EUROPEAN JOURNAL OF DEVELOPMENTAL PSYCHOLOGY
2008, 5 (1), 68–91

Psychology Press
Taylor & Francis Group

Intermodal matching of vision and audition in infancy: A proposal for a new taxonomy

Audrey Guihou and Jacques Vauclair
University of Provence, Aix-en-Provence, France

The aim of this review paper is first to devise a common framework for the various procedures used in experimental investigations of intermodal auditory–visual matching by human infants, and second to propose a new taxonomy of intermodal tasks in order to gain a better understanding of the perceptual-cognitive processing underlying these tasks and their relationships to other cognitive achievements such as the development of language. Based on an examination of the tasks used in the developmental literature, we suggest analysing them in terms of (a) their mode of presentation (simultaneous or sequential) for accessing the information in the two modalities (auditory and visual) and (b) the type of relation (amodal or arbitrary) between these two stimulus sources. A review of the literature in the light of this classification shows that most experimental studies employ parallel intermodal presentation (PIP) using both amodal and arbitrary relations between stimuli. Very few investigations have used sequential intermodal presentation (SIP) with amodal relations, and none have used SIP procedures with arbitrary relations. It is claimed here that this last approach might be the most appropriate one for furthering our understanding of (a) the categorical and semantic processing involved in intermodal matching and (b) how the developing infant learns to process words.

INTRODUCTION

From birth, infants are immersed in an environment of information characterized by multimodal properties. Any person or object can be simultaneously heard and seen, if not touched and smelled. Among the

Correspondence should be addressed to Audrey Guihou, Department of Psychology & Center for Research in Psychology of Cognition, Language and Emotion, University of Provence, Aix-en-Provence, France. E-mail: Audrey.Guihou@aixup.univ-aix.fr

Writing of this paper was made possible with a grant for the CNRS-OHLL Program (Origines de l'Homme, du Langage et des Langues).
The authors thank A. Streri for her comments on a previous version of this manuscript and an anonymous reviewer for constructive suggestions.

http://www.psypress.com/edp DOI: 10.1080/17405620600760409

range of sensory modalities that constitute the ingredients of the infant's first experiences, audition and vision are the most prominent. Studying how information from sensory channels is intermodally perceived and integrated represents one of the major challenges in the field of infant cognition (Lewkowicz & Lickliter, 1994). Although this field of investigation is a very active one, researchers in developmental psychology use a variety of procedures and refer to different concepts when addressing the question of how infants perceive and combine intermodal information. Consequently, this state of affairs makes it difficult to achieve an overall view of the domain and of the relationships between intermodal perception and language (speech) acquisition during infancy.

A number of authors (e.g., Gogate & Bahrick 1998; Gogate, Walker-Andrews, & Bahrick, 2001; Molfese, Morse, & Peters, 1990; Roy & Pentland, 2002) have highlighted the role of multimodal auditory and visual stimulation in language acquisition. This bimodal aspect can be seen in the importance of gestures in infant-directed speech (Iverson, Capirci, Longobardi, & Caselli, 1999; Zukow-Goldring, 1997), and in the fact that, before speaking, caregivers usually try to establish eye contact with the infant (Butterworth, 2001). It is therefore crucial to understand how infants make the connection between visual and auditory information during early development.

In this review paper we will: (1) propose a new taxonomy for intermodal tasks based on procedural criteria; (2) summarize and analyse some of the most relevant data in the literature on auditory–visual matching during infancy, using the new taxonomy; and (3) consider the implications of this methodological taxonomy for the cognitive processes underlying intermodal matching, and its impact on our knowledge of the development of matching abilities in infancy and how they are related to language skills.

DIVERSITY OF INTERMODAL TASKS AND TYPES OF MATCHING

Many researchers have explored the development of intermodal processes in infancy, especially in recent years, but, as we shall see later, it is difficult to compare available data within the same methodological framework. Two key problems in this area are the lack of a common lexicon for describing intermodal procedures and tasks, and the confusion of experimental procedures and psychological processes.

Intermodal tasks can be divided into two broad categories: (1) situations or tasks for which the individual has to transfer information from one modality to another, i.e., usually described as "cross-modal transfer" tasks (e.g., for infants see Mendelson & Ferland, 1982; for children see Blank & Bridger, 1964; for nonhuman primates see Savage-Rumbaugh, Sevcik, & Hopkins, 1988); (2) situations or tasks where an individual has to integrate stimuli via two different sensory channels so as to detect a link (or conflict) between the stimuli. These situations cover several concepts: *detection of auditory visual equivalence* (e.g., Lewkowicz, 1994), *detection of auditory–visual relations/correspondences for the same event* (e.g., Bahrick, 1983, 1987, 1992, 1994; Kuhl & Meltzoff, 1982, 1984; MacKain, Studdert-Kennedy, Spieker, & Sern, 1983), *intermodal matching* (e.g., Hashiya & Kojima, 1997, 2001; Hashiya, 1999, with nonhuman primates), and *learning of arbitrary auditory–visual associations* (e.g., Bahrick, 1988; Fenwick & Morongiello, 1998; Slater, Quinn, Brown, & Hayes, 1999; Brookes, Slater, Quinn, Lewkowicz, Hayes, & Brown, 2001).

It is important to note how difficult it is, in all these examples, to determine whether the terms listed above are used by the authors to depict their experimental procedures or the perceptual/cognitive processes involved in these situations. The variety of terms used to describe the tasks makes it difficult to compare studies of intermodal infant behaviour and, consequently, the abilities that these tasks attempt to assess. This is our main reason for proposing a new classification of these procedures and tasks based on a meta-analysis of the literature.

Several taxonomies of intermodal tasks and intermodal processes are available in the literature, although it is not always a straightforward matter to distinguish between them; some of the classifications emphasize processes, whereas others focus more on procedural differences.

The taxonomy proposed by Walker-Andrews (1994) concerns task features and, more specifically, the nature of the relationships between the two kinds of stimuli. This author distinguishes four categories. The first category, *amodal specification*, includes all tasks for which the infant has to match two objects or events on the basis of amodal properties such as synchrony, tempo, and intensity. The second category covers tasks for which the intermodal relation is described as *arbitrary and artificial*. An example of a device with this double feature is a cellular phone, which can vibrate, ring, or play melodies. Relationships between words and their referents also belong to this category. In the third category, the relation is *arbitrary but natural*, or *idiosyncratic but natural*, as in the association between a voice and a face. In the fourth it is "typical": this is the case for the sound produced by a heavy object hitting the ground. This relation is said to be typical because the sound is partially determined by the nature of the ground and the object.

Other authors have proposed classifications that revolve around the processes required to solve intermodal tasks. Blank and Bridger (1964) distinguished between *cross-modal equivalence* tasks, where young participants have to recognize the same object using various sensory modalities, and *cross-modal concept* tasks, which are more complex, since the same concept has to be used to distinguish problems involving dissimilar but analogous stimuli in different modalities. For example, in a cross-modal transfer task (Blank & Bridger, 1964), young participants had to use a piece of information acquired though one modality (concept of numerosity: one or two lights) to solve a problem presented in another modality (one or two sounds). The distinction between these two types of tasks is potentially useful because it can help differentiate two levels of conceptualization, even though amodal information is available in both situations.

