical or varied is not mentioned.

The basic finding reported is that by summing infants' looking time across each of the two test trial types separately and the last two habituation trials, infants habituated to audiovisual /va/ showed renewed visual interest to the audio /da/-visual /va/ (perceived as /da/ by adults), but not to the audio /ba/-visual /va/ (perceived as /va/ by adults). These findings suggest that infants may perceive the audio /ba/-visual /va/ as similar to the habituation stimulus of audiovisual /va/ and moreover, that they perceive the audio /da/-visual /va/ as /da/ and not as /va/.

One possible explanation raised by the authors for this pattern of results is that infants might not be integrating the audio /da/-visual /va/, but instead perceiving two separate syllables, a /da/ and a /va/, which might be interesting because it is odd. To test this hypothesis, Rosenblum et al. conducted a second experiment using a sequential preferential looking paradigm. Infants were presented with three trials of audio /ba/-visual /va/ and three trials of audio /da/-

visual /va/ in alternating order with starting stimulus counterbalanced across infants. Infants did not show a significant preference for audio /da/-visual /va/ over audio /ba/-visual /va/. This pattern of results in Experiment 2 suggests that infants do not find audio /da/-visual /va/ to be more interesting or odd than audio /ba/-visual /va/.

In Experiment 3, the authors test a second alternative explanation for the pattern of findings in Experiment 1 (renewed visual interest to audio /da/-visual /va/, but not to audio /ba/-visual /va/ following habituation to audiovisual /va/): that infants were attending only to the auditory stimulus and that they simply find an auditory /va/ to be more similar to an auditory /ba/ than to an auditory /da/. Indeed, Miller and Nicely (1955) find that at six different signal-to-noise ratios, adults more readily confuse an auditory /va/ with /ba/ than with /da/—a pattern consistent with the results of Rosenblum et al.'s Experiment 1. In Experiment 3, infants were habituated to a smiling face accompanied by an auditory /va/ and then tested on an auditory /da/ and an auditory /ba/ each presented twice, as in Experiment 1. Infants looked significantly longer to the stationary face on the auditory /ba/ trial following habituation to the auditory /va/, but infants did not look significantly longer to the stationary face on the auditory /da/ trial. There was no difference in looking time to the stationary face on the auditory /ba/ trial compared to the auditory /da/ trial.

The authors suggest that because infants in the Experiment 3 could discriminate an auditory /ba/ from an auditory /va/, that in Experiment 1 the failure to recover to an audio /ba/-visual /va/ after habituation to audiovisual /va/ is due to the fact that infants perceived the audio /ba/-visual /va/ as /va/. However, it is difficult to explain how infants habituated to audiovisual /va/ come to show renewed visual interest to an audio /da/-visual /va/—supposedly perceived as /da/ by adults—if they cannot discriminate an auditory /da/ and an auditory /va/ as shown in Experiment 3.

Rosenblum et al. present a fourth experiment designed to check that adults are able to identify accurately the auditory only tokens and to determine how adults perceive the incongruent audiovisual tokens. As expected, adults perceived the audio /ba/-visual /va/ as /va/ on the majority of trials (98.6% of the time) and audio /da/-visual /va/ as /da/ on the majority of trials (96.5% of the time). Of interest is the finding that for the auditory only tokens, adults were very accurate at identifying /da/ (99.9% correct), but less accurate at identifying /va/ (92.9% correct) and /ba/ (83.5% correct). The authors do not report whether accuracy on the three auditory only tokens differed statistically. Interestingly adults were least accurate on /va/ and /ba/—the only two syllables that infants in Experiment 3 were able to discriminate. Experiment 4, then, fails to shed light on why infants in Experiment 3 were unable to discriminate auditory /da/ and auditory /va/. Moreover, as infants and adults were not tested on the same task, it is not possible to compare directly the results of the two experiments. Thus, the reader is still left wondering why infants in Experiment 1 showed renewed visual interest to audio /da/-visual /va/ (perceived by adults as /da/) after habituation to audiovisual /va/ if infants cannot discriminate /va/ and /da/.

In summary, although both the Burnham and Dodd (in press) and Rosenblum et al. (1997) studies provide some preliminary evidence that infants of 4 months of age may be able integrating heard and seen speech, the studies are inconclusive and thus further research is necessary to clarify whether young infants integrate auditory and visible speech and to rule out alternative hypotheses.

The Present Research

The present series of experiments attempts to address the issue of whether the integration of auditory and visible speech occurs in young infants. It is possible

that experience producing consonants is a necessary prerequisite for the development of an underlying representation of visible speech which then supports the mandatory integration of heard and seen speech; therefore, mandatory integration may not occur in young infants who do not yet produce consonant sounds.

To test whether the integration of heard and seen consonants occurs in young infants, 4-month-old infants were tested in an habituation-dishabituation paradigm to see whether they would systematically show visual capture. Infants were habituated to one syllable and then tested on a new syllable. The logic of the habituation procedure is that infants should show a recovery in looking time only to a stimulus that is perceived as different from the familiarization stimulus.

Research to date provides only equivocal results for both matching and integration of consonant information across auditory and visual modalities. McKain et al. (1983) showed that infants of 5-6 months of age could match some consonant-vowel disyllables, but only when the visible articulation occurred in the right visual field. Burnham and Dodd (in press) and Rosenblum et al. (1997) have both provided some evidence for the integration of heard and seen consonants which is only partially supported by the data and alternative explanations for their findings have not been ruled out.

Moreover, previous research as reviewed above suggests that children are much less influenced by the visible component of an incongruent audiovisual syllable than are adults. Extrapolating down in age into the infancy period would lead us to conclude that it may be difficult to show systematic influence of the visual syllable across a group of infants.

One factor which might be important especially if integration is difficult to show in young infants is the sex of the infants. Sex differences in the rate of habituation or in the magnitude of recovery of visual interest to a novel stimulus

following habituation have been observed in some experiments (see Tighe & Powlison, 1978, for a review). The authors of the two previous attempts to show that infants integrate heard and seen speech did not report findings for male and female infants separately. Burnham and Dodd (in press) do not indicate the sex distribution of the infants in their study and do not analyze for potential sex differences. While Rosenblum et al. (1997) included roughly equal numbers of male and female infants in each of their experiments, they failed to include sex as a factor in their analyses. Thus, these two experiments do not shed light on whether both male or female infants show integration of heard and seen speech. In the present experiments, therefore, data from both male and female infants will be examined separately in order to determine whether both sexes show the same pattern of results.

EXPERIMENT 1

In this study, infants were habituated to either an audiovisual /bi/ or an audiovisual /vi/ and then tested on a novel syllable, an audio /bi/-visual /vi/ which typically results in visual capture for adults. If infants' perception is influenced by watching a speaker's lip and mouth movements, infants habituated to the audiovisual /bi/ should show renewed visual interest to the audio /bi/-visual /vi/ while infants habituated to audiovisual /vi/ should not. This pattern of results would suggest that infants perceive the test stimulus, an audio /bi/-visual /vi/, as /vi/—as do adults. Moreover, this would provide support for the hypothesis that there is an innately-specified underlying representation which supports the integration of auditory and visual information about speech.

On the other hand, if infants habituated to an audiovisual /bi/ fail to show renewed visual interest to the audio /bi/-visual /vi/, this pattern of results would suggest that infants' perception is not reliably influenced by visible speech at 4 months of age. Indirect support would be provided for an alternative account of the development of audiovisual speech perception such as that put forward by Meltzoff and Kuhl (1994) suggesting that experience babbling the relevant speech sounds is necessary, or that suggested by McGurk (1988) that articulatory skill and the acquisition of the phonology of the native language are necessary prerequisites.

An habituation-dishabituation paradigm (Horowitz, Paden, Bhana & Self, 1972) was chosen (as in Burnham & Dodd, in press, and in Rosenblum et al., 1997) because in this paradigm, the onset and offset of the stimuli are tied to the infant's looking behaviour; an infant can hear the speech *only* when she/he is looking at the speaker's face. Four-month-old infants were chosen because, by 4 months of age infants' visual acuity has developed sufficiently to resolve the detail of speech gestures, and previous research suggests that by 4 months of age

infants are able to detect: polymodal associations (e.g., wet sponges & squishing sounds), amodal properties (e.g., temporal synchrony) of nonspeech and speech, and match heard and seen vowel sounds. Moreover, both Burnham and Dodd (in press) as well as Rosenblum et al. (1997) attempted to show integration of auditory and visible speech in 4-month-old infants.

Method

Participants

The participation of parents and their infants was arranged by telephone. Parents initially expressed interest in participating in an experiment when contacted in person by a research assistant shortly after the birth of their child at a local maternity hospital; or parents phoned in response to a radio, television or newspaper public service announcement. To express our gratitude for their participation, infants were awarded an "Infant Scientist Degree" certificate and given a t-shirt bearing our Infant Scientist logo.

The final sample consisted of 60 infants (30 male, 30 female) ranging in age from 4 months 0 days to 4 months 31 days (M = 4 months 18 days, SD = 8 days). The male infants did not differ significantly in age from the female infants, $t(58) = 1.533$, $p > .1$ (M = 4 months 18 days, SD = 9 days and M = 4 months 17 days, SD = 8 days for male and female infants respectively). All infants were full term and, according to parent report, had had no more than one ear infection since birth and were healthy at the time of testing. All infants came from homes in which English was spoken at least 90% of the time.

Infants were randomly assigned to one of three conditions, two experimental and one control condition, such that there were 10 male and 10 female infants in each condition. An additional 32 infants were excluded for the following reasons:

cried (8), refused to participate (3), experimenter error (5), parent distracted infant (1), equipment failure (4), did not meet habituation criterion[15] (6), outlier[16] (5).

Stimuli

Audiovisual tokens of an adult female speaking syllables /vi/, /bi/ and /shu/[17] (used in the pretest and posttest trials—see below) were taken from a prerecorded laser disc produced by L.E. Bernstein and S.P. Eberhardt (Johns Hopkins University). Incongruent audiovisual tokens were made in the editing suite of the Computer Vision Laboratory (with the assistance of Marc Romanycia) at the University of British Columbia. An auditory /bi/ was laid down with a visual /vi/ such that the release of the /bi/ corresponded with the parting of the lips to form a /vi/. Recordings of congruent /bi/, /vi/ and /shu/ and incongruent audio /bi/-visual /vi/ were then transferred to a laser disc.

The incongruent stimulus—audio /bi/-visual /vi/—was chosen because the bilabial stop, /bi/, when paired with a labiodental, /vi/, typically gives rise to a reliable percept of /vi/ in adults (e.g., Manuel, Repp, Liberman & Studdert-Kennedy, 1983). And, pretesting with adults confirmed that our instance of audio /bi/-visual /vi/ typically results in a percept of /vi/: Seven out of 8 adults reported a percept of /vi/ on the majority of trials (96% of the trials) while only one participant consistently reported hearing /bi/. See Appendix C for details of pretesting with adults. As well, the /i/ vowel context was chosen as McGurk effects may be stronger in the /i/ vowel context rather than in the /a/ or /u/ vowel contexts (Green, Kuhl & Meltzoff, 1988).

Segments of 60 seconds of each audiovisual signal (congruent /shu/, /vi/, /bi/ and incongruent audio /bi/-visual /vi/) were made by copying a 2 second segment

---

[15]Criterion for habituation: 3 successive trials the mean of which is 65% of the mean of the initial 3 trials. Habituation was said to occur if this criterion was met within any 3 of 12 trials following the initial 3 trials.

[16] Infants who scored more than 2 standard deviations above the mean in average looking time to the last three habituation trials or more than 2 standard deviations above the mean in looking time on one of the dishabituation trials were considered outliers.

[17]The phonetic symbol for the 'sh' sound is /ʃ/.

30 times to make one long segment. This resulted in 4 segments—one for each syllable—with the same syllable repeated at 2 second intervals (e.g., bi, bi, bi.....bi). A 60 second recording was also made of a red flashing light (no sound).

Equipment

A custom designed Hypercard habituation program (by Marc Romanycia) run on a Mac IIfx controlled the delivery of audiovisual stimuli read from a Sony LDP 1550 laser disc to a 9 1/2 by 12 in Mitsubishi HC 3905 video monitor in the testing room. The video image of the woman's face, subtending a visual angle of 18° on the horizontal plane, approximately 8 1/4 in (from hairline to chin) by 6 in (cheek to cheek) was located approximately 18 inches directly in front of the infant seated in a Evenflo bucket-style car seat. Sound was delivered at approximately 65-68 dB via a Bose 101 speaker located directly beneath the monitor and hidden from view by a black curtain which obscured everything but the video screen and the videocamera lens. A Panasonic PV-S770-K videocamera was used to feed a video image of the infant to a JVC TM-13CA monitor located in the control room, allowing the experimenter to record looking time on-line. Also, a tape in the videocamera recorded the image for later reliability coding.

Procedure

The infant was positioned in an infant-seat directly in front of the television monitor with his/her parent seated off to the side. The parent was instructed to refrain from directing her/his infant's attention to the screen and to soothe the infant with touch should the infant become fussy or upset during the study. Furthermore, the parent was instructed that should the infant turn to look at her/him, that the parent should smile to reassure the infant, but avoid engaging the infant in any games. However, if the infant became really fussy or upset or for any other reason the parent wished to stop the study, the parent was instructed to remove the infant from the seat and testing would be terminated. During testing, the parent listened

to female vocal music delivered over headphones to mask the speech stimuli presented to her/his infant.

