Author's personal copy

# Intermodal Perception☆

**Lorraine E Bahrick,** Florida International University, Miami, FL, United States
**George Hollich,** Purdue University, West Lafayette, IN, United States

© 2017 Elsevier Inc. All rights reserved.



## Glossary

**Amodal** Information that is not specific to a particular sense modality, but is perceivable and completely redundant across two or more senses. For example, rhythm and tempo are redundant across vision and audition.

**Bimodal** Occurring in two sense modalities.

**Increasing specificity** Progressive differentiation of finer levels of stimulation as a result of perceptual experience.

**Intersensory** Across the senses; intermodal.

**Intersensory redundancy** The co-occurrence of amodal information (e.g., rhythm, intensity changes) across two or more sense modalities; this information is highly salient to infants.

**Invariant** A property or relation that remains constant across transformation.

**Localization** Using auditory, visual, or tactile information to determine the location in space of a speaker or object.

**Modality-specific** Information, such as color or timbre, that can be perceived through only one sense modality.

**Multimodal** Information that can be experienced through multiple-sense modalities.

**Perceptual narrowing** Progressive improvements in perceptual discrimination as a function of experience in a given domain and progressive decline in perceptual discrimination in related domains to which we are not exposed.

**Proprioception** Information about self-movement based on feedback from the muscles, joints, and vestibular system.

☆*Change History*: November 2015. LE Bahrick and G Hollich Updated acknowledgement, sections, Figures, Further Reading, Relevant Websites, and bibliographies.



Reference Module in Neuroscience and Biobehavioral Psychology, 2017, 1–15

Author's personal copy



**Salience hierarchy** Attention hierarchies that dictate what information is detected and processed first versus later in exploratory time. The most salient, or attention-getting information is processed first, the information that is next most salient is processed next, and so forth.

**Speech segmentation** Extracting words or other important units from the fluent stream of speech.

**Temporal synchrony** Changes in events (e.g., from different senses) that occur at the same moment in time. For example, there is temporal synchrony between the movements of the lips and sounds of speech in natural audiovisual speech.

## Introduction

Intermodal perception, the perception of unitary objects and events from concurrent stimulation to multiple senses, is fundamental to early development. Early sensitivity to temporal, spatial, and intensity patterns of events ("amodal" information) that are redundant across stimulation to different senses, guides infants' perceptual, cognitive, and social development. Intermodal perception develops rapidly across infancy. Even very young infants are sensitive to amodal information, allowing them to perceive unitary multimodal events by linking sights and sounds of speech, emotional expressions, and objects, as well as information across visual and tactile, olfactory, and proprioceptive stimulation. Perceptual development proceeds along a path of differentiation of increasingly more specific levels of stimulation and increasing economy of information pick up with experience.

## Historical Conceptions of Intermodal Perception: The Binding Problem

The world provides a richly structured, continuously changing array of stimulation to all the senses. Our senses provide overlapping and redundant information for objects and events in the environment. For example, we can both see the visual motion and hear the rhythmic sounds of a hammer tapping; the movements of our lips and faces correspond closely to the sounds of speech. How are these sights and sounds connected? Despite the fact that information about the world arrives through distinct sensory channels, we perceive a stable world of unitary objects and events rather than separate sights, sounds, and tactile impressions. How do we accomplish this? Why do we not perceive a "blooming buzzing confusion" of disconnected, continuously changing patterns of stimulation from the different senses, as suggested by William James in 1890?

Broadly speaking, there have been two different approaches to explaining the origins and development of sensory integration. From very early in development, we perceive information from the sights, sounds, and tactile impressions of a single event as either 1) integrated across the senses and we must later learn to separate it, or 2) we perceive it as separate and we must later learn to integrate it. Both approaches have had their proponents. Dating as far back as the time of Aristotle, philosophers and scientists have been intrigued and puzzled by the specificity of the different senses and the overlap among them. Aristotle postulated a "*sensus communis*" that allowed us to perceive the qualities of stimulation that were common across different senses ("common sensibles") such as rest, magnitude, number, form, and unity. These properties are remarkably similar to those that scientists today view as common across the senses ("amodal"), serving as the basis for perceiving unitary objects and events.