Another classification of intermodal transfer tasks, based on the underlying processes, was suggested by Bushnell (1994). This author distinguished two processes potentially involved in any intermodal task, *matching by recognition* and *matching by analysis*. According to Bushnell, matching by recognition refers to a global process that is used spontaneously and is based on the semantic memory of the stimuli, whereas matching by analysis concerns amodal correspondences between stimuli. The two processes can operate simultaneously, in which case there is a dual system of perceptual processing. But depending on what stimuli are presented and the infant's experience with those stimuli, observations have shown that one of the two processes dominates. This model is particularly useful for understanding the different steps involved in intermodal tasks, because it takes into account not only the nature of the stimulus, but also each infant's specific experience. However, as the author clearly explains, in most cases the two processes are activated in parallel.

More recently, Lewkowicz (2000) described three forms of intersensory integration processes: (1) *detection of amodal invariants*; (2) *intersensory association of modality-specific cues* (i.e., when the infant has to match two stimuli on the basis of arbitrary information); and (3) *nonspecific effect of stimulation in one modality on responsiveness to stimulation in another modality*. This last form has often been studied with adults. The most famous examples are the McGurk effect (McGurk & MacDonald, 1976), and the ventriloquism illusion (Radeau, 1994). With infants, similar phenomena have been reported for the processing of visual and auditory stimuli (e.g., Meltzoff & Kuhl, 1994; see below for additional information on the McGurk effect).

Each of the above proposals has its merits, but it seems that none of these classifications of intermodal relations offers a complete view of the way that the various relations are accessed by the infant. In particular, they do not take into account a feature that we consider crucial, namely, the way (simultaneous or sequential) information is presented in the two modalities. On the one hand, auditory and visual stimuli can be presented simultaneously to the infant. We will call this procedure *parallel intermodal presentation* (or PIP). On the other hand, there are tasks in which the infant has access to the first stimulus in one sensory channel, and to the second stimulus in another sensory channel. We will call this procedure *serial intermodal presentation* (or SIP).

Depending on the access mode (PIP or SIP), it is quite possible that the way the information is handled can affect the infant's intermodal performance, simply because the memory load is not the same in each mode. We will also attempt to show to what extent serial vs. parallel presentation is related to the nature and complexity of the underlying cognitive processes involved in the intermodal processing.

PROPOSAL FOR A NEW TAXONOMY OF AUDITORY–VISUAL MATCHING SITUATIONS

In this section, we propose a framework for classifying intermodal tasks by considering their degree of difficulty. It is based on the idea that different methodological approaches should target different cognitive processes, but the methodological section of this paper will not include an in-depth study of the relationships between procedures and tasks and their underlying cognitive processes. Instead, our main goal is to classify a significant number of studies published in this field so as to clarify the state of our knowledge on auditory–visual intermodal capacities in infancy. Our classification is built around two basic features. The first is the temporal organization of the stimuli as presented; the second is the type of relationship between the two categories of stimuli, i.e., amodal or arbitrary.

Temporal presentation of stimuli

A key characteristic in our classification scheme is the order of the stimuli in the two modalities. In some studies, the visual stimulus is presented at the same time as the auditory stimulus; in other studies, the stimuli are presented one after the other. This is why we believe it is useful to introduce the distinction between PIP and SIP procedures. The proposed distinction is not confined solely to vision and audition, but can be applied to other intermodal situations (e.g., haptic–visual). For examples, see the studies by Meltzoff and Borton (1979) and Streri and Gentaz (2003), who used a SIP procedure, and the study by Kaye and Bower (1994), who used a PIP procedure to examine cross-modal recognition of textures/shapes from hand to eyes.

The section that follows is devoted to describing and illustrating these two procedures.

Parallel intermodal presentation procedures

Most studies have used auditory–visual parallel presentation procedures, with tasks often requiring "intermodal perception" or "cross-modal perception" (e.g., Bahrick, 1987; Slater et al., 1999; Soken & Pick, 1992). Parallel presentation is present when an individual is requested to simultaneously integrate information perceived in different sensory modalities. It follows that there is parallel presentation only if one stimulus (object or event) is simultaneously available to different modalities, and if the individual has to consider these modalities as being related to the same source.

PIP procedures are always based on at least one amodal property, namely the simultaneity of the auditory perception of the sound and the visual perception of the sound's source. Simultaneity exists when the stimulus perceived via the first sensory channel is still present in the sensory memory of the infant when he/she perceives the same stimulus via the second channel. For example, when watching a cartoon on TV with two characters, an infant will associate the voice heard with the character whose lips are moving: this situation is a case of parallel intermodal presentation.

Serial intermodal presentation procedures

It can happen that the auditory perception of the stimulus is not concurrent with its visual perception. In this case, the stimuli are presented successively. For example, an infant could hear his/her mother's voice, and then a few seconds later, see both the mother and the father and focus his/her attention on the mother. This behaviour is rather complex because the individual has to abstract and memorize sensory information perceived in one modality, and match it with another element of information coming from a different sensory channel. This second situation is usually called *intermodal transfer* (e.g., Mendelson & Ferland, 1982) or *cross-modal transfer* (Gottfried, Rose, & Bridger, 1977), where cross-modal transfer is defined as "the ability to convey information that is acquired in one sensory modality to another" (Gottfried, Rose, & Bridger, 1977, p. 118).

Our classification refers to the fact that in *serial intermodal presentation* (SIP procedures), the individual is requested to sequentially associate certain properties obtained from different sensory modalities for the purposes of identifying and defining a single object or event. A typical SIP procedure would thus be to present an infant with a stimulus in one modality (e.g., the sound an object makes), and then after a lapse of time, to present the same stimulus in another modality (e.g., the sight of that object), in order to determine whether the infant can recognize it.

Time constraints seem to have been overlooked in the intermodal-processing, literature, except in the study by Streri (2000). This author proposed a distinction between *intermodal perception* and *intermodal transfer* using time as the criterion, i.e., whether sequential or simultaneous access to the information from both modalities was available. As explained below, time constraints in SIP and PIP procedures could be important because they may involve different levels of cognitive processing.

TYPES OF RELATIONSHIPS BETWEEN THE TWO STIMULI

In addition to the time criterion for the evaluation of intermodal tasks, a second criterion of interest is the nature of the relationship between the two sensory modalities. Given that this criterion has already been included in other classifications (see above, and Walker-Andrews, 1994), we will only comment briefly on the two main types of relationships that characterize intermodal presentation tasks: the amodal type and the arbitrary type.

Amodal relationships

In amodal relationships, the connection between the two stimuli is made by way of common properties. Gogate and Bahrick (1998) consider relations to be amodal when information is "*completely redundant across two or more senses*" (1998, p. 134). Usually, amodal properties are time-related (e.g., rhythm, rate, tempo, duration, etc.), although other properties can also be shared by the two stimulus categories (e.g., intensity, composition, location, or quantity; Bahrick, Hernandez-Reif, & Flom, 2005; for a review see Lewkowicz, 2000).