A flashing red light attracted the infant's attention to the screen. As soon as the infant fixated on the screen, the experimenter signaled that a trial was to begin by depressing a key on the computer in the control room adjacent to the testing room. A different 'timer' key was used to indicate the infants' looking time to the visual stimulus. At the onset of a trial, the experimenter depressed the timer key and did not release the key again until the infant looked away. The duration of each look was calculated on-line as indicated by the length of time the experimenter depressed the timer key. Releasing this key signaled the computer to turn off the stimulus and to turn on the red flashing light.

The first stimulus presented at the beginning of the session after the infant fixates the screen is called the pretest stimulus: an audiovisual /shu/. On the next trial and on all subsequent trials until habituation criteria are met, the habituation stimulus was presented. Habituation was said to occur when the average duration of three sequential looks at the visual stimulus decreased to 65% of the average duration of the initial three looks to the habituation stimulus. Any look which was less than 2 seconds in total duration was excluded as each audiovisual instance was 2 seconds in total duration. After habituation was reached or a maximum of 15 trials presented, a novel stimulus (the dishabituation stimulus) was presented. If the infant noticed that this stimulus was different from the previous one, then the infant should show renewed visual interest. The session ended after the posttest stimulus, an audiovisual /shu/–the same stimulus as used in the pretest.

The pretest trial functions to introduce infants to the task while the posttest trial functions to ensure that infants will show renewed visual interest to an audiovisual stimulus which differs from the habituation stimulus in both consonant and vowel, thereby indicating that they are still awake and 'playing the game.'

Reliability coding. A trained coder, blind to the condition, recoded one-quarter of the sessions from the videotapes made during testing. In this way, the reliability of the experimenter's on-line recordings of infants' looking time was determined.

Conditions

One group of infants was habituated to a congruent audiovisual /vi/. A second group of infants was habituated to a congruent audiovisual /bi/. Both groups of infants were presented with a dishabituation stimulus of an audio /bi/-visual /vi/. An additional group of infants was included as a control for spontaneous recovery of visual interest: half of the infants were presented with the congruent audiovisual /vi/ on both habituation and dishabituation trials and the other half were presented with the congruent audiovisual /bi/ on both habituation and dishabituation trials.

Predictions

For adults, the dishabituation stimulus of audio /bi/-visual /vi/ reliably results in a unified percept of /vi/. That is, adults report hearing /vi/ and not /bi/–their percept is "captured" by the visual information of the lips forming a /vi/ sound. If young infants' percepts are similarly captured by the visual information, I would expect that (1) infants habituated to audiovisual /bi/ would show renewed visual interest to the audio /bi/-visual /vi/, and (2) infants habituated to audiovisual /vi/ would fail to show recovery of interest. On the other hand, if infants of 4 months of age do not attend to the visual information and are instead more influenced by the auditory information, I would expect (1) infants habituated to /bi/ would fail to show renewed visual interest to the dishabituation stimulus of audio /bi/-visual /vi/, and (2) infants habituated to /vi/ would show renewed interest to the audio /bi/-visual /vi/ dishabituation stimulus. These sets of predictions can be seen in Table 2.

Table 2

Patterns of Results (Experiment 1).

Stimuli		Coherent patterns	
Habituation	Dishabituation	Auditory only pattern	Integration of auditory & visual
Experimental conditions:			
audiovisual /vi/	audio /bi/-visual /vi/	+	–
audiovisual /bi/	audio /bi/-visual /vi/	–	+
Control condition:			
audiovisual /vi/	audiovisual /vi/	–	–
audiovisual /bi/	audiovisual /bi/		

Note: + denotes renewed visual interest on the dishabituation trial and – denotes no renewed visual interest.

Results and Discussion

Reliability Coding

A trained coder, blind to condition, measured infants' looking times on each trial for 15 randomly selected infants (25% of the sample) for a total of 151 trials. The experimenter's and the coder's looking times were significantly correlated, $r = .997$, $p < .0001$, indicating that the experimenter's on-line coding was reliable.

Habituation Trials

To confirm that infants did indeed show a decrement in looking across trials, looking time data for the criterion trials during habituation (the first three and last three habituation trials) were submitted to a mixed design analysis of variance with Sex (male, female) and Condition (audiovisual /vi/, audiovisual /bi/, control) as

between-subjects factors and Habituation Trial (1, 2, 3, 4, 5, 6) as a repeated measure.

The main effects for Sex, $F(1, 54) = 1.960$, $p > .1$, and Condition, $F(2, 54) = 0.09$, $p > .1$, were not significant, and the Sex by Condition interaction, $F(2, 54) = 0.527$, $p > .1$, was also not significant. There was as expected a significant decrement in looking time across the trials as indicated by the main effect for Habituation Trial, $F(5, 270) = 30.501$, $p < .0001$. Planned orthogonal polynomial contrasts revealed that the decrement in looking had a significant linear component, $F(1, 270) = 144.084$, $p < .0001$, as well as a significant quadratic component, $F(1, 270) = 7.798$, $p < .01$.

A significant Habituation Trial by Sex interaction, $F(5, 270) = 2.284$, $p < .05$, was obtained, as can be seen in Figure 1. Protected t tests indicated that female infants looked significantly longer than did male infants on the first trial, $t(270) = 3.650$, $p < .0005$, but not on any of the other trials (all ps $> .1$). The Habituation Trial by Condition, $F(10, 270) = 0.232$, $p > .1$, and Habituation Trial by Sex by Condition, $F(10, 270) = 0.688$, $p > .1$, interactions were not significant.

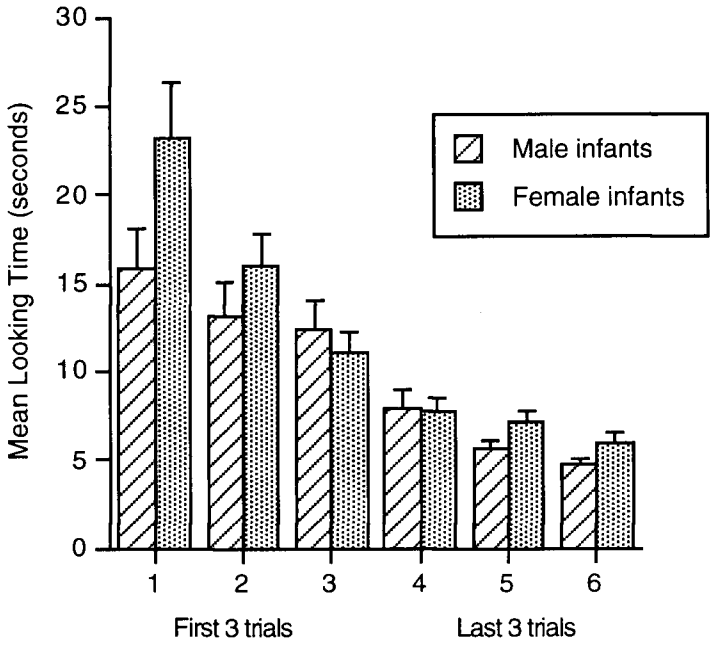

Figure 1. Mean Looking Time During Habituation on Criterion Trials (Experiment 1). Error bars represent SEM.

Trials to criterion. The minimum number of trials required was six and the maximum allowable number of trials was 15. On average infants reached criterion in 7.8 trials (SD = 2.3 trials; range 6 - 14 trials). Female infants did not differ from male infants in the number of trials to criterion according to parametric, t (58) = 0.166, $p > .1$, and non-parametric, $U = 416.5$, $p > .1$ tests (M = 7.8 trials, SD = 2.2, and M = 7.7 trials, SD = 2.4, for male and female infants respectively).

Posttest Trial

To assess whether infants would show renewed visual interest to the posttest trial after habituation had occurred, looking time data[18] were submitted to a mixed design analysis of variance with Sex (male, female) and Condition (audiovisual /vi/, audiovisual /bi/, control) as between-subjects factors and Trial (average, posttest) as a repeated measure. No significant main effects were

[18]Looking time scores for two infants were not available because one infant began to cry at this trial and the computer failed during the posttest trial for a second infant. To avoid having unequal n's, the average dishabituation looking time scores were used in place of the missing scores.

observed for either Sex, $F(1, 54) = 0.046$, $p > .1$, or Condition, $F(2, 54) = 0.052$, $p > .1$. Likewise the interaction between Sex and Condition was not significant, $F(2, 54) = 0.047$, $p > .1$.

A significant main effect for Trial was observed, $F(1, 54) = 8.410$, $p < .003$, one tailed, indicating that infants looked longer to the posttest stimulus ($M = 10.0$ s, $SD = 9.0$ s) than to the average of the last three habituation trials ($M = 6.5$ s, $SD = 2.6$ s). Neither the Trial by Sex, $F(1, 54) = 0.260$, $p > .1$, Trial by Condition, $F(2, 54) = 0.004$, $p > .1$, nor Trial by Sex by Condition, $F(2, 54) = 0.288$, $p > .1$, interactions were significant.

This analysis confirms that regardless of condition or sex, infants overall showed renewed visual interest to the posttest stimulus of /shu/. Infants were still attending to the stimuli and 'playing the game' by the end of the testing session.

Dishabituation Trial: Overall Analysis

To assess whether infants showed renewed visual interest to the dishabituation trial, looking time scores were submitted to a mixed design analysis of variance with Sex (male, female) and Condition (audiovisual /vi/, audiovisual /bi/, control) as between-subjects factors and Trial (average, dishabituation) as a repeated measure.

A significant main effect for Sex was observed, $F(1, 54) = 4.135$, $p < .05$ indicating that female infants ($M = 7.6$ s, $SD = 4.0$ s) looked longer overall than did male infants ($M = 6.0$ s, $SD = 3.2$ s). No significant main effect for Condition was observed, $F(2, 54) = 1.250$, $p > .1$. The Sex by Condition interaction was not significant, $F(2, 54) = 0.424$, $p > .1$, suggesting that female infants looked longer than male infants regardless of condition.

The main effect for Trial did not yield a significant effect, $F(1, 54) = 1.072$, $p > .1$. The Trial by Sex interaction, $F(1, 54) = 1.621$, $p > .1$, and the Trial by Condition

interaction, $F(2, 54) = 0.925$, $p > .1$, were not significant. Likewise the three-way interaction, Trial by Sex by Condition, was not significant, $F(2, 54) = 0.209$, $p > .1$,

Dishabituation Trial: Individual Conditions by Sex of Infants

Planned contrasts between means were carried out to address whether either male infants or female infants showed longer looking to the dishabituation trial than in the average of the last three habituation trials in each of the three conditions individually.

Control condition. For the control condition, when the dishabituation stimulus is the same as the habituation stimulus, it was expected that neither male nor female infants would show longer looking to the dishabituation trial compared to the average of the last three trials. Indeed, as shown in Figure 2, the planned contrasts confirmed that neither female infants, $F(1, 54) = 0.034$, $p > .1$, nor male infants, $F(1, 54) = 0.642$, $p > .1$, showed longer looking to the dishabituation trial than to the average of the last three habituation trials. The permutation test for paired replicates (for details see Siegel & Castellan, 1988) also indicates that neither female infants nor male infants showed renewed visual interest on the dishabituation trial, both ps $> .1$

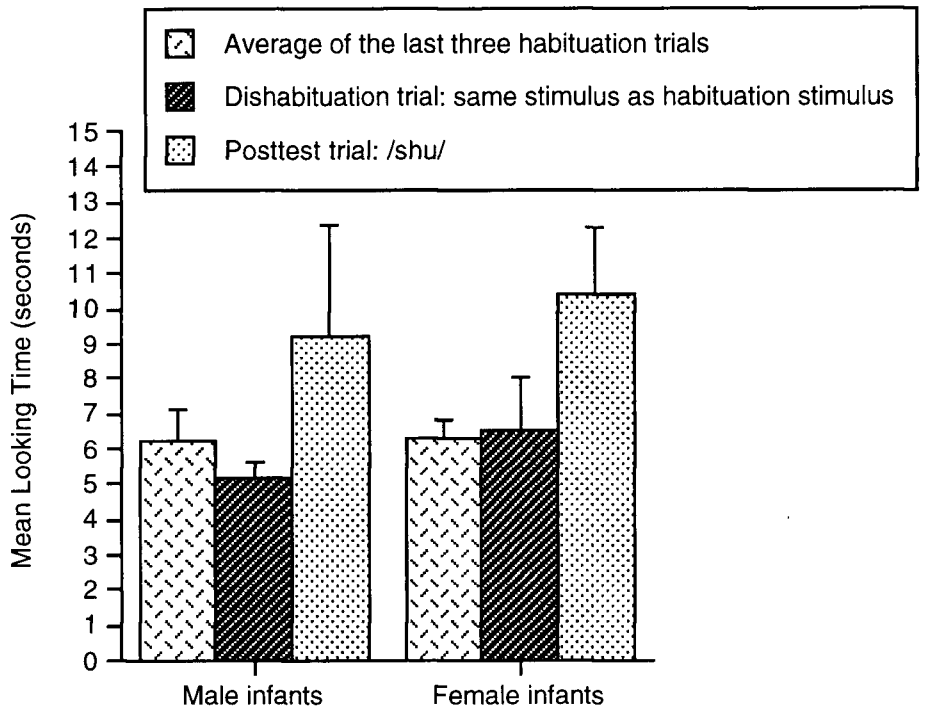

Figure 2. Mean Looking Time in the Control Condition as a Function of Trial and Sex (Experiment 1). Error bars represent SEM.