Centuries later, philosophers such as Locke and Berkeley promoted a different view of intermodal perception. They argued that we must learn to integrate and interpret sensations across separate sensory channels before meaningful perception of objects and events was possible. This "constructivist" view provided the foundation for many modern theories of perception. According to this view, sensations from different receptors must be integrated and organized in the brain. This view, in turn, posed a "binding" problem for perception in that mechanisms must be discovered that translate information from different sensory codes into a common language. The constructivist view dominated thinking about the development of perception through most of the twentieth century, including the views of Piaget, a pioneer in theories of cognitive development. According to Piaget, infants must learn to integrate and coordinate information across the senses through a gradual process of association across development. Integration was thought to occur through interacting with objects and experiencing concurrent stimulation from different senses and associating, assimilating, or calibrating the different senses to one another. From this perspective, not until after the first half year of life did infants begin to integrate touch and vision. Prior to this activity-based achievement, infants perceived a world of unrelated tactile impressions and visual images that would shrink and expand, appear and disappear capriciously.

Not until 1966 when J.J. Gibson published his seminal work on the "ecological" view of perception, was the constructivist view seriously questioned. In a clear departure from traditional views, Gibson proposed that the existence of different forms of sensory stimulation was not a problem for perception. Rather, it provided an important basis for our ability to perceive coherent multimodal objects and events. He asserted that the sensory systems work in concert, as a unified perceptual system, to detect "invariant" aspects of stimulation. A critical type of invariant for intermodal perception is "amodal" information. Amodal information is information that is common across more than one sense, similar to the concept of the *sensus communis*

Author's personal copy



formulated by Aristotle. Temporal and spatial aspects of stimulation are typically conveyed by multiple senses and are fundamental dimensions of amodal information. For example, the sights and sounds of hands clapping are temporally synchronous, share a common rate and rhythm, and occur across a common spatial location. Thus, detection of amodal information provides a solution to the age-old binding problem.

We now know from a large body of research on the development of intermodal perception generated since the 1970s that young infants detect a wide array of amodal properties of events across visual, acoustic, tactile, and proprioceptive stimulation. Inspired in large part by Gibson's ecological approach to perception, we have discovered that the ability to detect amodal information is a foundation for perceiving meaningful multimodal events in the first months of life. This ability provides a radical and efficient solution to the binding problem created by the constructivist views. That is, if infants detect amodal information in early development, then there is no need to integrate and coordinate separate sources of sensory information. Rather, by detecting information that is common to multiple senses, a naïve perceiver can explore unitary events in a coordinated manner, without binding.

In the next sections, we outline some terms and definitions relating to intermodal perception as well as discuss some key principles of intermodal development. We then discuss fetal and neural development and their relation to these key principles. Next we describe what is known about the development of infants' sensitivity to audiovisual, visual-tactile, visual-motor stimulation, and other sensory combinations as well as how each of these types of information provides the basis for infants to learn about different aspects of the environment.

## Key Terms and Definitions

### What Is Intermodal Perception?

Intermodal perception (also called intersensory or multimodal perception) is the perception of unitary objects or events that make information simultaneously available to more than one sense. For example, a bouncing ball can be seen, heard, and touched. This information is typically spatially collocated and temporally coordinated. The sights, sounds, and tactile impressions come from the same location and share the same temporal pattern. Since most objects and events are experienced through multiple senses, most everyday perception is intermodal.

### Amodal Versus Modality Specific Information

Objects and events make two different types of information available, amodal and modality-specific. Amodal information, as discussed above, is information that can be conveyed through more than one sense and is not specific to a particular sense modality. It is information that is redundant across different senses. For example, the rhythm and tempo of the bouncing ball can be detected visually or acoustically. Amodal information can be characterized along three primary dimensions: time, space, and intensity. Since all events occur over time and space and have a characteristic intensity pattern, all events provide amodal information.

Sensitivity to amodal information is fundamental for the development of event perception. Consistent with Gibson's ecological approach perceiving temporal synchrony, common rhythm, tempo, or intensity patterns between the sights and sounds of the bouncing ball allows the infant to experience a unified, multimodal event without intermodal knowledge to guide this process.

In contrast to amodal information, all events also provide modality-specific information. Modality-specific information is information that can be perceived only through a particular sense modality. For example, color and pattern can be perceived only visually, timbre only acoustically, and temperature only through touch. Perception of modality-specific information allows us to differentiate between a blue and a red ball or between the faces or voices of two different individuals.