Arbitrary relationships

Intermodal relations are said to be *arbitrary* or *modality-specific* when "*information can be conveyed primarily though one sense modality alone*" (Bahrick et al., 2005). In this kind of relationship "*information across the senses is not redundant, and may not predictably occur together in nature*" (Gogate & Bahrick, 1998, p. 135). These authors proposed several examples of arbitrary relationships, such as "*the relationship between words and their referents*" (Gogate & Bahrick, 1998, p. 135) and "*the relation between the color of an object and its sound at impact*" (Gogate & Bahrick, 1998, p. 135). Arbitrary information is not immediately detected and thus needs to be learned. Generally, such relationships concern modality-specific properties. For example, the association between a person's voice and face is acquired through experience. But these are not the only modality-specific properties that can be arbitrarily related. A modality-specific property may be related to an arbitrary one. As an example, an infant can quickly learn the association between the sound of a toy and its shape, the latter being an amodal property that can be perceived visually (e.g., when the parent manipulates the toy) as well as by haptic exploration (by the infant). A final case concerns the arbitrary association of two amodal properties. It is unlikely that such relationships occur naturally, but they can be experimentally manipulated. An example would be an intermodal learning experiment requiring making the connection between one amodal property of a given object (e.g., its size) and another of its amodal properties (e.g., its shape only haptically perceived).

This type of relationship between two stimuli is likely to require a higher degree of conceptualization than amodal matching, because the infant has to match two stimuli that have no shared amodal properties, except for cases of temporal synchronies (see below). A typical example is matching a face to a voice heard a few seconds earlier; in this case an arbitrary SIP procedure would be used. An example of a task using an arbitrary PIP would be an experiment where the infant simultaneously sees his/her mother and an unknown woman and hears each voice, one at a time (Spelke & Owsley, 1979). Situations involving arbitrary relations have an obvious ecological reality, so it is surprising that this type of stimulus is rarely found in the developmental literature as a case of intermodal mapping (except for early language comprehension). According to Oviatt (1980), Hirsh-Pasek and Gollinkoff (1996), and Werker Cohen, Lloyd, Casasola, and Stager (1998), an example of simple auditory–visual matching is that of associating a single word with the matching object or person before reaching the stage where the symbolic meaning of the word is understood. This is why we consider certain word–object associations to be based on PIP procedures. However, note that both kinds of properties (amodal and arbitrary) are available in most intermodal events or situations. For example, speech is an audio-visual activity that combines several kinds of information, since it simultaneously provides amodal information (temporal synchrony, rhythm, intensity) and also arbitrary relations via different kinds of face–voice combinations.

With our two procedural categories (PIP and SIP) and the distinction between amodal and arbitrary relations, we can now differentiate four types of auditory–visual intermodal tasks.

REVIEW OF AUDITORY–VISUAL INTERMODAL TASKS

Applying the taxonomy we devised for analysing the literature, we can see that most experimental studies on auditory–visual intermodal abilities in infants have used *amodal and simultaneous intermodal presentation*.

Studies using amodal and parallel intermodal presentations

Over the past forty years or so, a number of studies have provided strong evidence of an early link between auditory and visual perception during infancy. One of the first observations of the precocious or innate co-ordinated functioning of these sensory modalities was reported by

Wertheimer (1961), who conducted a series of tests on his newborn daughter. After lightly rattling keys close to the baby's ear, Wertheimer observed that the infant turned her eyes towards the source of the sound. Note that McGurk, Turnure, and Creighton (1977) were unable to replicate Wertheimer's findings. However, since then, a number of studies have confirmed this early integration of vision and audition (e.g., Butterworth & Castillo, 1976; Field, DiFranco, Dodwell, & Muir, 1979; Muir & Field, 1979, 1980). More recently, several research projects have been conducted to determine the conditions needed for this intersensory perception to emerge. These studies (most of which are cited below) have shown that by 2 or 3 months of age infants are able to detect some amodal invariants across the different sensory perceptions. The most important invariants are the temporal synchrony and spatial co-location of visual and auditory stimuli.

PIP tasks based on temporal invariants

The co-ordination of visual and auditory perception appears to be highly dependent on the temporal cues linking the two stimuli. Several lines of evidence indicate that very young infants are sensitive to this type of information. Bahrick (1987, 1988), for example, showed two different film clips to two age groups (3.5- and 4.5-month-old infants) while playing the soundtrack that matched only one of them. Although the findings did not provide any evidence of mapping in the 3.5-month-olds, the 4.5-month-olds displayed a visual preference for the film corresponding to the sound heard. This shows that the older infants perceived the synchrony, or absence of it, between a visual and an auditory event. With an intermodal looking-preference procedure, Pickens, Field, Nawrocki, Martinez, Soutullo, and Gonzalez (1994) demonstrated the same ability to match a face to a voice on the basis of temporal synchrony in 3- and 7-month-old infants. However, 5-month-olds and pre-term 3.5- and 7-month-olds failed.

Other studies have shown that infants can detect: (a) simultaneity between the sight of an object hitting a surface and the noise of the impact (Bahrick, 1983, 1992; Spelke, 1979, 1981); (b) temporal synchrony between the occurrence of a sound and the trajectory change of the moving object (Spelke, Born, & Chu, 1983); and (c) synchrony or lack of synchrony between the visual rhythm (impact) of a moving object hitting a surface and the rhythm of the sound produced by the falling object (Spelke, 1979). Other evidence has shown that infants are also able to use multimodal temporal information to determine the substance and composition of an object bouncing off a surface (Bahrick, 1983, 1987, 1988, 1992, 2001).

These findings suggest that temporal information is a significant criterion for early auditory–visual intermodal perception. Lewkowicz (2000) published an exhaustive review of research conducted in this field, and proposed a developmental model of intermodal perception, which postulates that intersensory synchrony is the basis for the development of intermodal mapping.

PIP tasks based on spatial co-location

In addition to the temporal criterion, another type of amodal cue can favour intermodal mapping, i.e., the co-location of the two forms of stimuli. Co-location refers to the fact that the sound and the visual stimulus come from the same spatial reference point. In an experiment conducted by Humphrey, Tees, and Werker (1979), 4-month-old infants did not take this factor into account in an intersensory integration task. But in another experiment, it was an essential factor for success among 6-month-olds, even though the picture and the sound were synchronous (Lawson, 1980). Spelke (1976) provided observations supporting the hypothesis that 4-month-old infants are able to match the sound of a simple natural event to the appropriate film clip, without spatial cues. Since Spelke, several studies have used an intermodal preference procedure (e.g., recently, Houston-Price, Plunkett, & Harris, 2005).

Other data have confirmed that spatial co-location is an essential criterion for intermodal perception in infancy. For example, 2- to 8-month-old infants were able to correctly perform a PIP task when the sound and sight were temporally and spatially concordant (Morongiello, Fenwick, & Nutley, 1998); integration was also achieved with spatial congruence but not with temporal congruence only. In cases of spatial co-location but temporal asynchrony, however, infants in the same experiment did not succeed in the intermodal integration of the stimuli. Morongiello, Fenwick, and Chance (1998) reported that newborns expected a sound and a visual stimulus that occurred at the same time to be located in the same place. This phenomenon can be seen in adults, with the sensory illusion of ventriloquism (see Radeau, 1994).