Based on this finding, it can be inferred that any instances of significantly greater looking to the dishabituation trial compared to the last three habituation trials in the experimental conditions are due to infants perceiving a change in the stimulus rather than due to spontaneous recovery of visual interest after habituation criteria are met.

Audiovisual /vi/. If infants are like adults, infants habituated to audiovisual /vi/ should not show renewed visual interest to the audio /bi/-visual /vi/ because it is perceived as /vi/. On the other hand, if infants' perception are only influenced by the auditory component, then the dishabituation stimulus of audio /bi/-visual /vi/ would be perceived as /bi/. In this case, infants habituated to audiovisual /vi/ should show renewed visual interest to the dishabituation stimulus.

For infants habituated to the audiovisual /vi/, neither female infants, $F(1, 54) = 0.548$, $p > .1$, nor male infants, $F(1, 54) = 0.108$, $p > .1$, showed longer looking to

the dishabituation trial compared to the average of the last three habituation trials, as shown in Figure 3. Permutation tests for paired replicates were also not significant for either male of female infants, both ps > .1. These results suggest that both groups of infants may have perceived the dishabituation stimulus, audio /bi/-visual /vi/, as /vi/.

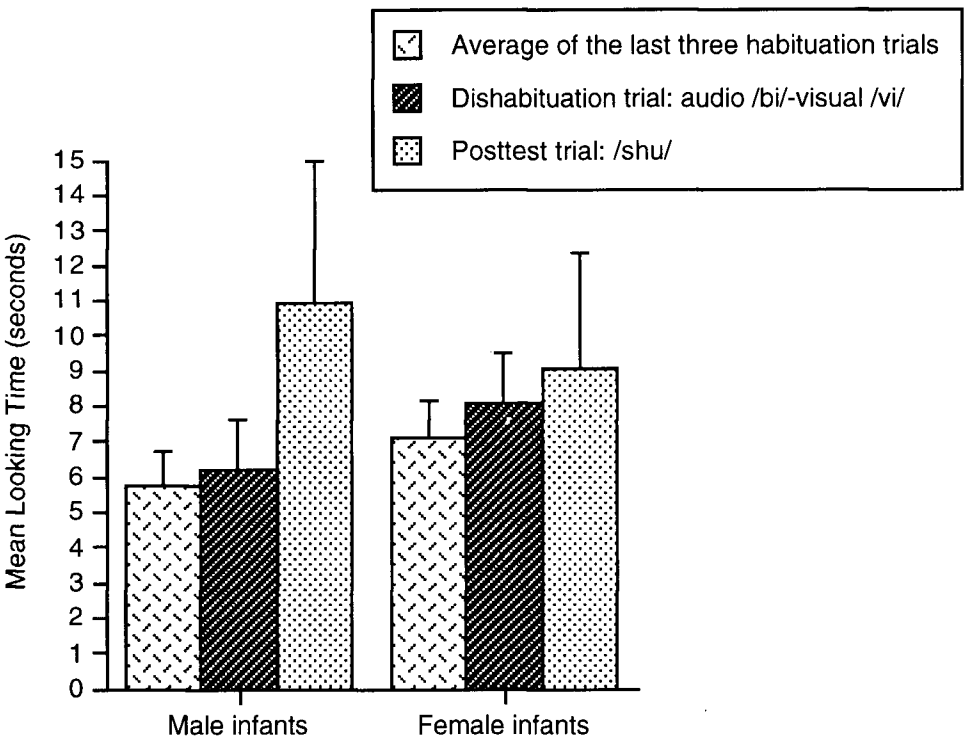

Figure 3. Mean Looking Time in the Audiovisual /vi/ Condition as a Function of Trial and Sex (Experiment 1). Error bars represent SEM.

Audiovisual /bi/. If infants, like adults, perceive the audio /bi/-visual /vi/ as /vi/, then infants habituated to audiovisual /bi/ should show renewed visual interest on the dishabituation trial. On the other hand, if infants attend only to the auditory component, infants habituated to the audiovisual /bi/ should not show renewed visual interest.

For infants habituated to the audiovisual /bi/, female infants looked significantly longer to the dishabituation stimulus, audio /bi/-visual /vi/, compared to

the average of the last three habituation trials, $F(1, 54) = 3.583$, $p < .05$, whereas male infants did not, $F(1, 54) = 0.034$, $p > .1$, as shown in Figure 4. Permutation tests for paired replicates also indicated that while female infants showed renewed visual interest, $p < .05$, male infants did not, $p > .1$. The results of this condition suggest that female infants, but not male infants, may have perceived the audio /bi/-visual /vi/, as /vi/–as do adults.

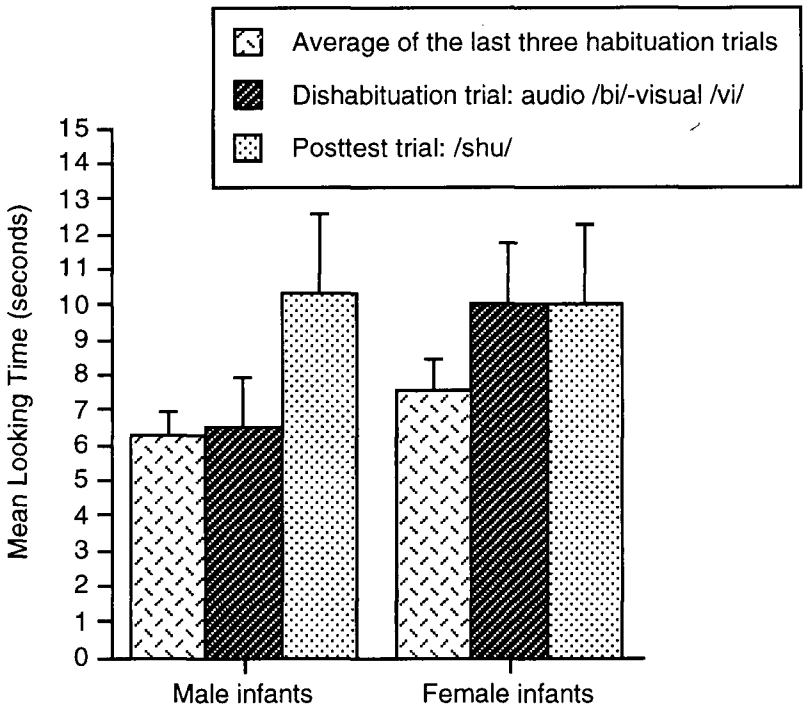

Figure 4. Mean Looking Time in the Audiovisual /bi/ Condition as a Function of Trial and Sex (Experiment 1). Error bars represent SEM.

Supplementary Analysis

Researchers interested in individual differences in infancy have found the average looking time per trial, measured during habituation studies, to be a reliable index of infants' information processing abilities; moreover, this measure can be used to predict performance during childhood on a variety of cognitive tasks (see Colombo & Mitchell, 1990, and McCall & Carriger, 1993, for reviews). Although the

present experiments were not designed with the prediction of performance as a goal, average looking time per trial was calculated to see if this variable could help explain the obtained sex difference.

The average looking time per habituation trial was calculated by summing each infant's looking time during habituation trials (i.e., excluding the pretest, dishabituation trials and the posttest) and dividing the total looking time by the number of habituation trials completed until habituation criterion was met. Although female infants and male infants did not differ in average looking time according to an unpaired t test, t (58) = -1.504, $p > .1$, a Mann-Whitney test, U = 321, $p < .06$, approached significance. The average look for female infants (M = 12.6 s, SD = 4.6 s) tended to be longer than for male infants (M = 10.8 s, SD = 5.1 s). Average looking time was not significantly correlated with age (days over 4 months), $r = .103$, $p > .1$.

Summary

The planned contrasts computed to determine whether male and female infants showed increased looking to the dishabituation stimulus compared to the average of the last three habituation trials suggest (1) infants do not show increased looking in the control condition, (2) both male and female infants habituated to /vi/ do not show increased looking to an audio /bi/-visual /vi/ and (3) female infants, but not male infants, habituated to /bi/ show a significant increase in looking time to an audio /bi/-visual /vi/. Taken together, these findings suggest that female infants may perceive the audio /bi/-visual /vi/ as /vi/—an adult-like pattern of results. The failure to find renewed visual interest to the dishabituation stimulus in some conditions was not due merely to loss of interest or boredom with the 'game' as infants showed renewed visual interest to the posttest stimulus regardless of sex or condition.

Although the female infants were not significantly older than male infants, they did look longer than male infants on the first habituation trial, and their average looking time per trial tended to be longer than for male infants (as indicated by the Mann-Whitney test which approached significance, $p < .06$). Although the general tendency is for shorter average looking times in visual habituation studies to be correlated with superior performance on other cognitive tasks (see Colombo & Mitchell, 1990; McCall & Carriger, 1993), a recent report suggests that this trend does not always hold.

Krinsky-McHale, Devenny and Bornstein (1997) found that short lookers at 5 months of age in a single visual exemplar habituation study had more advanced language skills at the age of 2 years than did toddlers who were long lookers at 5 months of age; however, when the habituation stimulus was complex (different facial expressions or different geometric patterns), long lookers at 5 months of age had more advanced language skills at the age of 2 years than did toddlers who were short lookers at 5 months of age. Krinsky-McHale et al. suggest that infants who spend too little time processing a complex stimulus at 5 months of age may be at a disadvantage later, as they have a tendency to ignore or fail to attend to all the relevant information.

The habituation stimuli used in the present experiments might well be considered to be more like the complex stimuli than like the single stimuli used by Krinsky-McHale et al. (1997). Although in the present study the habituation stimuli did not differ from trial to trial, the habituation stimuli were bimodal rather than unimodal. Thus, there were in fact *two* stimuli per trial: one auditory and one visual. In this respect, the habituation stimuli in this study are more complex than a single modality stimulus.

In the present study, the female infants showed a coherent pattern of results across the conditions which is consistent with the predicted adult-like pattern of

responding whereas the male infants did not show this pattern. Based on the average looking time data, the female infants appear to be like the longer lookers in the Krinsky-McHale et al. (1997) study. Because they spent more time processing the habituation stimulus they likely encoded both the auditory and the visual component of the stimulus.

Male infants, on the other hand, did not show a coherent pattern of responding across the conditions: they did not show either an adult-like or an auditory-only pattern of responding. One possibility is that some of the male infants encoded only the auditory syllable while other male infants encoded only the visual syllable. The encoding of different features would lead to two types of responding on the dishabituation trial which effectively cancel each other out. Given that male infants showed shorter looking on the first habituation trial and shorter average looking times during habituation than female infants, it is possible that they failed to encode *both* the auditory and the visual syllable.

In sum, female infants, but not male infants showed an adult-like pattern of responding while male infants did not. Although it is unlikely that the female infants habituated to audiovisual /bi/ may have showed renewed visual interest on the dishabituation trial simply because they noticed a change in the speaker's mouth movements, the pattern of results obtained for the female infants is also consistent with this explanation.

EXPERIMENT 2

To tease apart the two explanations for the finding that female infants habituated to /bi/ showed renewed visual interest to a dishabituation stimulus of audio /bi/-visual /vi/, infants in Experiment 2 were habituated to either an audiovisual /vi/ or an audiovisual /bi/—as in Experiment 1—and then presented with a dishabituation stimulus of an audio /vi/-visual /bi/. This incongruent audiovisual pairing typically results, for adults, in 'auditory capture'; adults perceive /vi/. Thus if only a change in the visual component of an incongruent audiovisual stimulus, independent of a change in perception of the speech sound, is necessary for infants to show renewed visual interest, infants habituated to an audiovisual /vi/ should show renewed visual interest to the dishabituation stimulus of audio /vi/-visual /bi/.

Method

Participants

Participation was solicited in the same manner as in the previous experiments. All infants were full term and, according to parent report, had had no more than one ear infection since birth and were healthy at the time of testing. The final sample consisted of 40 infants[19] (20 male, 20 female) ranging in age from 4 months 4 days to 5 months 1 day (M = 4 months 17 days, SD = 9 days). Male infants and female infants did not differ significantly in age, $t(38) = 1.250$, $p > .1$ (M = 4 months 18 days, SD = 8 days, and M = 4 months 17 days, SD = 9 days, for male

[19] Although the total number of infants was greater in Study 1 because there was a control condition in that study, the same number of male and female infants per test condition was used in both studies.

and female infants respectively). All infants came from homes in which English was spoken at least 90% of the time.

Infants were randomly assigned to two conditions (20 infants in each) with equal numbers of male and female infants in each condition. An additional 37 infants were omitted for the following reasons: cried (14), experimenter error (7), mother stopped the experiment (2), parent distracted infant (1), equipment failure (1), did not meet habituation criterion (8), outlier (4).