### Intersensory Redundancy

When the same amodal information (such as rhythm, tempo, or intensity patterns) is simultaneously available through different senses, this is called "intersensory redundancy." In audiovisual stimulation, intersensory redundancy entails synchronous, collocated sights and sounds, as in the natural stimulation from a bouncing ball or clapping hands. Intersensory redundancy is highly salient to infants, both human and animal. This redundancy can capture and direct attentional selectivity at the expense of other information. For example, in exploring an event, a young infant might notice the redundant aspects such as its rhythm and tempo, but not the appearance of the object or the specific nature of its sounds. This redundancy is so salient that young infants can use it to select one of two superimposed movies while ignoring the other. When the soundtrack to one event, such as a hand-clapping game, is turned on, young infants can selectively follow the flow of action, even when it is superimposed with another similar silent event such as a hand striking the keys of a xylophone (see Fig. 1).

Thus, in early development we perceive both amodal and modality-specific information. Amodal information is typically more salient, particularly when it is detected redundantly across two or more senses, and plays an important role in guiding perceptual development. The nature of its role in perceptual development is described in more detail in the section ahead.

Author's personal copy



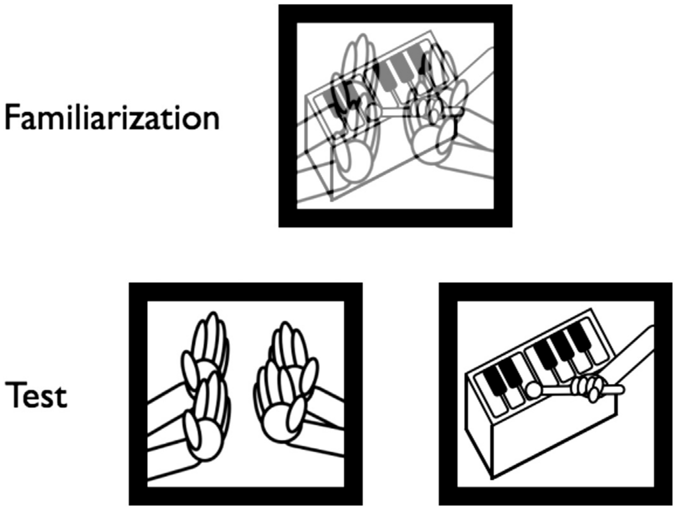


**Figure 1** Superimposed movies presented to 4-month-old infants. During familiarization, when the soundtrack to one event was turned on, creating intersensory redundancy, infants were able to selectively follow one event and ignore the other. During the test trials, infants then watched the "novel" unattended event. Adapted from Bahrick, L.E., Walker, A.S., Neisser, U., 1981. Selective looking by infants. Cogn. Psychol. 13, 377–390.

## Key Principles of Intermodal Development: A Foundation for Early Perceptual Development

### Sensitivity to Amodal, Redundant Information Guides Perceptual Development

Our senses provide us with a constantly changing flow of multimodal stimulation. How do we learn to select and attend to relevant stimulation and ignore the vast amount of stimulation that is irrelevant to our needs, goals, and actions? Moreover, how do infants, with no knowledge of the world, determine which sights, sounds, and tactile impressions belong together and constitute coherent events, and which are unrelated and should be ignored? With a limited attentional capacity we can selectively attend to only a small portion of the stimulation available, while the vast majority of concurrent stimulation remains in the background. For decades, scientists and philosophers have puzzled over how this process is guided and develops into economical patterns of attention and perception that are increasingly more in line with that of adult perceivers.

Perception of redundant, amodal information across the senses provides one answer to these questions. That our senses provide redundant information about objects and events is no excess of nature. Instead, this highly salient intersensory redundancy is fundamental for guiding the development of attention, perception and cognition in infancy. It provides a reliable basis for focusing early selective attention on patterns of stimulation that belong together and constitute unitary events and for ignoring stimulation from unrelated events. For example, if infants focus on the redundancy between faces and voices during speech such as the synchrony, rhythm, tempo, and intensity changes common to the voice and movements of the face/lips, this ensures that infants attend to stimulation from unified events (a person speaking) and not to concurrent sounds or the movements of an unrelated person or object nearby. A large body of infant research (reviewed briefly in the sections ahead) has accrued since the 1970s and reveals that early perception is quite organized and there is no evidence that infants experience the blooming, buzzing confusion postulated by William James more than 100 years ago. Sensitivity to redundancy across the senses thus promotes attention to unified events in the presence of competing sounds and motions. Once unitary events are differentiated from the flux of multimodal stimulation, the process of perception and learning can proceed in a meaningful way.