Other studies have shown that while spatial and temporal cues dominate, other amodal information may influence the intermodal perception of auditory–visual events. These are cases of non-physical invariants, mostly in the area of interpersonal communication. Several studies have shown that infants perceive the emotions expressed by their interlocutor in an amodal way. One of the most typical studies in this field was conducted by Walker (1982) with 5- to 7-month-olds tested on a visual preference task. The participants were shown two film clips, one with a woman speaking cheerfully, the other with a woman speaking sadly. At the same time, the participants heard the soundtrack of one of the movies. The results showed that the infants preferred the film that matched the soundtrack. These results were later confirmed by Walker-Andrews (1986, 1988) and Soken and Pick (1992). Soken and Pick emphasized the role of movement and temporal synchrony in this type of task, thus suggesting that the role played by information of a purely emotional nature is minimal.

With the same approach, auditory – visual PIP tasks have been used to study early speech perception. The McGurk effect clearly illustrates the multimodal perception of speech in adulthood: an individual watching a face articulating the syllable "ga" and simultaneously hearing the syllable "ba" will perceive the syllable "da" (McGurk & MacDonald, 1976). Some data indicate that this kind of multimodal speech perception is also found in infants, even before language acquisition. Using lip-reading techniques, it has been shown that 5-month-old infants can recognize the temporal synchrony between the sounds emitted and the speaker's lip movements (Kuhl, Williams, & Meltzoff, 1991; MacKain et al., 1983; Meltzoff & Kuhl, 1994). Moreover, this kind of intermodal mapping can already be performed by newborns (Chen, Striano, & Rakoczy, 2004).

Starkey, Spelke, and Gelman (1983) demonstrated infants' use of another amodal element of information, i.e., quantity or numerosity. The authors tested 6- to 8-month-old infants for intersensory numerical evaluation: infants were shown coloured pictures of two or three objects (e.g., a key or a pillow). In different experiments, the infant heard two or three drumbeats, either simultaneously (Experiments 2, 3, and 4) or serially (Experiment 5). The results (for both stimulus-presentation conditions) indicated that participants displayed a visual preference for pictures showing the same number of objects as the number of drumbeats heard. These results are in line with the findings obtained by Lawson and Turkewitz (1980), who presented different numbers of visual stimuli to newborns and found that the newborns spent more time looking at the visual stimuli when the same number of sound stimuli (white-noise bursts) were presented.

In sum, the studies and findings reported in this review of the literature demonstrate a clear and early link between visual and auditory perception in infancy.

Studies using arbitrary and parallel intermodal presentation tasks

Arbitrary PIP tasks refer to: (a) tasks for which the stimuli are connected by arbitrary relationships, namely relationships with modality-specific features that are learned (e.g., the sight of a violin corresponding to its sound); and (b) tasks in which the features of the stimuli are accessed simultaneously. Few examples of tasks based on arbitrary matching are available in the literature, because the arbitrary link is often reinforced by an amodal one. A good example illustrating this combination of amodal and arbitrary information can be found in Spelke (1976). In her study, the author presented 4-month-old infants with the sound of some daily routine events while they were watching two film clips. The results showed that the infants mainly looked at the film clip that matched the sound they could hear at the same time. This study can be classified as falling into the PIP task category. However, it is likely that the infants mainly used the temporal synchrony of the events, and, thus, that both amodal and arbitrary relations were available. Another experiment based on arbitrary relations but also including amodal cues is Spelke & Owsley (1979), in which 3- to 7-month-olds looked at either their father or their mother at the same time as they heard his/her tape-recorded voice. The participants looked at their parents but did not look at any speaking adult who was not their parent. In this study, it was not stated whether the voices were concurrent with the movement of the face, so one cannot contend that only arbitrary relationships were used as a basis for matching. The same difficulty of determining whether infants rely on amodal or arbitrary links to perform audiovisual matching arose in Experiment 1 of Patterson and Werker (2002). These authors tested 4.5- to 8-month-old infants' abilities in phonetic and gender matching, on a task where a male or female articulating the vowels "a" and "i" were seen and heard. As in the previous examples, the relationships between the sound and the visual display (shape of the mouth) can be interpreted as amodal, because the infants could have matched the "shape" of the sound with the shape of the lips.

Another field of research using arbitrary PIP tasks is the exploration of infants' ability to match faces and voices on the basis of emotions. For example, Walker (1982) and Walker-Andrews (1986, 1988) tested 2- to 7-month-olds on their ability to detect the correspondence between the emotional tone of a voice and one of two facial expressions simultaneously shown to them. The 2-month-olds displayed no evidence of matching. With the 4-month-olds, there was matching for certain expressions. The 5- and 7-month-olds were able to match all four expressions shown (happy, sad, angry, neutral). These results provide a strong case either for early competence in solving arbitrary PIP tasks, or for an early ability to detect invariant amodal visual and vocal affective information. It is difficult to clearly distinguish between the two, although other researchers (e.g., Soken & Pick, 1992; Walker, 1982) have shown that amodal cues play a role in this type of task. Walker (1982), for example, showed that the matching of vocal and facial expressions was inhibited in the 7-month-old group when visual stimuli were presented upside down. While amodal information such as temporal synchrony and rate information was still available, the emotional information proved difficult to perceive. More surprisingly, using an intermodal looking paradigm, Del Pino, Gogate, and Bahrick (1998) showed that 6.5- and 7.5-month-old infants were able to match their own voices with their own faces, even though the vocalizations and faces were not presented synchronously. Finally, Brookes et al. (2001) demonstrated the ability of 3-month-olds to learn face–voice associations, when, during familiarization trials, the relationships between the voices and faces were emphasized, notably by voice/lip synchrony, spatial co-location, and other procedural features.

Few cases of a non-ambiguous PIP task can be found in the literature. An example is the task used by Lyons-Ruth (1977). This author familiarized 3.5- to 4-month-old infants with an object that made a sound, and then presented the sound along with either the corresponding object or another object. The infants were found to spend more time looking at the non-corresponding object. A study of two groups of infants (4- and 6-month-olds) who had learned to associate a toy with a simultaneous sound showed that both age groups could properly perform the task (Fenwick & Morongiello, 1998). However, to succeed, the 6-month-olds, but not the 4-month-olds, needed amodal information (e.g., precise spatial co-location) in addition to the arbitrary information. A second example can be found in Patterson and Werker's (2002) Experiment 2. These authors tested the ability of 4.5- to 8-month-old infants to match two faces presented visually (one male, one female) and one voice. No evidence of gender matching between a face and a voice was found at 4.5 months of age, but the ability appeared in the 8-month-old group. More recently, Bahrick et al. (2005) conducted a study, demonstrating an early ability to detect arbitrary face–voice relationships in a parallel presentation by 4-month-old infants but not 2-month-olds. These authors also had the infants perform a task involving the recognition and memorization of arbitrary face–voice relationships presented in parallel. The ability to perform this task appears to emerge between 4 and 6 months of age.

Several authors have investigated a particularly interesting case of auditory–visual association learning, namely the acquisition of word–object associations. Werker et al. (1998) explored the ability of 14-month-old infants to perform an arbitrary PIP task. For the habituation phase, the infants were presented with a word–object pair. Once they had detected and learned the word–object association, they were presented with simultaneous word–object pairs, either the same as or different from the ones learned. In this experiment, 8- to 12-month-olds failed, but 14-month-olds demonstrated some ability to rapidly learn arbitrary auditory–visual associations.