Stimuli

The stimuli used were the same as those used in the previous study with one exception: an incongruent audio /vi/-visual /bi/ was used instead of audio /bi/-visual /vi/. This stimulus was created in the same fashion as the audio /bi/-visual /vi/ used in the previous study. Pilot testing with eight adults indicated that for all of the adults, this stimulus was perceived as /vi/—auditory capture—98% of the time (see Appendix C).

Equipment

The same equipment was used as in the previous study.

Procedure

The procedure was identical to that used in the previous study.

Conditions

One group of infants was habituated to a congruent audiovisual /vi/. A second group of infants was habituated to a congruent audiovisual /bi/. Both groups of infants were presented with a dishabituation stimulus of an audio /vi/-visual /bi/.

Predictions

For adults, the dishabituation stimulus of audio /vi/-visual /bi/ reliably results in a unified percept of /vi/. That is, adults report hearing /vi/ and not /bi/–their percept is "captured" by the auditory information. If young infants' percepts are

similarly captured by the auditory information I would expect that (1) infants habituated to /bi/ would show renewed visual interest to the audio /vi/-visual /bi/, and (2) infants habituated to /vi/ would fail to show recovery of interest. And, if infants show renewed visual interest when the dishabituation stimulus represents a change in the speaker's mouth movements, independent of a change in percept of the speech, then (1) infants habituated to /vi/ should show renewed visual interest to the dishabituation stimulus of audio /vi/-visual /bi/, and (2) infants habituated to /bi/ should not show renewed interest to the audio /vi/-visual /bi/ dishabituation stimulus. These sets of predictions can be seen in Table 3.

Table 3

Patterns of Results (Experiment 2)

Stimuli		Coherent patterns	
Habituation	Dishabituation	Visual only	Integration of auditory & visual
audiovisual /vi/	audio /vi/-visual /bi/	+	–
audiovisual /bi/	audio /vi/-visual /bi/	–	+

Note: + denotes renewed visual interest on the dishabituation trial and – denotes no renewed visual interest.

Results and Discussion

Reliability Coding

A trained coder, blind to condition, measured infants' looking times on each trial for 10 randomly selected infants (25% of the sample) for a total of 106 trials.

The experimenter's and the coder's looking times were significantly correlated, r = .999, $p < .0001$, indicating that the experimenter's on-line coding was reliable.

Habituation

To confirm that infants did indeed show a decrement in looking across trials, looking time data for the criterion trials during habituation (the first three and last three habituation trials) were submitted to a mixed design analysis of variance with Sex (male, female) and Condition (audiovisual /vi/, audiovisual /bi/) as between-subjects factors and Habituation Trial (1, 2, 3, 4, 5, 6) as a repeated measure.

The main effects for Sex, $F(1, 36) = 0.115$, $p > .1$, and Condition, $F(1, 36) = 0.03$, $p > .1$, were not significant, and the Sex by Condition interaction, $F(1, 36) = 0.238$, $p > .1$, was also not significant. There was as expected a significant decrement in looking time across the trials as indicated by the main effect for Habituation Trial, $F(5, 180) = 18.515$, $p < .0001$. Planned orthogonal polynomial contrasts revealed that the decrement in looking had a significant linear component, $F(1, 180) = 85.517$, $p < .0001$, as well as a significant quadratic component, $F(1, 180) = 5.203$, $p < .025$.

No significant Habituation Trial by Sex interaction, $F(5, 180) = 0.074$, $p > .1$, was observed in this study as can be seen in Figure 5. The Habituation Trial by Condition, $F(5, 180) = 0.412$, $p > .1$, and Habituation Trial by Sex by Condition, $F(5, 180) = 0.582$, $p > .1$, interactions were also not significant.

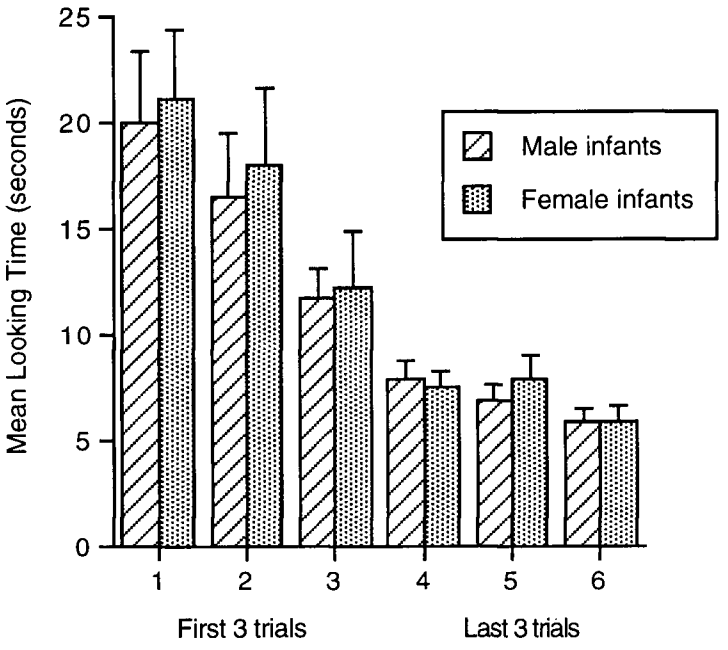

Figure 5. Mean Looking Time During Habituation on Criterion Trials (Experiment 2). Error bars represent SEM.

Trials to criterion. Infants on average required 7.2 trials (SD = 2 trials) to meet criterion. Female infants and male infants did not differ in the number of trials to criterion, t (38) = 0.392, $p > .1$, and, $U = 185$, $p > .1$. (M = 7.4 trials, SD = 1.9 and M = 7.1 trials, SD = 2.1, for male and female infants respectively).

Posttest trial

To assess whether infants would show renewed visual interest to the posttest trial after habituation had occurred, looking time data were submitted to a mixed design analysis of variance with Condition (audiovisual /vi/, audiovisual /bi/) and Sex (male, female) as between-subjects factors and Trial (average, posttest) as a repeated factor.

The main effects for Sex, $F(1, 36) = 0.575$, $p > .1$, and Condition, $F(1, 36) = 1.218$, $p > .1$, were not significant; likewise, the interaction between Sex and Condition, $F(1, 36) = 0.023$, $p > .1$, was not significant.

However, a significant main effect for Trial was obtained, $F(1, 36) = 15.714$, $p < .0003$, one tailed, indicating that infants looked longer to the posttest trial (M = 12.0 s, SD = 9.3 s) than to the average of the last three habituation trials (M = 7.0 s, SD = 2.5 s). The Trial by Sex, $F(1, 36) = 0.795$, $p > .1$, Trial by Condition, $F(1, 36) = 1.566$, $p > .1$, and Trial by Sex by Condition, $F(1, 36) = 0.037$, $p > .1$, interactions were all not significant.

In summary, this analysis confirms that regardless of condition or sex, infants overall showed renewed visual interest to the posttest stimulus of /shu/. Infants are still attending to the stimuli and 'playing the game' by the end of the testing session.

Dishabituation Trial: Overall Analysis

To assess whether infants would look longer to the dishabituation stimulus, audio /vi/-visual /bi/, following habituation to either an audiovisual /vi/ or an audiovisual /bi/, looking time data were submitted to a mixed design analysis of variance with Sex (male, female) and Condition (audiovisual /vi/, audiovisual /bi/) as between-subjects factors and Trial (average, dishabituation) as repeated measure.

The main effects for Sex, $F(1, 36) = 0.926$, $p > .1$, and Condition, $F(1, 36) = 0.459$, $p > .1$, and the Sex by Condition interaction, $F(1, 36) = 0.669$, $p > .1$, were not significant. The main effect for Trial, $F(1, 36) = 0.324$, $p > .1$ was also not significant. The Trial by Sex, $F(1, 36) = 1.467$, $p > .1$, and Trial by Condition, $F(1, 36) = 0.256$, $p > .1$, interactions were not significant. Likewise, the Trial by Sex by Condition interaction was also not significant, $F(1, 36) = 0.087$, $p > .1$.

Dishabituation Trial: Individual Conditions by Sex of Infants

Planned contrasts were carried out to determine whether either male or female infants looked longer to the dishabituation trial than to the average of the last three dishabituation trials in each condition separately.

Audiovisual /bi/. If infants, like adults, perceive the incongruent audio /vi/-visual /bi/, as /vi/, then infants habituated to an audiovisual /bi/ should look longer to the dishabituation stimulus than to the average of the last three habituation trials. As shown in Figure 6, contrasts revealed that neither female infants, $F(1, 36) = 0.045$, $p > .1$, nor male infants, $F(1, 36) = 1.662$, $p > .1$, looked longer to the dishabituation stimulus than to the average of the last three habituation trials. Permutation tests for paired replicates also indicated that neither male infants nor female infants showed renewed visual interest, both ps $> .1$. This pattern of results suggests that infants may not have perceived the audio /vi/-visual /bi/, as /vi/.

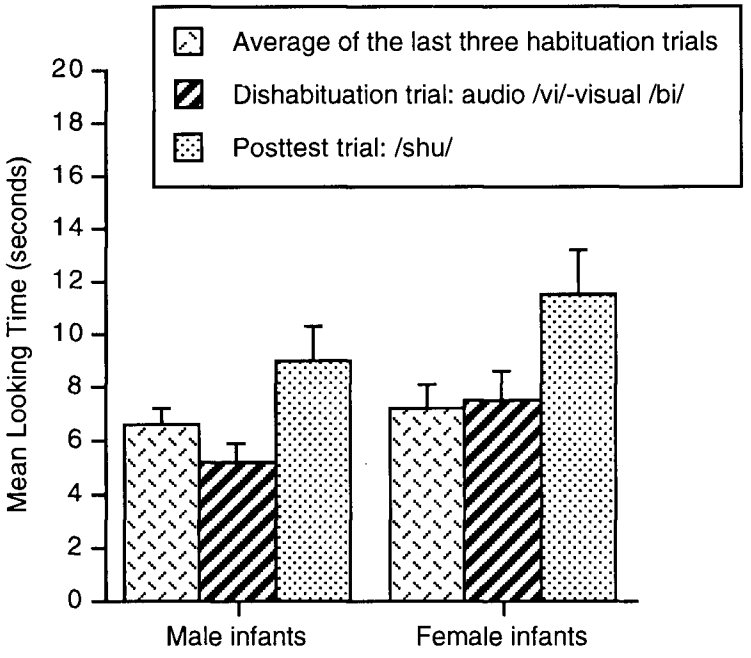

Figure 6. Mean Looking Time in the Audiovisual /bi/ Condition as a Function of Trial and Sex (Experiment 2). Error bars represent SEM.

Audiovisual /vi/. If infants will look longer to a dishabituation stimulus that represents a change in the visual component, regardless of the perceived speech sound, then infants habituated to an audiovisual /vi/ should look longer to the dishabituation stimulus than to the average of the last three habituation trials. As

shown in Figure 7, contrasts revealed that neither male infants, $F(1, 36) = 0.241$, $p > .1$, nor female infants, $F(1, 36) = 0.187$, $p > .1$, looked longer to the dishabituation stimulus than to the average of the last three habituation trials. Permutation tests for paired replicates also indicated that neither male infants nor female infants showed renewed visual interest, both $ps > .1$. The pattern of results obtained in this condition suggests that infants will not look longer to a dishabituation stimulus if it represents only a change in the visual component.

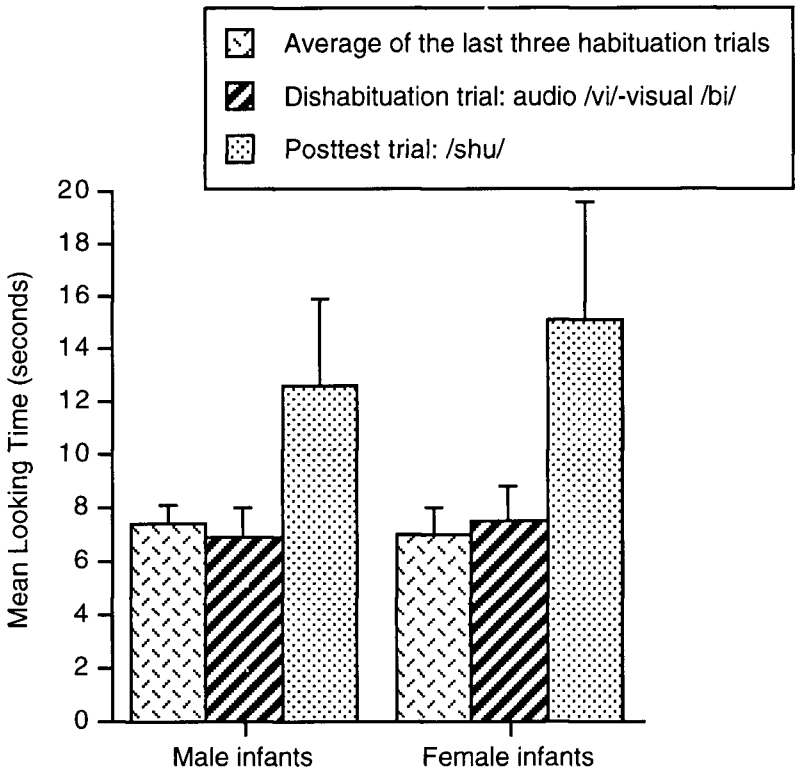

Figure 7. Mean Looking Time in the Audiovisual /vi/ Condition as a Function of Trial and Sex (Experiment 2). Error bars represent SEM.