### The Intersensory Redundancy Hypothesis

The "intersensory redundancy hypothesis" (IRH), a theory of selective attention, describes how attention to redundancy drives and organizes the development of intermodal and unimodal perception of objects and events across infancy and childhood, and in turn, provides a solid foundation for more complex skills of social, language, and cognitive functioning. The IRH begins with the well established finding that information that is redundant across the senses is highly salient to young organisms—both humans and animals. It then describes the consequences of this attentional salience for perception of objects and events across development. Detection of intersensory redundancy has both a facilitating and an interfering effect on perception of different types of information. It facilitates attention to amodal information and impairs attention to modality-specific information, and these effects are most evident in early development. In other words, redundancy across the senses causes amodal information to become "foreground" (facilitating effect) and other aspects of stimulation (i.e. modality-specific information) to become "background" (imparing effect). Sensitivity to this redundancy allows infants to follow one video event by hearing its synchronous soundtrack and selectively attending to the common tempo, rhythm and intensity changes across sight and sound, even when the event is superimposed with another video event (as seen in Fig. 1). In contrast, when stimulation is

Author's personal copy



unimodal (visual or auditory only), no highly salient redundancy is available to "capture attention," and attention is then promoted to modality-specific aspects of stimulation such as color, pattern, pitch, or timbre. Thus, the salience of intersensory redundancy gives rise to "attentional salience hierarchies" which guide and organize early attention and perception in a meaningful way. Salience hierarchies describe what information is detected first versus later in exploratory time and across development. They exert the most influence when attention and processing resources are most limited, such as in early development, or in later development when tasks are difficult in relation to the skills of the perceiver.

The IRH consists of four basic principles. The first two are illustrated in Fig. 2. 1) *Intersensory facilitation of amodal properties*: Amodal properties such as rhythm, tempo, and intensity are detected more easily and earlier in development when they are detected redundantly in multimodal stimulation than in unimodal stimulation (Fig. 2, upper and lower left quadrants). Only after first detecting salient amodal information such as the rhythm of a ball bouncing, would young infants then detect less salient (modality-specific) information such as the color or shape of the ball. In contrast, during unimodal exploration, such as watching a silent event or speaking on the telephone, attention is not captured by salient intersensory redundancy. 2) *Unimodal facilitation of modality-specific properties*: Modality-specific properties such as pitch, timbre, color, and visual pattern are detected more easily and earlier in development when they are detected in unimodal stimulation than in multimodal stimulation (Fig. 2, lower and upper right quadrants). Only after first detecting pitch and timbre of a person's voice on the phone, would infants then detect less salient amodal information such as it's rhythm and tempo. 3) *Developmental improvement*: with increasing efficiency of attention across development, infant attention moves through the salience hierarchies more quickly and infants can detect both amodal and modality-specific properties in both unimodal as well as multimodal stimulation. For example, older infants may now detect both the rhythm of the ball bouncing as well as its color and shape in a single episode of exploration. 4) *Task difficulty*: Intersensory and unimodal facilitation are most evident for tasks of relatively high difficulty in relation to the skills of the perceiver. Thus, both intersensory and unimodal facilitation should be evident across the life span for tasks that are difficult.

A variety of research findings with infants support these principles. Young infants show evidence of intersensory facilitation (principle 1) across a wide range of events. In the earliest studies illustrating intersensory facilitation, five-month-olds were shown a toy hammer tapping in a rhythmic pattern. Infants were able to detect a change in the rhythm and a change in the tempo of the hammer if they experienced the event bimodally (they could both see and hear the synchronous tapping of the hammer). However, they were unable to detect the change in rhythm or tempo if they experienced the event unimodally (they could either hear or see the hammer) or asynchronously (the movements of the hammer were out of synchrony with the sounds). It was not until a few months later, when attention became more flexible, that infants were able to perceive the change in rhythm and tempo in unimodal auditory or visual stimulation alone (evidence of principle 3, developmental improvement). A similar developmental pattern was found in a study of infant sensitivity to serial order information. When three distinctive objects, each with a distinctive sound, appeared in sequence, young infants detected the order of their occurrence when they were seen and heard together, but not when they were seen or heard alone. Only older infants detected serial order information in unimodal visual or unimodal auditory stimulation. Similar findings were observed for infant perception of emotion (see Social Development ahead), qualities that are amodally specified by a blend of tempo, rhythm and intensity information. Thus, perception of amodal information such as rhythm, tempo, emotion, and serial order information initially develops in the context of multimodal stimulation, and later in development, perception is extended to unimodal stimulation.