Slater, Brown, and Badenoch (1997) reported that 2-day-old infants were able to learn arbitrary intermodal relations. The authors familiarized the newborns with auditory–visual associations, e.g., a vertical red line with the sound "mum" uttered by a male voice, and a diagonal green line and the word "teat" spoken by a female voice. The auditory and visual stimuli were presented simultaneously. After the familiarization phase, the same associations or new combinations were presented. Learning was assessed through the preference displayed for the new associations. Other demonstrations of arbitrary auditory–visual PIP by infants were reported by Gogate and Bahrick (1996): 7-month-old infants were able to learn arbitrary pairings of vowel sounds and objects when the stimuli were presented simultaneously. In another study Bahrick (2004) had 3.5-month-old infants map the pitch of a sound-making object to the shape/colour of that object, in one of three experimental conditions. But no evidence of mapping was found.

The results from this set of studies suggest that by an early age (3 or 4 months), infants can learn arbitrary relationships between stimuli presented simultaneously, such as a voice and a face, a vowel heard and a face seen, or a noise heard and made by a toy seen. But this ability is difficult to demonstrate, and it is not always clearly established whether the relationships used by the infants were solely arbitrary in nature.

Studies using amodal and serial presentation tasks

An example of a SIP task based on non-arbitrary relations is the one used in the study by Wagner, Winner, Cichetti, and Gardner (1981). The findings of this study showed that 6- to 14-month-old infants were able to match a visual event and an auditory event on the basis of relationships that were neither arbitrary, nor temporal, nor physical, but solely based on "metaphorical similarities". For example, the infants appeared to match a broken line with a pulsing tone and a continuous line with a continuous tone, and an upward arrow was matched with an ascending tone, and a downward arrow, with a descending tone.

Lewkowicz and Turkewitz (1980) studied visual preferences for lights of different intensities in newborns following exposure to bursts of sounds or white noises. The newborns without prior exposure to the sounds preferred intermediate intensity lights, whereas the infants with prior exposure looked more at the lowest intensity lights. The authors interpreted these preferences as evidence that newborns respond to quantitative variations in stimulations. From the above results, we can infer that newborns are sensitive to at least one type of amodal information to which they have serial access (i.e., intensity). In a similar experiment, Gardner, Lewkowicz, Rose, and Karmel (1986) first exposed newborns to a fixed number of sounds and visual stimuli differing in frequency (created by white fluorescent lamps and a noise generator emitting at frequencies of 2 Hz, 8 Hz or no sound at all). The participants were tested later with visual stimuli presented in pairs: for each trial, only one element of the pair had the same frequency as the pre-exposed stimulus. The newborns' visual preferences were influenced by prior exposure and sound frequency in all modalities. This experiment showed that exposure to a given auditory frequency affected subsequent preferences for certain visual frequencies. Gardner et al. (1986) concluded that a form of supramodal perception is present in human newborns, as previously suggested by Bower (1974) and Gibson and Walker (1984).

In another SIP task, Mendelson and Ferland (1982) played a 60-second recording of syllables repeated at two different rhythms (regular or irregular) to 4-month-olds. Immediately after this auditory presentation, the participants were shown a silent film clip with a puppet opening its mouth at either a "familiar" or "novel" rhythm. The participants reacted to the mismatch between the auditory and the visual rhythm by looking longer at the visual display. Mendelson and Ferland (1982) interpreted their results as a demonstration that the participants had perceived the link between the auditory and visual temporal information.

In another experiment, 6- to 9-month-old infants were tested for their ability to detect correspondences between visual and auditory stimuli (Starkey, Spelke, & Gelman, 1990). One of the experiments of these authors (Experiment 5) had three phases, a pretest, a familiarization phase, and a test. On the pretest, participants heard six sequences of sounds (differing in number) and their spontaneous preferences were assessed. The familiarization phase consisted of the visual presentation of slides with a given number of objects that infants are used to encountering in their daily lives. For the test phase, the same sequences of sounds as those used during the pretest were presented. The authors reported that the "infants looked longer towards the sound source when it produced numerically familiar sequences than when it produced numerically novel sequences" (Starkey, Spelke, & Gelman, 1990, pp. 118–119). It was concluded that infants are able to detect non-simultaneous intermodal correspondences between an initially presented set of auditory stimuli and its subsequent display in the visual modality. We consider this example and the others cited above as further evidence that infants can perform SIP tasks with amodal stimuli.

Studies using arbitrary and serial presentation tasks

Arbitrary and serial presentation tasks target the ability to detect an arbitrary relation between a stimulus perceived and memorized in one modality and a second stimulus from the same source accessed via another modality. It is important to note that the second stimulus is not present when the information is processed for the first stimulus.

The developing infant tends to increasingly take into account information that is more and more distant from his/her position in space. The infant's environment becomes richer and more complex, and arbitrary information separated in both time and space must be processed. This type of intermodal combination can be observed in a number of daily activities of the growing infant. Several contexts offer such opportunities for intersensory integration.

For example, it is assumed that infants have had ample opportunity to observe that objects falling to the ground make sounds and that the types of the sounds depend on the nature or composition of the object (Bahrick, 1987). Consider the following scene: an infant (in his/her mother's arms in the bathroom) hears the sound of a glass breaking coming from the kitchen. A few seconds later, the mother and the infant arrive in the kitchen. When the infant sees the broken pieces of glass, will he/she match the visual scene with the auditory memory of the sound of the breaking glass heard earlier? Referring to the intermodal infant literature, we find that the answer to this question is far from clear. As noted above, few studies have used matching procedures with arbitrary relations. This fact is even more striking with respect to serial presentations of stimuli, as in arbitrary SIP tasks. To our knowledge, no study in the field of auditory–visual integration has been conducted on this topic, although some investigations have been carried out and have demonstrated serial matching between auditory–visual stimuli by infants under conditioning techniques (e.g., Werker & Tees, 1983). Systematic studies are needed to investigate the way that infants spontaneously develop these abilities under natural conditions.

Speech comprehension can be seen as a case requiring amodal information, but also, and probably more importantly, the integration of arbitrary auditory–visual intermodal information. Word understanding involves matching a sound (the word) with the visually perceived object designated by the word, but there is no amodal property common to both the sound and the object (Geschwind, 1965).

The development of matching between auditory and visual information in infancy has been widely studied. Nonetheless, our review of the literature revealed an imbalance among the different types of procedures proposed in the various studies. This imbalance may mean that certain developmental issues involved in the interplay between intersensory matching abilities and other cognitive functions, such as those related to language acquisition, have been overlooked. Of the four types of procedures distinguished in our review, parallel presentation procedures appear to have been used more frequently than serial ones. Amodal relationships also appear more often than arbitrary ones. To our knowledge, there are no studies in the developmental literature on the use of arbitrary SIP procedures.

IMPLICATIONS OF THE TAXONOMY

The taxonomy of intermodal procedures and the review of most of the available literature in the field should improve our understanding of the development of auditory–visual mapping abilities in infancy. In particular, we believe that the new taxonomy will help in comparing the different studies in the literature.