Supplementary Analysis

As in Experiment 1, average looking time per trial during habituation was calculated to determine if male and female infants differed on this measure. In this experiment, male and female infants did not differ in the average looking time per

trial, $t(38) = -0.266$, $p > .1$, and, $U = 197$, $p > .1$ ($M = 12.2$ s, $SD = 4.9$ s, and $M = 12.6$ s, $SD = 6.2$ s, for male and female infants respectively). As in Experiment 1, average looking time per trial was not correlated significantly with age (days over 4 months), $r = -203$, $p > .1$.

Summary

The pattern of results obtained from the analysis of the posttest trial indicates that infants, regardless of condition or sex, showed renewed visual interest to the posttest trial. Based on this result, it is likely that the failure to show renewed visual interest to the dishabituation trial is due to infants' perception of that stimulus and not due to boredom or disinterest in the game itself. Although the failure to observe renewed visual interest on the dishabituation trial in both conditions is difficult to explain if this study is considered on its own, it is possible to interpret these findings in light of the positive findings of the previous study.

The results of the dishabituation analysis suggest that infants may not have perceived the dishabituation stimulus, audio /vi/-visual /bi/, as /vi/, as do adults. If it was the case that infants, unlike adults, showed visual capture, thus perceiving the audio /vi/-visual /bi/ as /bi/ rather than as /vi/, I would have expected infants habituated to an audiovisual /vi/ to show renewed visual interest on the dishabituation trial—and they did not. Thus it is not clear from this study how infants perceived this incongruent audiovisual syllable.

What does seem to safe to conclude is that infants were not attending solely to the auditory component as might be expected given that preschool children show a greater reliance on the auditory than on the visual component of audiovisual syllables (e.g., Desjardins et al., in press). If they were attending exclusively to the auditory component and ignoring the speaker's visible articulation, infants habituated to the audiovisual /bi/ would have show renewed

visual interest to the audio /vi/-visual /bi/ as it represents a change in the speech sound delivered through the auditory channel.

What is also clear from this study is that a change in the speaker's mouth movements—regardless of whether a change in speech sound is perceived—is not sufficient for either male or female infants to show renewed visual interest on the dishabituation trial. This finding lends some credence to the explanation for the results of the previous study: that is, infants showed renewed visual interest to audio /bi/-visual /vi/ following habituation to audiovisual /bi/ because it represented a change in perceived speech sound rather than a change in the visual stimulus alone.

Studies 1 and 2 together suggest that at least female infants' perception of speech can be influenced by both the auditory and the visual components of speech. What is not apparent from these two experiments is whether infants, particularly female infants, *consistently* integrate consonants across heard and seen speech. In fact, the findings of Experiment 2 suggest that they may not always integrate auditory and visible speech; in this study female infants did not appear to perceive audio /vi/-visual /bi/ as /vi/, unlike adults.

EXPERIMENT 3

This experiment was designed to address further whether the integration of heard and seen speech is mandatory. Do infants consistently perceive the incongruent audio /bi/-visual /vi/ as /vi/ as do adults? To answer this question, infants were habituated to an auditory /bi/-visual /vi/ which typically results in visual capture for adults; adults perceive this stimulus as /vi/. Then infants were tested on two novel syllables: an audiovisual /bi/ and an audiovisual /vi/. If infants' perception of the heard syllable is consistently influenced by watching a speaker's lip and mouth movements, they should look longer to the audiovisual /bi/ than to the audiovisual /vi/. This pattern of results would suggest that the integration of the auditory and visual signals is mandatory, thus, an audio /bi/-visual /vi/ is reliably perceived as /vi/—as it is for adults.

Two other patterns of results are also possible. If infants attend solely to the auditory component of the habituation stimulus, in this case /bi/, then infants should look longer on the test trials to the audiovisual /vi/ than to the audiovisual /bi/. If infants fail to show a preference between the audiovisual /vi/ and the audiovisual /bi/, this would indicate that infants may perceive the audio /bi/-visual /vi/ as /bi/ on some trials and as /vi/ on other trials. Both these patterns would be consistent with integration not being a mandatory process for young infants.

Method

Participants

The participation of parents and their infants was solicited in the same manner as in the two previous experiments. All infants were full term and, according to parent report, had had no more than one ear infection since birth and

were healthy at the time of testing. All infants came from homes in which English was spoken at least 90% of the time. The final sample consisted of 16 infants[20] (8 male, 8 female) ranging in age from 4 months 2 days to 5 months 2 days (M = 4 months 14 days, SD = 9 days). Male and female infants did not differ significantly in age, $t(14) = -1.059$, $p > .1$ (M = 4 months 12 days, SD = 6 days, and M = 4 months 17 days, SD = 12 days, for male and female infants respectively). Infants were randomly assigned to one of two orders with equal numbers of male and female infants in each. An additional 25 infants were excluded for the following reasons: cried (8), refused to participate (3), experimenter error (6), did not meet habituation criterion (6), outlier (2).

Stimuli

The stimuli used were identical to those used in Experiment 1.

Equipment

The same equipment was used as in the previous experiments.

Procedure

The procedure was the same as that used in the previous experiments with one exception: two dishabituation trials were presented instead of one.

Conditions

All infants were habituated to the incongruent audio /bi/-visual /vi/. When habituation criteria were met, infants were presented with a congruent audiovisual /vi/ and a congruent audiovisual /bi/ on the next two trials. Half the infants were presented with audiovisual /vi/ followed by audiovisual /bi/ while the other half were presented with the stimuli in the reverse order.

---

[20] A smaller number of infants per order was used in this study as both the test stimuli are presented to each infant rather than to different infants.

Predictions

For adults, the audio /bi/-visual /vi/ reliably results in a unified percept of /vi/. That is, adults report hearing /vi/ and not /bi/—their percept is "captured" by the visual information of the lips forming a /vi/ sound. If young infants' percepts are similarly captured by the visual information on the majority of trials, I would expect that having been habituated to the audio /bi/-visual /vi/, infants would look longer to the audiovisual /bi/ than to the audiovisual /vi/. If infants of 4 months of age do not systematically integrate auditory and visual information, I would not expect to see a strong preference for one test syllable over the other. If infants are influenced only by the auditory information, I would expect infants to look longer to the audiovisual /vi/ than to the audiovisual /bi/. These sets of predictions can be seen in Table 4.

Table 4

Patterns of Results (Experiment 3).

Auditory only	Integration of auditory and visual	Mixed: Integration and auditory only
greater looking to /vi/ than to /bi/	greater looking to /bi/ than to /vi/	equal looking to /bi/ and /vi/

Results & Discussion

Reliability Coding

A trained coder, blind to condition, measured infants' looking times on each trial for 10 randomly selected infants (25% of the sample) for a total of 106 trials. The experimenter's and the coder's looking times were significantly correlated, $r = .999$, $p < .0001$, indicating that the experimenter's on-line coding was reliable.

Habituation

To confirm that infants did indeed show a decrement in looking across trials, looking time data for the criterion trials during habituation (the first three and last three habituation trials) were submitted to a mixed design analysis of variance with Sex (male, female) and Order of dishabituation trial (audiovisual /bi/ first, audiovisual /vi/ first) as between-subjects factors and Habituation Trial (1, 2, 3, 4, 5, 6) as a repeated measure.

The main effects for Sex, $F(1, 12) = 2.421$, $p > .1$, and Condition, $F(1, 12) = 0.899$, $p > .1$, were not significant, and the Sex by Condition interaction, $F(1, 12) = 0.429$, $p > .1$, was also not significant. There was, as expected, a significant decrement in looking time across the trials as indicated by the significant main effect for Habituation Trial, $F(5, 60) = 5.373$, $p < .0005$. Planned orthogonal polynomial contrasts revealed that the decrement in looking had a significant linear component, $F(1, 60) = 23.204$, $p < .0001$, but not a significant quadratic component, $F(1, 60) = 0.766$, $p > .1$—unlike in the two previous experiments.

No significant Habituation Trial by Sex interaction, $F(5, 60) = 0.359$, $p > .1$, was observed in this study. The Habituation Trial by Order, $F(5, 60) = 0.787$, $p > .1$, and Habituation Trial by Sex by Order, $F(5, 60) = 0.272$, $p > .1$, interactions were also not significant.

As can be seen in Figure 8, although it appears as if female infants tended to look longer than male infants on the first three habituation trials; these differences were not significant, Tukey's HSD, $ps > .05$. Mann-Whitney tests confirmed that the female infants did not look longer than the male infants on the first trial, $U = 30$, $p > .1$, or on the second trial, $U = 22$, $p > .1$; however, the difference between males and females approached significance on the third trial, $U = 15$, $p < .08$.

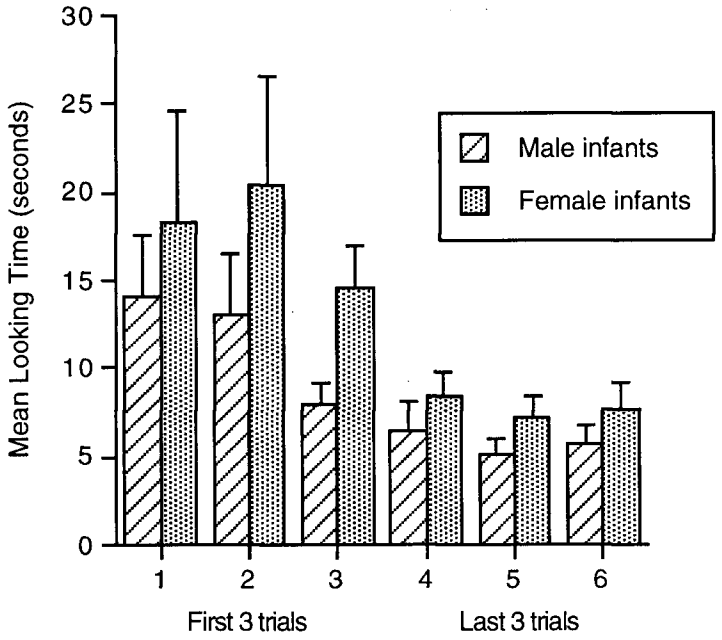

Figure 8. Mean Looking Time During Habituation on Criterion Trials (Experiment 3). Error bars represent SEM.

Trials to criterion. On average, infants required 7.3 trials to meet habituation criterion. Trials to criterion did not differ significantly for male and female infants, $t(14) = 0.816$, $p > .1$, and $U = 24.5$, $p > .1$ ($M = 7.75$ trials, $SD = 2.6$ and $M = 6.9$ trials, $SD = 1.6$ for male and female infants respectively).

Posttest trial

To assess whether infants would show renewed visual interest to the posttest trial after habituation had occurred, looking time data were submitted to a mixed design analysis of variance with Sex (male, female) and Order of dishabituation trial (audiovisual /bi/ first, audiovisual /vi/ first) as between-subjects factors and Trial (average, posttest) as a repeated measure.

The main effects for Sex, $F(1, 12) = 0.889$, $p > .1$, and Order, $F(1, 12) = 0.327$, $p > .1$, were not significant; similarly, the interaction was not significant, $F(1, 12) = 1.536$, $p > .1$. There was a significant main effect for Trial, $F(1, 12) = 3.911$, $p < .04$, one tailed, indicating that infants did look longer to the posttest trial ($M = 10.9$

s, SD = 8.6 s) than to the average of the last three habituation trials (M = 6.7 s, SD = 3.0 s). The Trial by Sex, F(1, 12) = 0.051, p > .1, Trial by Order, F(1, 12) = 0.235, p > .1, and Trial by Sex by Order, F(1, 12) = 1.000, p > .1, interactions were all not significant.

In summary, this pattern of results indicates that regardless of condition or sex, infants showed renewed visual interest to the posttest stimulus of /shu/ after habituation criteria had been met. Infants will show renewed interest to a stimulus perceived as different from the habituation stimulus and by the end of the session, infants are still attending to the stimuli.

Dishabituation Trials: Overall Analysis

To determine whether infants looked longer to one of the dishabituation trials than to the other, looking time data on the two dishabituation trials were submitted to a mixed design analysis of variance with Sex (male, female) and Order of dishabituation trial (audiovisual /bi/ first, audiovisual /vi/ first) as between-subjects factors and Trial (audiovisual /bi/, audiovisual /vi/) as a repeated measure.

The main effects for Sex, F(1, 12) = 1.236, p > .1, and Order, F(1, 12) = 0.002, p > .1, were not significant. Likewise the Sex by Order interaction was not significant, F(1, 12) = 0.041, p > .1. The main effect for Trial, F(1, 12) = 2.062, p > .1 was not significant; however a significant Trial by Sex interaction was observed F(1, 12) = 8.410, p < .02, indicating that male infants looked relatively longer to the audiovisual /bi/ than to the audiovisual /vi/ while the reverse was true for female infants. This interaction can be seen in Figure 9. The Trial by Order, F(1, 12) = 0.711, p > .1, and Trial by Sex by Order, F(1, 12) = 0.279, p > .1, interactions were not significant.