In this section, we will first examine differences in the underlying cognitive processes used in PIP and SIP tasks and then the difference between arbitrary and amodal associations. The four types of tasks distinguished will be compared and ranked in order of difficulty.

In the literature on auditory–visual matching tasks, it is important to distinguish tasks that tap sensory memory (understood as a passive recording process; Estes, 1988) from those that imply active processing (attention, representation; Atkinson & Shiffrin, 1968). In an intermodal task that only requires the activation of sensory memory, both stimulus-related elements of information are available and matching can be performed immediately. However, in a task for which only partial information on an object or event is available, an amodal or plurimodal representation is required for matching to another representation later (i.e., the representation of the stimulus in the other modality). Thus, memory-storage's time constraints should play some role in the accuracy and success of intermodal matching when the two stimuli are not simultaneously present.

Unfortunately, very few studies are available on early features and age-related changes in sensory memory. In adults, iconic memory has a span of about 200–400 ms (Van der Heijden, 1981) and echoic memory has a span of 2 to 4 seconds for behavioural responses (Darwin, Turvey & Crowder, 1972). Concerning neural responses, the time span of auditory memory is shorter in newborns (about 800 ms) than in children or adults (Cheour et al., 2002). In children, sensory transmission reaches its adult form by the age of 8 (Leblanc, Muise, & Blanchard, 1992). When auditory information is presented less than 300 ms after visual information, the task is said to rely on a PIP procedure. In such cases, the visual stimulus probably does not need to be represented and stored in short- or long-term memory.

The rationale for stressing the distinction between PIP and SIP tasks is the idea that memory is not used in the same way for the two presentation conditions. In a typical SIP task, the mere fact that there is a time lapse before access to the second stimulus means that the information from the first stimulus has to be stored in memory. In a typical PIP situation, however, it can be hypothesized that when the two stimuli can be accessed simultaneously; there will be a reduction in the memory load or no memory load at all. In a PIP task, the co-occurrence (without co-location) of two stimuli helps the participant establish the necessary link between the two elements of information. This situation probably occurs most frequently under natural conditions. But in a SIP situation, because of the time lapse, the link between the two stimuli has to be discovered or constructed by the individual.

Tasks based on SIP procedures are therefore likely to involve deeper cognitive processing than PIP tasks. Given that in a PIP task, the infant can directly and simultaneously perceive both stimuli, we can consider that intermodal matching is facilitated. But in tasks requiring a SIP procedure, the infant has to memorize the stimulus available in the first modality, and, after a variable time lapse, compare the mnemonic trace of the previous stimulus to the new stimulus presented in the new modality. This difference in cognitive processing between the two categories of tasks leads us to postulate that intermodal matching should emerge later in infancy for SIP procedures than it does for PIP procedures. It can therefore be argued that tasks using SIP procedures may be more complex than tasks using PIP procedures.

The many studies involving amodal and arbitrary SIP procedures provide strong evidence of an early ability to match auditory and visual stimuli when the stimuli are presented simultaneously. Other studies have provided evidence of early intermodal matching in the absence of temporal synchrony. In the category of studies reviewed and labelled as using SIP procedures or tasks, the auditory and visual stimuli are presented in succession. But we noticed a marked imbalance in the literature, with a limited number of SIP studies conducted with young participants. This phenomenon may be due to the difficulty very young infants have in performing such tasks.

Among the various procedures reviewed, amodal PIP procedures have the most critical amodal feature, i.e., temporal synchrony, for providing the integration of auditory and visual information (see the model proposed by Lewkowicz, 2000). Other properties of amodal PIP procedures are numerosity, rate, spatial co-location, etc., which vary across studies and task demands. Experiments involving amodal PIP relationships have amply demonstrated the early emergence of abilities that rely on such procedures, although these abilities appear somewhat later than the ability to match stimuli on the basis of an amodal relation (at about 4 months of age vs. at birth, but see Slater, Brown, & Badenoch, 1997, for an exception).

Based on a similar line of reasoning, we can consider amodal relationships as being recognized earlier than arbitrary relationships in the development of the infant. This hypothesis has been formulated in several studies of auditory–visual or other modalities (Bahrick, 1994, 2004; Gogate & Bahrick, 1998; Hernandez-Reif & Bahrick, 2001; Slater et al., 1999). This earlier sensitivity in the processing of amodal relationships led Lickliter and Bahrick (2001) and Bahrick and Lickliter (2002) to develop the "intersensory redundancy hypothesis", which focuses on the effect of redundant amodal information on perceptual development in infancy (Bahrick, Flom, & Lickliter, 2002; Bahrick, Lickliter, & Flom, 2004).

Thus, given that the differences in the procedures presented in this section involve differences in the cognitive processing used by infants, we propose the following correspondence between these kinds of tasks and their underlying processes: (a) tasks with parallel presentation of stimuli linked by an amodal relation could involve what is usually labelled "intermodal perception"; (b) tasks with serial presentation of stimuli linked by an amodal relationship could involve what is usually called "intermodal matching or transfer"; and (c) tasks based on stimuli linked by a non-modality-specific relationship, either parallel or serial presentation, would correspond to what is usually labelled "intermodal matching".

In the following section, we will examine a possible developmental hierarchy for infant success on these different kinds of tasks.

If the distinctions we have outlined above have some relevance, it would be interesting to find out if there is a developmental hierarchy in the abilities tested by the different procedures. The four proposed categories could thus be ranked according to their degree of difficulty, which probably depends on the amount of amodal information available in the experimental context or the environment. Amodal PIP tasks could be seen as the simplest of the four categories, because the redundancy of amodal information should facilitate intermodal matching. It could also be hypothesized that arbitrary PIP tasks are easier to solve than amodal SIP tasks, since, in the former case, the infant can rely on at least one amodal cue, i.e., the temporal synchrony of the stimuli. In an amodal SIP task, amodal information is obviously accessible (e.g., numerosity), but is less accessible than the information provided concurrently in time in an amodal PIP task. Consequently, SIP tasks should be more difficult than PIP tasks. The most difficult tasks should be arbitrary SIP tasks for which amodal information is minimal. In such cases, the infant has to perform some form of categorical or semantic processing to establish the intermodal relationships between the two types of information. The task-difficulty order proposed above needs to be confirmed, or invalidated, through the accumulation of new empirical findings.

CONCLUSION

The taxonomy that we have proposed in the present paper could help clarify discussions on the relations between language and intermodal auditory–visual matching. We have mentioned some arguments in favour of a close link between the two abilities. Recent experimental approaches indicate that forms of intermodal matching appear very early in infants and that these capacities are completely independent of language. However, we have shown that different levels of intermodal matching must be considered and that their respective relations with language must be analysed carefully.

As explained above, tasks evaluated via an arbitrary serial presentation mode should be the most appropriate ones for assessing early linguistic abilities such as word comprehension. Experimental investigations using SIP procedures should thus provide useful information on the age at which infants efficiently perform these tasks. This type of knowledge would obviously provide insight into the role of auditory – visual matching abilities in the emergence of language (and vice versa).