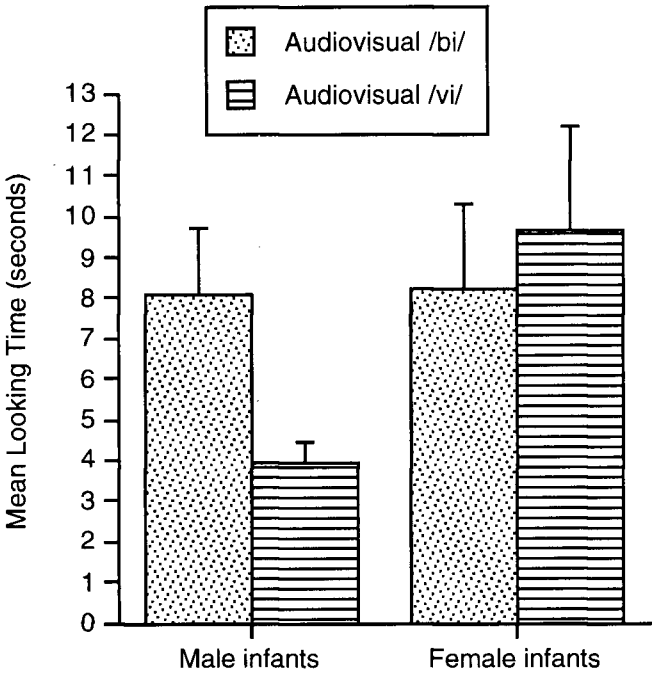

Figure 9. Mean Looking Time on Dishabituation Trials as a Function of Sex (Experiment 3). Error bars represent SEM.

Dishabituation Trials: Sex Differences

Contrast analyses were carried out to determine whether male infants looked significantly longer to one of the two stimuli and similarly whether female infants looked longer to one of the two stimuli. As shown in Figure 9, for the male infants, a significant difference was obtained, $F(1, 12) = 9.395$, $p < .025$, indicating that male infants looked longer to the audiovisual /bi/ than to the audiovisual /vi/. The permutation test for paired replicates also indicated that male infants looked significantly longer to audiovisual /bi/ than to audiovisual /vi/, $p < .04$.

For the female infants, a significant difference was not obtained with parametric testing, $F(1, 12) = 1.064$, $p > .1$. However, the permutation test for paired replicates was significant, $p < .02$, indicating that female infants longer to the audiovisual /vi/ than to the audiovisual /bi/. Seven out of 8 female infants showed this pattern while one female infant looked equally long to both stimuli.

Supplementary Analysis

As in Experiment 1, in an effort to explain the obtained sex difference, average looking time per habituation trial was calculated. Although the average looking time for female infants (M = 13.2 s, SD = 5.3 s) appears to be greater than for male infants (M = 9.2 s, SD = 4.7 s), this difference did not even approach significance, $t(14) = -1.582$, $p > .1$ and, $U = 18$, $p > .1$. As in the previous experiments average looking time was not significantly correlated with age (days over 4 months), $r = .-.036$, $p > .1$. Unlike in Experiment 1, the average looking time measure does not help explain why it is that the male infants and not female infants showed the predicted adult-like pattern of results.

Summary

The pattern of results obtained with male infants of longer looking to the audiovisual /bi/ than to the audiovisual /vi/ suggests that male infants perceived the habituation stimulus, audio /bi/-visual /vi/, as /vi/. In both Studies 1 and 2 male infants did not show renewed visual interest when the lip movements of the speaker changed. Thus, it is likely that the male infants in this study did not simply look longer at the audiovisual /bi/ than at the audiovisual /vi/ because the visible of the articulation of the audiovisual /bi/ was novel. Male infants looked longer at the audiovisual /bi/ than at the audiovisual /vi/ because they perceived the audio /bi/-visual /vi/ as /vi/.

The basic finding with the female infants is that there is a small but consistent preference for the audiovisual /vi/ over the audiovisual /bi/ with 7 of 8 showing a preference, and one showing no preference. The fact that female infants do not show a strong preference in terms of the *magnitude* of difference between looking to audiovisual /vi/ compared to audiovisual /bi/ is consistent with the hypothesis that female infants do not integrate all the instances of the incongruent stimulus.

If female infants had not integrated the auditory and visible syllables at all, I would have expected an auditory only pattern reflected in much longer looking to audiovisual /vi/ than to audiovisual /bi/. The fact that there was only a very small difference in looking time to the two stimuli suggests that infants likely perceived the audio /bi/-visual /vi/ as /bi/, but on some occasions as /vi/.

GENERAL DISCUSSION

The Present Research

Previous research has demonstrated that adults, and to a lesser extent, young children are influenced by both the speech sound they hear and the mouth movements they observe. The current research was designed with the purpose of determining whether young infants' perception of speech is similarly influenced by watching a speaker's mouth movements. To test this question, three experiments were designed in which infants of 4 months of age were habituated to a woman's face and voice repeating a syllable and then tested with the same face and voice producing a different syllable which represented either a change in the auditory component or a change in the visual component.

In Experiment 1, infants were habituated to either an audiovisual /bi/ or an audiovisual /vi/ and then presented with an audio /bi/-visual /vi/ (perceived as /vi/ by adults) in the experimental conditions, or the habituation stimulus presented again in the control condition. Both male and female infants in the control condition, as expected, did not show renewed visual interest on the dishabituation trial. Male infants habituated to the audiovisual /bi/ failed to show renewed interest to the audio /bi/-visual /vi/ and male infants habituated to audiovisual /vi/ also failed to show renewed visual interest to the audio /bi/-visual /vi/. This pattern of responding is not consistent with either attention to only the auditory component or with integrating the auditory and visual components. One possible explanation for this pattern of results is that some of the male infants may have been attending to only the auditory, some to only the visual component, while others may have been attending to both. Such a combination would result in no clear pattern over all.

Female infants, on the other hand, did show a consistent pattern of results. Female infants habituated to audiovisual /bi/ showed renewed visual interest to audio /bi/-visual /vi/ while female infants habituated to audiovisual /vi/ did not. This pattern is consistent with two possible explanations: either the female infants were integrating the auditory and visible speech or they were attending to only to the visual component and hence showed renewed visual interest to the changing lip movements of the speaker.

Experiment 2 was designed to rule out the latter explanation. Both male and female infants habituated to an audiovisual /vi/ failed to show renewed visual interest on the dishabituation trial to an audio /vi/-visual /bi/. This dishabituation stimulus is perceived as /vi/ and represents a change in the lip-movements but not a change in the syllable 'heard' by adults. By failing to show renewed visual interest, infants were indicating that a change in the visual stimulus is not sufficient to elicit renewed visual interest. Thus, I feel confident that female infants in Experiment 1 habituated to audiovisual /bi/ showed renewed visual interest to audio /bi/-visual /vi/ because they perceived it as /vi/ and not because they were attending solely to the visible speech.

Contrary to expectation, in Experiment 2, neither male nor female infants habituated to audiovisual /bi/ showed renewed visual interest to audio /vi/-visual /bi/. If infants had integrated the auditory and visible syllables they should have heard this mismatched audiovisual stimulus as /vi/ as do adults. Even if infants were simply attending to only the auditory component—as might be expected given that preschool children attend more to the auditory component than do adults (e.g., Desjardins et al., in press)—they should have shown renewed visual interest in this condition. Thus, it is unclear how infants 'heard' the audio /vi/-visual /bi/.

As a reminder, it cannot be argued that failure to show renewed visual interest on the dishabituation trial was due to infants having fallen asleep or lost

interest in the game. In both these experiments (and in Experiment 3), infants showed renewed visual interest on the posttest trial to an audiovisual /shu/ which differs from the habituation stimuli in both consonant and vowel. Thus, any failures to find renewed visual interest to the dishabituation stimuli are likely due to infants' as a group not perceiving the dishabituation stimulus as novel.

In an attempt to explain the obtained sex difference, average looking time per trial during habituation was calculated. Analyses of looking time data showed differences between female and male infants in Experiment 1 only. Female infants looked longer on the first habituation trial and the average looking time per trial tended to be greater compared to male infants. Recall that longer looking time per trial during habituation to a complex stimulus at 5 months of age is associated with more advanced language skills at 2 years of age (Krinsky-McHale et al., 1997).

One possible explanation for the finding that female infants show integration in Experiment 1 is that the female infants may be more advanced in their language development than male infants. During toddlerhood, female children tend to be more advanced in their language skills than male children (e.g., Huttenlocher, Haight, Bryk, Seltzer & Lyons,1991; Morisset & Barnard, 1995).

In Experiment 3, I further addressed the possibility that although some infants can integrate auditory and visible speech that integration is not mandatory for young infants. To test this hypothesis, male and female infants were habituated to audio /bi/-visual /vi/, and then presented with two novel dishabituation stimuli, audiovisual /vi/ and audiovisual /bi/. Male infants showed a strong preference for the audiovisual /bi/ over the audiovisual /vi/ while female infants showed only a weak preference for the audiovisual /vi/ over the audiovisual /bi/.

The finding for male infants of a strong preference for audiovisual /bi/ over audiovisual /vi/ is consistent with infants' perceiving the habituation stimulus, audio /bi/-visual /vi/ as /vi/. This result is surprising when seen in context of the previous

two experiments in which male infants did not show integration of audio /bi/-visual /vi/ (Experiment 1) and did not show renewed visual interest to a change in lip movements (Experiment 2).

The results of Experiment 3 suggest that female infants do not always integrate the auditory and visual components. When habituated to an audio /bi/-visual /vi/, female infants did not show a strong preference for the audiovisual /bi/ over the audiovisual /vi/—as would have been expected had infants systematically integrated the habituation stimulus. This suggests that the female infants did not consistently perceive the audio /bi/-visual /vi/ as /vi/. In fact, female infants tended to look longer at the auditory /vi/ than the auditory /bi/, suggesting that they more often heard the incongruent habituation stimulus as /bi/. That the magnitude of this preference for audiovisual /vi/ over audiovisual /bi/ was very small (1.4 s difference) suggests that at least on some trials, female infants may have heard the audio /bi/-visual /vi/ as /vi/.

Analysis of looking time per trial during habituation did not yield any sex differences in this experiment unlike in Experiment 1. Thus, measures of looking time do not clearly differentiate across experiments which group of infants will show integration and which will not.

Why only the female infants in Experiment 1 and only the male infants in Experiment 3 showed clear integration of auditory and visible speech is not obvious. However, Experiments 1 and 3 did differ in design: in Experiment 1 infants were habituated to the matched audiovisual syllable whereas in Experiment 3 infants were habituated to the mismatched audiovisual syllable. Tighe and Powlison (1978) in their meta analysis of infant visual habituation research find that experiments showing sex differences in opposite directions sometimes differ procedurally in systematic ways (e.g., intertrial interval, trial length, number of habituation trials). Although the variables they cite do not differ between the

current experiments, the nature of the habituation stimuli do (matched vs. mismatched).

Taken together, these three experiments provide evidence that both male and female infants are able to integrate auditory and visible speech, but that they do not do so all the time: Integration of the auditory and visible components is not mandatory for young infants. Female infants showed integration of auditory and visible speech in Experiment 1, but not in Experiment 3; for male infants the reverse is true.

Relationship to Previous Research

The finding that at least some infants of 4 months of age can integrate auditory and visible speech concurs with the conclusions put forward by Burnham and Dodd (in press) and Rosenblum et al. (1997). Unfortunately, in neither of those papers did the authors report on possible sex differences. Recall, that Burnham and Dodd did not use either renewed visual interest or differential looking to the dishabituation stimuli, but rather two derived measures. Thus, their results are not directly comparable to the present study. Nevertheless, Burnham and Dodd found that infants familiarized to an audio /ba/-visual /ga/ have relatively higher derived integration scores and lower derived auditory scores than do infants familiarized to an audiovisual /ba/. On this basis, Burnham and Dodd argue that integration of auditory and visual syllables can occur. The authors do not show whether infants perceive an audio /ba/-visual /ga/ as /da/ or as /ða/. Thus, the reader is left with a certain amount of ambiguity about what speech sound the infants heard. In the present series of experiments it is suggested that female infants in Experiment 1 and male infants in Experiment 3 perceive an audio /bi/-visual /vi/, as /vi/.

Some evidence that infants are able to integrate auditory and visible speech is presented by Rosenblum et al. (1997). Their main finding is that infants familiarized to an audiovisual /va/ do *not* show renewed visual interest to an audio /ba/-visual /va/ (perceived as /va/ by adults) while they do show renewed visual interest to an audio /da/-visual /va/ (perceived as /da/ by adults). The present research differs from the that study in terms of the question motivating the research. By presenting only the lower half of the face and synthesized speech, Rosenblum et al. address whether it is possible to force infants to integrate auditory and visible speech. In fact, the authors indicate that pilot testing with a full face lead infants to be distracted hence they chose to use only the lower portion of the speaker's face. By using a full face and real speech, the present experiments address the question of whether, under more ecologically valid conditions, infants typically will integrate auditory and visible speech.

Moreover, the present research goes beyond addressing the question of whether infants can integrate heard and seen speech, to address the question of whether the integration of auditory and visible speech is mandatory for infants. And the findings of Experiments 1 and 3 suggest that it is not mandatory; only some infants show the effect. Interestingly, there are hints in the Burnham and Dodd (in press) report that integration may not be mandatory. They found that infants familiarized to audio /ba/-visual /ga/ looked longer on the first two familiarization trials than did infants who were familiarized to audiovisual /ba/. It is possible that infants looked longer in the audio /ba/-visual /ga/ condition because they initially noticed the discrepancy between the auditory and visual stimuli.