It would also be valuable to have future studies explore the ability of infants to perform arbitrary SIP tasks, and to compare their results with the performance levels reported for other types of auditory – visual matching tasks (e.g., arbitrary PIP procedures).

The benefit of our approach could be twofold. First, the new taxonomy proposed has the advantage of providing a common framework for viewing the various intermodal tasks used in infant research, along with their perceptual and cognitive constraints. Second, the approach should help improve our understanding of the cognitive requirements (e.g., mastery of arbitrary relations, representational and mnemonic levels) for speech and language acquisition in general.

Manuscript received 14 February 2005
Revised manuscript received 29 March 2006
First published online 12 July 2007

REFERENCES

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. Spence & J. Spence (Eds.), *The psychology of learning and motivation: II* (pp. 89 – 195). New York: Academic Press.

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

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

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

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

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

Bahrick, L. E. (2001). Increasing specificity in perceptual development: Infants' detection of nested levels of multimodal stimulation. *Journal of Experimental Child Psychology*, *79*, 253 – 270.

Bahrick, L. E. (2004). The development of perception in a multimodal event. In G. Bremner & A. Slater (Eds.), *Theories of infant development* (pp. 90 – 120). Oxford, UK: Blackwell.

Bahrick, L. E., Flom, R., & Lickliter, R. (2002). Intersensory redundancy facilitates discrimination of tempo in 3-month-old infants. *Developmental Psychology*, *41*, 352 – 363.

Bahrick, L. E., Hernandez-Reif, M., & Flom, R. (2005). The development of infant learning about specific face – voice relations. *Developmental Psychology*, *41*, 541 – 552.

Bahrick, L. E., & Lickliter, R. (2002). Intersensory redundancy guides early perceptual and cognitive development. In R. Kail (Ed.), *Advances in child development and behavior* (Vol. 30, pp. 153 – 187). New York: Academic Press.

Bahrick, L. E., Lickliter, R., & Flom, R. (2004). Intersensory redundancy guides the development of selective attention, perception, and cognition in infancy. *Current Directions in Psychological Science*, *13*, 99–102.

Blank, M., & Bridger, W. H. (1964). Cross-modal transfer in nursery-school children. *Journal of Comparative and Physiological Psychology*, *58*, 277–282.

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

Brookes, H., Slater, A., Quinn, P. C., Lewkowicz, D., Hayes, R., & Brown, E. (2001). Three-month-old infants learn arbitrary auditory–visual pairings between voices and faces. *Infant and Child Development*, *10*, 75–82.

Bushnell, E. W. (1994). A dual-processing approach to cross-modal matching: Implications for development. In D. J. Lewkowicz & R. Lickliter (Eds.), *The development of intersensory perception: Comparative perspectives* (pp. 19–38) Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Butterworth, G. (2001). Joint visual attention in infancy. In G. Bremner & A. Fogel (Eds.), *Handbook of developmental psychology* (pp. 213–240). Malden, MA: Blackwell.

Butterworth, G., & Castillo, M. (1976). Co-ordination of auditory and visual space in newborn human infants. *Perception*, *5*, 155–160.

Chen, X., Striano, T., & Rakoczy, H. (2004). Auditory–oral matching behavior in newborns. *Developmental Science*, *7*, 42–47.

Cheour, M., Ceponiene, R., Leppanen, P., Alho, K., Kujala, T., Renlund, M. et al. (2002). The auditory sensory memory trace decays rapidly in newborns. *Scandinavian Journal of Psychology*, *43*, 33–39.

Darwin, C. J., Turvey, M. T., & Crowder, R. C. (1972). An auditory analogue of the Sperling partial report procedure: Evidence for brief auditory storage. *Cognitive Psychology*, *3*, 255–267.

Del Pino, N., Gogate, L., & Bahrick, L. E. (1998). Intermodal self-perception: Infants match their voices with their faces. *Infant Behavior and Development*, *21*, 375.

Estes, W. K. (1988). Human learning and memory. In R. C. Atkinson et al. (Eds.), *Stevens' handbook of experimental psychology: Vol. 2. Learning and cognition* (2nd ed., pp. 351–415). Chichester, UK: Wiley.

Fenwick, K. D., & Morongiello, B. A. (1998). Spatial co-location and infants' learning of auditory–visual associations. *Infant Behavior and Development*, *21*, 745–760.

Field, J., DiFranco, D., Dodwell, P., & Muir, D. W. (1979). Auditory–visual coordination in 2–12-month-old infants. *Infant Behavior and Development*, *2*, 113–122.

Gardner, J. M., Lewkowicz, D. J., Rose, S. A., & Karmel, B. Z. (1986). Effects of visual and auditory stimulation on subsequent visual preferences in neonates. *International Journal of Behavioral Development*, *9*, 251–263.

Geschwind, N. (1965). Disconnection syndrome in animals and man. *Brain*, *88*, 237–294.

Gibson, E. J., & Walker, A. S. (1984). Development of knowledge of visual–tactual affordances of substance. *Child Development*, *55*, 453–460.

Gogate, L. J., & Bahrick, L. E. (1996). Learning of arbitrary vowel–object relations by 7-month-old infants. *Infant Behavior and Development*, *19*, 479.

Gogate, L. J., & Bahrick, L. E. (1998). Intersensory redundancy and 7-month-old infants' memory for arbitrary syllable–object relations. *Infancy*, *2*, 219–231.

Gogate, L. J., Walker-Andrews, A. S., & Bahrick, L. E. (2001). The intersensory origins of word comprehension: An ecological-dynamic systems view. *Developmental Science*, *4*, 1–37.

Gottfried, A. W., Rose, S. A., & Bridger, W. H. (1977). Cross-modal transfer in human infants. *Child Development*, *48*, 118–123.

Hashiya, K. (1999). Auditory visual intermodal recognition of conspecifics by a chimpanzee (*Pan troglodytes*). *Primate Research*, *15*, 333–342.

Hashiya, K., & Kojima, S. (1997). Auditory visual intermodal matching by a chimpanzee (*Pan troglodytes*). *Japanese Psychological Research*, *39*, 182–190.

Hashiya, K., & Kojima, S. (2001). Acquisition of auditory–visual intermodal matching-to-sample by a chimpanzee (*Pan troglodytes*): Comparison with visual–visual intramodal matching. *Animal Cognition*, *4*, 231–239.

Hernandez-Reif, M., & Bahrick, L. E. (2001). The development of visual–tactile perception of objects: Amodal relations guide learning about arbitrary relations. *Infancy*, *2*, 51–72.

Hirsh-Pasek, K., & Gollinkoff, R. M. (1996). *The origin of grammar*. Cambridge, MA: MIT Press.

Houston-Price, C., Plunkett, K., & Harris, P. (2005). Word-learning wizardry at 1;6. *Journal of Child Language*, *32*, 175–189.

Humphrey, K., Tees, R. C., & Werker, J. F. (1979). Auditory–visual integration of temporal relations in infants. *Canadian Journal of Psychology*, *33*, 347–352.

Iverson, J. M., Capirci, O., Longobardi, E., & Caselli, M. C. (1999). Gesturing in mother–child interactions. *Cognitive Development*, *14*, 57–75.