Hence the finding in the present experiments that infants do not always integrate the auditory and visual stimuli is indirectly supported by the findings of Burnham and Dodd (in press). Although it is unclear what the relationship is between matching of auditory and visual stimuli and the integration of auditory and

visual stimuli, the finding with slightly older infants (5-6 months of age) that matching of heard and seen consonants only occurs for some stimuli and only in the right visual field (McKain et al., 1983) suggests that young infants' representation of visible speech may not be sufficiently detailed to support mandatory integration.

The present research forms a novel contribution to the literature on the development of audiovisual speech perception. These experiments do not suffer from the methodological limitations of the previous attempts (Burnham & Dodd, in press; Rosenblum et al., 1997) to show integration of auditory and visible speech in infants. The present research provides a novel contribution by showing that although young infants can integrate auditory and visible speech, integration is not mandatory in young infants. These studies, then, suggest that the mechanism which supports the integration of auditory and visible speech may be due to an innate link between perception and production as specified in the revised motor theory of speech perception (Liberman & Mattingly, 1985). but experience is required for the mechanism to be more fully elaborated.

Future Directions

It is quite possible that experience producing consonants may be necessary before integration becomes mandatory. Desjardins et al. (in press) found that in a preschooler sample, producing consonants correctly was related to the perception of visible speech. Preschoolers who made production errors on consonants were poorer at lip-reading consonants in a visual only condition and showed less visual capture in an audiovisual speech perception task than did preschoolers who did not make production errors. Desjardins et al. suggest that producing consonants correctly may serve to enhance or develop more fully the underlying representation

of the visible articulation. It is conceivable, then, that the integration of auditory and visible speech becomes more mandatory over the course of development as infants and children gain experience producing consonants correctly.

Certainly this hypothesis is consistent with one put forward by Meltzoff and Kuhl (1994) regarding vowel perception. They suggest that producing vowel sounds helps to develop an auditory-articulatory intermodal map of speech. It is this map which allows infants, in a matching task, to look at the correct articulation of a heard vowel. Meltzoff and Kuhl do not address whether this intermodal map also supports the integration of heard and seen speech.

The hypothesis that production and perception might be intimately linked provides an avenue for further research. It would be interesting to see whether the number of consonants produced correctly is correlated with the extent to which young children are influenced consistently by the visible syllable in an audiovisual speech perception task. It would be of also be of interest to compare young children who are able to produce both /v/ and /b/, for example, with a group of agemates who cannot yet produce /v/. Does the ability to produce /v/ and /b/ influence the extent to which young children are influenced by a visible /bi/ or a visible /vi/?

Another avenue for future research is to explore the relationship between integration and matching. Can both male and female infants of 4 months of age detect the visible articulation which corresponds to a heard /bi/ or /vi/? Do infants who correctly match the heard syllable with the visible articulation show visual capture in an audiovisual speech perception task (such as Experiment 1)? Does the ability to match precede the ability to integrate developmentally? Moreover, how does the production of consonants relate to both the ability to match and the ability to integrate? What is the relationship between production, matching and integration?

Conclusions

The experiments presented here show that young infants between 4 and 5 months of age are able to integrate an incongruent audiovisual stimulus. Although infants are *able* to integrate heard and seen speech, they do not do so all the time—integration is not mandatory for young infants. These young infants have had limited experience hearing speech produced and observing other's speak, nevertheless, they are able to integrate auditory and visible speech. These findings then provide some support for the existence of an innate mechanism which makes it *possible* for infants to integrate heard and seen speech (Liberman & Mattingly, 1985).

However, that fact that infants do not always integrate heard and seen speech indicates that there is a role for experience. Although it is not clear from the present research exactly what kind(s) of experience are necessary, I put forward the hypothesis that the infant/child's own experience producing speech may serve to enhance the underlying representation of visible speech, which in turn may cause the integration of auditory and visible speech to become mandatory as development proceeds. Although future experiments are required to provide evidence for this hypothesis, the present series of experiments is the first step forwards investigating the role of experience in the development of audiovisual speech perception.

REFERENCES

Atkinson, J. (1984). Human visual development over the first 6 months of life: A review and a hypothesis. Human Neurobiology, 3, 61-74.

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

Bertelson, P., Vroomen, J., Wiegeraad, G., & de Gelder, B. (1994). Exploring the relation between McGurk interference and ventriloquism. Proceedings of 1994 International Conference on Spoken Language Processing (ICLP94), 2, 559-562.

Binnie, C.A., Montgomery, A.A. & Jackson, P.A. (1974). Auditory and visual contributions to the perception of consonants. Journal of Speech and Hearing Research, 17, 619-630.

Bower, T.G.R. (1974). Development in Infancy. San Francisco: Freeman.

Burnham, S. & Dodd, B. (in press). 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.

Campbell, R. (1992). The neuropsychology of lip-reading. In V. Bruce, A. Cowey, A.W. Ellis & D.I. Perrett (Eds.), Processing the Facial Image. Oxford: Clarendon Press.

Campbell, R., Garwood, J., Franklin, S., Howard, D., Landis, T. & Regard, M. (1990). Neuropsychological studies of the auditory-visual fusion illusion. Neuropsychologia, 28, 787-802.

Colombo, J., & Mitchell, D.W., (1990). Individual differences in early visual attention: Fixation time and information processing. In J. Colombo & J. Fagen (Eds.), Individual Differences in Infancy: Reliability, Stability, Prediction (pp. 193-227). Hillsdale, NJ: LEA.

Cutler, A. (1989). Straw Modules. Behavioral and Brain Sciences, 12, 760 -762.

de Gelder, B., Vroomen, J. & van der Heide, L. (1991). Face recognition and lip-reading in autism. Special Issue: Face recognition. European Journal of Cognitive Psychology, 3, 69-86.

Desjardins, R.N., Rogers, J. & Werker, J.F. (in press). An exploration of why preschoolers perform differently than do adults in audiovisual speech perception tasks. Journal of Experimental Child Psychology.

Dodd, B. (1977). The role of vision in the perception of speech. Perception, 6, 31-40.

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

Dodd, B. (1987). The acquisition of lip-reading skills by normally hearing children. In B. Dodd & R. Campbell (Eds.), Hearing by Eye: The Psychology of Lip-Reading (pp. 163-175). London, England: LEA.

Fodor, J. (1983). The Modularity of Mind. Cambridge, MA: MIT Press.

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

Gerdemann, A. (1994). Temporal Incongruity & the McGurk Effect. Unpublished Master's Thesis, University of Arizona.

Gibson, E.J. (1969). Principles of Perceptual Learning and Development. New York: Appleton-Century-Crofts.

Goodale, M.S. (1988). Modularity in visuomotor control: From input to output. In Z.W. Pylyshyn (Ed.), Computational Processes in Human Vision: An Interdisciplinary Perspective, pp. 262-285. Norwood, NJ: Ablex.

Green, K.P. (1995). The Influence of an Inverted Face on the McGurk Effect. Unpublished manuscript.

Green, K.P., Kuhl, P.K., Meltzoff, A.N. & Stevens, E.B. (1991). Integrating speech information across talkers, gender, and sensory modality: Female faces and male voices in the McGurk effect. Perception & Psychophysics, 38, 269-276.

Hockley, N.S. (1994). The Development of Audiovisual Speech Perception. Unpublished Master's thesis, McGill University.

Hockley, N.S. & Polka, L. (November, 1994). A Developmental Study of Audiovisual Speech Perception Using the Mcgurk Paradigm. Poster presented as the 12th Meeting of the Acoustical Society of America, Austin, Texas.

Horowitz, F.D., Paden, L., Bhana, K. & Self, P. (1972). An infant-control procedure for studying infant visual fixations. Developmental Psychology, 7, 90

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M. & Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. Developmental Psychology, 27, 236-248.

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

Krinsky-McHale, S.J., Devenny, D.A. & Bornstein, M.H. (1997, April). Prediction of Language Acquisition from Infant Habituation Measures. Poster presented at the Society for Research in Child Development, Washington, DC.

Kuhl, P.K. & Meltzoff, A.N. (1982). The bimodal perception of speech in infancy. Science, 218, 1138-1141.

Kuhl, P.K. & Meltzoff, A.N. (1984). The intermodal representation of speech in infants. Infants Behavior and Development, 7, 361-381.

Ladefoged, P. (1993). A Course in Phonetics (3rd ed.). Fort Worth: Harcourt Brace Jovanovich.

Legerstee, M. (1990). Infants use multimodal information to imitate speech sounds. Infant Behavior and Development, 13, 343-354.

Lewkowicz, D.J. (1994). Development of intersensory perception in human infants. In D.J. Lewkowicz & R. Lickliter (Eds.), The Development of Intersensory Perception: Comparative Perspectives (pp. 165-203). Hillsdale, NJ: LEA

Lewkowicz, D.J. (in press). 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.

Liberman, A.M., Cooper, F.S., Shankweiler, D.P., & Studdert-Kennedy, M. (1967). Perception of the speech code. Psychological Review, 74, 431-461.

Liberman, A.M. & Mattingly, I.G. (1985). Motor theory of speech perception revised. Cognition, 21, 1-36.

Manuel, S.Y., Repp, B.H., Liberman, A.M. & Studdert-Kennedy, M. (1983, November). Exploring the "McGurk Effect". Paper presented at 24th annual meeting of the Psychonomic Society, San Diego.

Massaro, D.W. (1984). Children's perception of visual and auditory speech. Child Development, 55, 1777-1788.

Massaro, D.W. (1987). Speech Perception by Ear and Eye: A Paradigm for Psychological Inquiry. Hillsdale, NJ: LEA.

Massaro, D.W. (1989). Multiple book review of Speech Perception by Ear and Eye: A Paradigm for Psychological Inquiry. Behavioral and Brain Sciences, 12, 741-794.

Massaro, D.W. (1994). Modularity of information, not processing. Current Psychology of Cognition, 13, 97-102.

Massaro, D.W., Cohen, M.M., Gesi, A., Heredia, R. & Tsuzaki, M. (1993). Bimodal speech perception: An examination across languages. Journal of Phonetics, 21, 445-478.

Massaro, D.W., Cohen, M.M., Smeele, P.M.T. (1995). Cross-linguistic comparisons in the integration of visual and auditory speech. Memory & Cognition, 23, 113-131.

Massaro, D. W., Thompson, L.A., Barron, B. & Laren, E. (1986). Developmental changes in visual and auditory contributions to speech perception. Journal of Experimental Child Psychology, 41, 93-113.

McCall, R.B. & Carriger, M.S. (1993). A meta-analysis of infant habituation and recognition memory performance as predictors of later IQ. Child Development, 64, 57-79.

McGurk, H. (1988, March 2). Developmental Psychology and the Vision of Speech. Inaugural Lecture, University of Surrey.

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

McKain, K., Studdert-Kennedy, M., Spieker, S. & Stern, D. (1983). Infant intermodal speech perception is a left-hemisphere function. Science, 219, 1347-1349.

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. 335-369). Hillsdale, NJ: LEA

Miller, G.A. & Nicely, P.E. (1955). An analysis of perceptual confusions among some English consonants, Journal of the Acoustical Society of America, 27, 338-352.

Mills, A.E. (1987). The development of phonology in the blind child. In B. Dodd & R. Campbell (Eds.), Hearing by Eye: The Psychology of Lip-Reading (pp. 145-161). London, England: LEA.

Morisset, C.E., Barnard, K.E. & Booth, C.L. (1995). Toddlers' language development: Sex differences within social risk. Developmental Psychology, 31, 851-865.

Munhall, K.G., Gribble, P., Sacco, L. & Ward, M. (1996). Temporal constraints on the McGurk effect. Perception & Psychophysics, 58, 351-362.

Ojemann, G.A. (1988). Effect of cortical and subcortical stimulation on human language and verbal memory. In F. Plum (Ed.) Language, Communication and the Brain, pp. 101-115. New York: Raven Press.

Ojemann, G.A. (1991). Cortical organization of language. The Journal of Neuroscience, 11, 2281-2287.

Piaget, J. (1952). The Origins of Intelligence in Children. New York: International Universities Press.

Radeau, M. (1994). Auditory-visual spatial interaction and modularity. Current Psychology of Cognition, 13, 3-51.

Rosenblum, L.D. (1994). How special is audiovisual speech integration? Current Psychology of Cognition, 13, 110-116.

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

Saldaña, H.M. & Rosenblum, L.D., & Osinga, T. (1992). Visual influences of heard speech syllables with a reduced visual image. Journal of the Acoustical Society of America, 92, 2340.

Sams, M., Aulanko, R., Hamalainen, M., Hari, R., Lounasmaa, O.V., Lu, S-T. & Simola, J. (1991). Seeing speech: Visual information from lip movements modifies activity in the human auditory cortex. Neuroscience Letters, 127, 141-145.

Sekiyama, K. (1997). Cultural and linguistic factors in audiovisual speech processing: The McGurk effect in Chinese subjects. Perception & Psychophysics, 59, 73-80.