Kaye, K. L., & Bower, T. G. R. (1994). Learning and intermodal transfer of information in newborns. *Psychological Science*, *5*, 286–288.

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. *Infant Behavior and Development*, *7*, 361–381.

Kuhl, P. K., Williams, K., & Meltzoff, A. N. (1991). Cross-modal speech perception in adults and infants using nonspeech auditory stimuli. *Journal of Experimental Psychology: Human Perception and Performance*, *17*, 829–840.

Lawson, K. R. (1980). Spatial and temporal congruity and auditory–visual integration in infants. *Developmental Psychology*, *16*, 185–192.

Lawson, K. R., & Turkewitz, G. (1980). Intersensory function in newborns: Effect of sound on visual preferences. *Child Development*, *51*, 1295–1298.

Leblanc, R. S., Muise, J. S., & Blanchard, L. (1992). Backward visual masking in children and adolescents: Sensory transmission, accrual rate and asymptotic performance. *Journal of Experimental Child Psychology*, *53*, 105–114.

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

Lewkowicz, D. J. (2000). The development of intersensory temporal perception: An epigenetic systems/limitations view. *Psychological Bulletin*, *126*, 281–308.

Lewkowicz, D. J., & Lickliter, R. (Eds.). (1994). *The development of intersensory perception: Comparative perspectives*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

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

Lickliter, R., & Bahrick, L. E. (2001). The salience of multimodal sensory stimulation in early development: Implications for the issue of ecological validity. *Infancy*, *2*, 451–467.

Lyons-Ruth, K. (1977). Bimodal perception in infancy: Response to auditory–visual incongruity. *Child Development*, *48*, 820–827.

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

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

McGurk, H., Turnure, C., & Creighton, S. J. (1977). Auditory–visual coordination in neonates. *Child Development*, *48*, 138–143.

Meltzoff, A. N., & Borton, R. W. (1979). Intermodal matching by human neonates. *Nature*, *282*, 403–404.

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* (pp. 335–370). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
Mendelson, M. J., & Ferland, M. B. (1982). Auditory–visual transfer in four-month-old infants. *Child Development*, *53*, 1022–1027.
Molfese, D. L., Morse, P. A., & Peters, C. J. (1990). Auditory evoked responses to name four different objects: Cross-modal processing as a basis for infant language acquisition. *Developmental Psychology*, *26*, 780–795.
Morongiello, B. A., Fenwick, K. D., & Chance, G. (1998). Cross-modal learning in newborn infants: Inferences about properties of auditory–visual events. *Infant Behavior and Development*, *21*, 543–554.
Morongiello, B. A., Fenwick, K. D., & Nutley, T. (1998). Developmental changes in associations between auditory–visual events. *Infant Behavior and Development*, *21*, 613–626.
Muir, D., & Field, J. (1979). Newborn infants orient to sounds. *Child Development*, *50*, 431–436.
Muir, D., & Field, J. (1980). Newborn infants orient to sounds. *Annual Progress in Child Psychiatry & Child Development*, *1980*, 85–94.
Oviatt, S. L. (1980). The emerging ability to comprehend language: An experimental approach. *Child Development*, *51*, 97–106.
Patterson, M. L., & Werker, J. F. (2002). Infants' ability to match dynamic and gender information in the face and voice. *Journal of Experimental Child Psychology*, *81*, 93–115.
Pickens, J., Field, T., Nawrocki, T., Martinez, A., Soutullo, D., & Gonzalez, J. (1994). Full-term and preterm infants' perception of face–voice synchrony. *Infant Behavior and Development*, *17*, 447–455.
Radeau, M. (1994). Auditory–visual spatial interaction and modularity. *Current Psychology of Cognition*, *13*, 3–51.
Roy, D. K., & Pentland, A. L. (2002). Learning from sights and sounds: A computational model. *Cognitive Science*, *26*, 113–146.
Savage-Rumbaugh, S., Sevcik, R. A., & Hopkins, W. D. (1988). Symbolic cross-modal transfer in two species of chimpanzees. *Child Development*, *59*, 617–625.
Slater, A., Brown, E., & Badenoch, M. (1997). Intermodal perception at birth: Newborn infants' memory for arbitrary auditory–visual pairings. *Early Development and Parenting*, *6*, 99–104.
Slater, A., Quinn, P. C., Brown, E., & Hayes, R. (1999). Intermodal perception at birth: Intersensory redundancy guides newborn infants' learning of arbitrary auditory–visual pairings. *Developmental Science*, *2*, 333–338.
Soken, N. H., & Pick, A. D. (1992). Intermodal perception of happy and angry expressive behaviors by seven month old infants. *Child Development*, *63*, 787–795.
Spelke, E. S. (1976). Infants' intermodal perception of events. *Cognitive Psychology*, *8*, 553–560.
Spelke, E. S. (1979). Perceiving bimodally specified events in infancy. *Developmental Psychology*, *15*, 626–636.
Spelke, E. S. (1981). The infant's acquisition of knowledge of bimodally specified events. *Journal of Experimental Child Psychology*, *31*, 279–299.
Spelke, E. S., Born, W. S., & Chu, F. (1983). Perception of moving, sounding objects by four-month-old infants. *Perception*, *12*, 719–732.
Spelke, E. S., & Owsley, C. J. (1979). Intermodal exploration and knowledge in infancy. *Infant Behavior and Development*, *2*, 13–27.

Starkey, P., Spelke, E. S., & Gelman, R. (1983). Toward a comparative psychology of number. *Cognition*, *39*, 171–172.

Starkey, P., Spelke, E. S., & Gelman, R. (1990). Numerical abstraction by human infants. *Cognition*, *36*, 97–127.

Streri, A. (2000). Pour une certaine unité des sens à la naissance. *Devenir*, *12*, 59–79.

Streri, A., & Gentaz, E. (2003). Cross-modal recognition of shape from hand to eyes in human newborns. *Somatosensory and Motor Research*, *20*, 11–16.

Van der Heijden, A. H. C. (1981). *Short-term visual information forgetting*. London: Routledge.

Wagner, S., Winner, E., Cichetti, D., & Gardner, H. (1981). "Metaphorical" mapping in human infants. *Child Development*, *52*, 728–731.

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

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

Walker-Andrews, A. S. (1988). Infants' perception of the affordances of expressive behaviors. *Advances in Infancy Research*, *5*, 173–221.

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

Werker, J. F., & Tees, R. C. (1983). Developmental changes across childhood in the perception of non-native speech sounds. *Canadian Journal of Psychology*, *37*, 278–286.

Werker, J. F., Cohen, L. B., Lloyd, V. L., Casasola, M., & Stager, C. L. (1998). Acquisition of word–object associations by 14-month-old infants. *Developmental Psychology*, *34*, 1289–1309.

Wertheimer, M. (1961). Psychomotor co-ordination of auditory and visual space at birth. *Science*, *134*, 1692.

Zukow-Goldring, P. (1997). A social ecological realist approach to the emergence of the lexicon: Educating attention to amodal invariants in gesture and speech. In C. Dent-Read, & P. Zukow-Goldring (Eds.), *Evolving explanations of development: Ecological approaches to organism–environment systems* (pp. 199–250). Washington, DC: American Psychological Association.