Sekiyama, K. & Tohkura, Y. (1993). Inter-language differences in the influence of visual cues in speech perception. Journal of Phonetics, 21, 427-444.

Siegel, S. & Castellan, N.J., Jr. (1988). Nonparametric Statistics for the Behavioral Sciences (2nd ed.). New York: McGraw-Hill.

Siva, N., Stevens, E.B., Kuhl, P.K. & Meltzoff, A.N. (1995). A comparison between cerebral-palsied and normal adults in the perception of auditory-visual illusions. Journal of the Acoustical Society of America, 98, 2983.

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

Spelke, E.S. & Cortelyou, A. (1981). Perceptual aspects of social knowing: Looking and listening in infancy. In M.E. Lamb & L.R. Sherrod (Eds.), Infant Social Cognition: Empirical and Theoretical Considerations. Hillsdale, NJ: LEA (pp.-61-84).

Summerfield, Q. (1991). Visual perception of phonetic gestures. In I.G. Mattingly (Ed.), Modularity and the motor theory of speech perception. Hillsdale, NJ: LEA (pp. 117-137).

Tees, R.C. (1994). Early stimulation history, the cortex, and intersensory functioning in infrahumans: Space and time. In D.J. Lewkowicz & R. Lickliter (Eds.), The Development of Intersensory Perception: Comparative Perspectives (pp. 107-131). Hillsdale, NJ: LEA

Tighe, T.J. & Powlison, L.B. (1978). Sex differences in infant habituation research: A survery and some hypotheses. Bulletin of the Psychonomic Society, 12, 337-340.

Turkewitz, G. (1994). Sources of order for intersensory functioning. In D.J. Lewkowicz & R. Lickliter (Eds.), The Development of Intersensory Perception: Comparative Perspectives (pp. 3-17). Hillsdale, NJ: LEA

Turkewitz, G., Lewkowicz, D.J. & Gardner, J.M. (1983). Determinants of infants' perception. In J.S. Rosenblatt, R.A. Hinde, C. Beer & M.C. Busnel (Eds.), Advances in the Study of Behavior (Vol. 13, pp. 39-62). New York: Academic Press.

Wagner, S.H. & Sakovits, L.J. (1986). A process analysis of infant visual and cross-modal recognition memory: Implication for an amodal code. In L.P. Lipsitt & C. Rovee-Collier (Eds.), Advances in Infancy Research: Vol. 4. (pp. 195-217). Norwood, NJ: Ablex.

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

Walker, S., & Bruce, V., & O'Malley, C. (1995). Facial identity and facial speech processing: Familiar faces and voices in the McGurk effect. Perception & Psychophysics, 57, 1124-1133.

Walker-Andrews, A. (1994). Taxonomy for intermodal relations. In D.J. Lewkowicz & R. Lickliter (Eds.), The Development of Intersensory Perception: Comparative Perspectives (pp. 39-56). Hillsdale, NJ: LEA

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

Werker, J.F., McGurk, H. & Frost, P.E. (1992). La langue et les lévres: Cross-language influences on bimodal speech perception. Canadian Journal of Psychology, 46, 551-568.

APPENDIX A.

Fuzzy Logical Model of Perception (FLMP)

Massaro (e.g., 1987, 1989, 1994) has developed a model of the evaluation and integration of information which employs fuzzy logic: "a continuously valued logic that represents the truth of propositions in terms of truth values that range between zero (false) to one (true)" (Massaro, 1989, p. 742). In audiovisual speech perception, the signals from each modality are evaluated independently for certain features and assigned a value representing the degree to which the signal supports each of the relevant alternatives. For example, a visual /da/ is evaluated for degree to which lip opening is present and an auditory /da/ for the onset of F2 and F3 (slightly falling). The degree to which each stimulus supports the relevant features is determined. Then, the resulting values are combined in the integration stage resulting in an overall degree to which the combined sources match one of two alternatives. A third stage follows in which a decision or classification of the stimulus occurs.

A variety of demonstrations by Massaro involving natural and synthesized speech and faces on a /ba-da/ continuum support his contention that FLMP accounts better for human performance than a categorical model in which a decision is made about the phonetic category of the stimulus in each of the channels before integration occurs; or an auditory dominance model in which the effect of visual information is greatest when the auditory information is ambiguous (see Massaro, 1987).

In the FLMP, information from the two channels is initially evaluated independently. However Bernstein (1989) argues that judgments about the presence of particular features in the signal such as voicing cannot be made before integration of information from both channels occurs. For example, the feature

voicing involves an assessment of the relative timing of two events perceived through different modalities; the onset of glottal pulsing is detected through the auditory channel while the release of the consonant is observed through the visual channel. Thus, the combination of the two sources of information must precede evaluation of the voicing feature.

Massaro argues that "modularity predicts that the processes responsible for audiovisual speech perception should be unique" (1989, p.741). But, FLMP can be used to model, for example, letter and word recognition, and visual depth perception as well as audiovisual speech perception. Any domain in which multiple sources of information are evaluated, integrated and classified with respect to stored representations can be modeled accurately with FLMP according to Massaro. What is modular about speech is not the processing, but the information (Massaro, 1994).

Massaro's argument hinges on the notion of independence of information, yet this is precisely what is meant by 'encapsulation' in the Fodorian module. As Cutler (1989, p. 761) cogently states, "what domain-specificity does *not* entail is that the operations of a modular system need be unlike the operations of other systems in every respect. Indeed it would be astonishing if this were so." If Cutler's assessment is correct, then Massaro does not provide a critical challenge to the modularity hypothesis with respect to speech perception.

APPENDIX B

Developmental Models of Intersensory Processing

Developmental models of intersensory processing fall into one of two categories: integration or differentiation models. Integration models specify that the neonate has no capacity for intermodal perception; she hears, sees, tastes, smells, and feels things, but has no capacity to relate what she sees to what she hears etc.. It is only over the course of development that the capacity to integrate information across channels arises. Most often associated with this position is Piaget (1952) who argued that through acting on the world (initially through innate reflexes) and through reciprocal assimilation the infant is able eventually to come to perceive unitary objects and events.

Differentiation models are the exact opposite of integration models. Bower (1974) holds the most extreme position: neonates are said to be unable to distinguish between the senses at all—there is a "primitive unity." The Intensity hypothesis (e.g., Turkewitz, Lewkowicz & Gardner, 1983) specifies that young infants respond to intensities across modalities rather than to specific amodal properties such as duration. More recently, Turkewitz (1994) has developed this position further: The sequential onset of the different modalities reduces the amount of competing sensory information and thus makes development of a structure organizing the input easier. When combined with intensity-based responding, this allows the development of the capacity to recognize equivalent information in different modalities. The Gibsonian (1969) position is that neonates are able to extract amodal invariants from the environment and to coordinate information from different modalities. Development consists of learning to perceive finer and finer relations between the senses.

APPENDIX C

Pretesting with Adults

In order to determine whether the chosen incongruent audiovisual stimuli elicited the expected responses of visual capture (/vi/ responses to audio /bi/-visual /vi/) and auditory capture (/vi/ responses to audio /vi/-visual /bi/), the stimuli were presented to adult participants before testing with infants began.

Method

Participants

Eight adults (5 females, 3 males) between the ages of 21 and 35 years volunteered to participate. All reported no known hearing loss and all had normal or corrected vision.

Stimuli

A total of 14 different audiovisual syllables were presented, 8 of which had mismatched (incongruent) auditory and visual syllables. Only four of the test stimuli are relevant to the current report: audiovisual /vi/, audiovisual /bi/, audio /bi/-visual /vi/, audio /vi/-visual /bi/. Each of the 14 stimuli was randomly presented once per block. Ten blocks, each with different random orders were presented, for a total of 140 trials.

Equipment

The same basic equipment was used as in the infant studies with the exception that the adults viewed the speaker's face on a monitor in the control room and listened to the audio portion via headphones.

Procedure

Adults were tested individually. The adult was seated at a distance of approximately 3 feet from the monitor with headphones over his/her ears. The experimenter sat adjacent to the adult but off to the side so that she could not see the visual stimulus; she pressed a button to begin each trial when the adult was looking at the monitor. Participants were instructed first to tell the experimenter what syllable the speaker said, and second to give a rating of how confident he or she was that the reported syllable was the one heard. A seven point rating scale was used with 1 corresponding to "not at all confident" and 7 corresponding to "very confident"; 4 was the mid-point of the scale, a "neutral" response. Two stimuli not used in the test trials were presented initially to allow the adult to become familiarized to the task. The experimenter recorded the adult's responses.

The testing session lasted approximately 30 minutes. Participants were asked at the end of each block whether they wanted to take a brief break, or continue testing immediately.

Participants were told that the experimenter wanted to pretest some syllables in order to choose the best ones for use in an audiovisual speech perception study with infants; they were not told that some of the stimuli were auditory-visual mismatches. After testing was over, adults were fully debriefed about the nature of the stimuli.

Results and Discussion

Only the results of the four test stimuli relevant to the current infant studies will be presented.

Identification responses

The number of correct responses out of 10 for each of the four test stimuli was entered into a one-way analysis of variance. The syllable /vi/ (i.e., visual capture) was considered a correct response for the audio /bi/-visual /vi/. The syllable /vi/ (i.e., auditory capture) was also considered a correct response for the audio /vi/-visual /bi/. The main effect for Syllable was not significant, F (3, 21) = 1.346, $p > .1$, suggesting there was no difference in the number of correct responses. Planned orthogonal contrasts revealed that adults did not report more 'correct' responses on the two matched audiovisual stimuli than on the two mismatched audiovisual stimuli, $F(1,21) = 2.127$, $p > .1$. Moreover adults' accuracy at identifying the audiovisual /bi/ did not differ from that of the audiovisual /vi/, $F(1, 21) = 0.019$, $p > .1$. Adults did not differ in the number of /vi/ response to audio /bi/-visual /vi/ and audio /vi/-visual /bi/, $F(1, 21) = 1.891$, $p > .1$.

Adults correctly identified the audiovisual /bi/ as /bi/ 100% of the time, and audiovisual /vi/ as /vi/ 99% of the time. On average, adults heard audio /vi/-visual /bi/ as /vi/ 96% of the time. Audio /bi/-visual /vi/ was heard as /vi/ 84% of the time on average; one adult consistently heard the audio /bi/-visual /vi/ as /bi/. The remaining seven adults heard the audio /bi/-visual /vi/ as /vi/ 96% of the time.

Confidence ratings

Adults' confidence ratings for the four stimuli were averaged for each stimulus separately and entered into a one-way analysis of variance. Only ratings for /bi/ or /vi/ responses to each of the four stimuli were included. The main effect for syllable was significant, $F(1, 21) = 9.529$, $p < .0005$. Planned orthogonal contrasts indicated that adults' were more confident in their responses for the matched audiovisual stimuli than for the mismatched audiovisual stimuli, $F(1, 21) = 26.166$, $p < .0001$. However, confidence ratings did not differ between the two

matched audiovisual stimuli, $F(1, 21) = 2.435$, $p > .1$, nor between the two mismatched audiovisual stimuli, $F(1, 21) = 0.010$, $p > .1$.

Although there were significant differences in confidence ratings between the matched and mismatched audiovisual stimuli ($M = 6.1$, $SD = 0.5$ for audiovisual /bi/; $M = 6.7$, $SD = 0.3$ for audiovisual /vi/; $M = 5.0$, $SD = 1.1$ for audio /bi/-visual /vi/; $M = 5.1$, $SD = 5.1$ for audio /vi/-visual /bi/), the average confidence ratings were on the positive end of the scale (4 = neutral, 7 = very confident). Interestingly, the one participant who reported hearing audio /bi/-visual /vi/ as /bi/, rather than as /vi/ as expected, gave a confidence rating of 3.1 on average for that test syllable.

Participants comments

After testing was completed, adults were asked to tell the experimenter if they noticed whether the auditory and visual component did not match on any of the trials, and on what proportion of trials mismatches occurred. Responses varied: none or some towards the end ($n = 2$), 5% ($n = 1$), 20% ($n = 3$), 30% ($n = 1$), and over 50% ($n = 1$) mismatches. All but one adult underestimated the number of mismatched auditory and visual syllables (57%).

Summary

The identification responses suggested that the audio /bi/-visual /vi/ typically is heard as /vi/ by adults (for 7 of 8 adults on 96% of the trials), and that audio /vi/-visual /bi/ is also typically heard as /vi/ by adults (on 96% of the trials). The rate of visual capture in the first instance is similar to that reported by Manuel et al. (1983) for audio /ba/-visual /va/ (98% /va/ responses) while the rate of auditory capture obtained in this sample is much higher than that obtained by Manuel et al. for audio /va/-visual /ba/ (45% /va/ responses). The fact that strong /vi/ percepts were attained for each of the incongruent audiovisual pairings suggests that these tokens are suitable for the infant studies.

The difference in confidence ratings between the matched and mismatched audiovisual tokens suggests two possibilities. Either the syllable 'heard' as a result of integration is not a good exemplar of that consonant category, or although integration of the auditory and visible speech is mandatory, the perceptual system 'noticed' a mismatch between the auditory and visible speech. Further research is needed in order to determine which of these explanations accounts for the confidence ratings. These results suggest, however, that including a confidence rating in adult studies of audiovisual speech perception might prove to be a useful and informative measure.