Young infants also show evidence of unimodal facilitation, principle 2, and developmental improvement for this principle. In contrast to perception or redundant information, perception of nonredundant modality-specific information (specific to a single sense modality, such as the pitch of a voice, the configuration of a face, the color or orientation of an object) develops first in

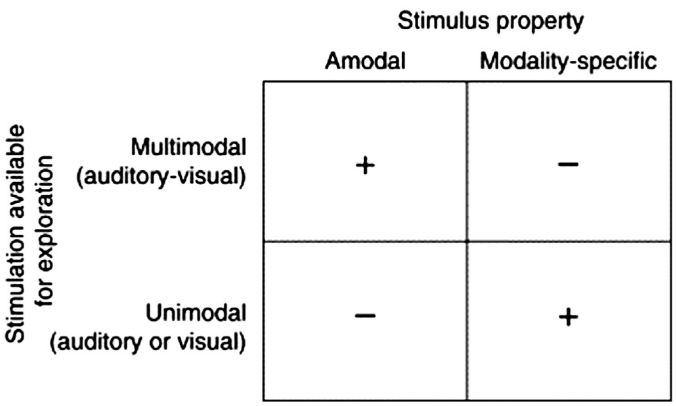


**Figure 2** Principles 1 and 2 of the intersensory redundancy hypothesis (Intersensory and unimodal facilitation): stimulus properties (amodal vs. modality-specific) and the type of stimulation available for exploration (multimodal vs. unimodal) together determine whether attention and perceptual processing of a stimulus property are facilitated (+) or inhibited (−). From Bahrick, L.E., Lickliter, R., 2002. Intersensory redundancy guides early perceptual and cognitive development. In: Kail, R., (Ed.), Advances in Child Development and Behavior, vol. 30. Academic Press, New York, p. 166.

Author's personal copy



unimodal stimulation and is later extended to multimodal stimulation. For example, 5-month-old infants who were shown a toy hammer tapping downward versus upward were able to detect a change in the orientation of the hammer's motions only if they viewed the hammer tapping silently (unimodal, nonredundant) but not if they viewed it and could hear its synchronous sounds (bimodal, redundant). Later in development, by 8 months of age, infants were able to detect the change in orientation in both unimodal and bimodal stimulation. Similar patterns are found for infant discrimination of individual faces and voices, discussed later (see Social Development). These findings illustrate the dual role of intersensory redundancy, both facilitating detection of amodal information and impairing detection of modality-specific information, particularly in early development when attentional resources are most limited. They also suggest that domains that provide a great deal of multimodal, redundant stimulation such as social interaction, speech, live music, dance, or stimulation from the self, can highlight and promote earlier learning about salient amodal aspects of stimulation. In contrast, domains that involve exploration through only a single sense modality, such as viewing a silent object or event, listening to music or speech with no visual accompaniment, are more likely to promote learning about nonredundant, modality-specific aspects of stimulation.

Evidence for the role of task difficulty (principle 4) has also been found. Older infants show detection of amodal tempo information in both bimodal and unimodal stimulation from a toy hammer tapping (developmental improvement). However, they revert back to patterns of intersensory facilitation (detecting tempo changes only in bimodal redundant stimulation) when the tempo discrimination task is made more difficult. Moreover, evidence for both the principles of intersensory and unimodal facilitation has now been found for older infants and children, and these principles may also apply to adults when learning novel content or when attention is taxed or divided.

The principles of the IRH are thought to be general principles of intermodal development, applicable across a wide range of event types and tasks. They were derived from convergent findings from studies of both human infants and non-human animal embryos and infants. Even nonhuman animals show intermodal and unimodal facilitation. Bobwhite quail embryos and chicks demonstrate a remarkable advantage for the perception of amodal information when it is redundantly presented as compared with when it is available in unimodal stimulation alone. For example, quail embryos learned the temporal characteristics of a maternal call four times faster and remembered it four times longer when the call was presented redundantly, in synchrony with a flashing light, as compared with no visual stimulation or with an asynchronous presentation. These convergent findings across species demonstrate that the principles of intersensory and unimodal facilitation are broad principles of intersensory functioning, generalizability across developmental periods, tasks, events, and response types.

The formulation of the IRH was guided by the ecological theory of perception developed by James Gibson in the 1960s and adapted by Eleanor Gibson, who formulated the developmental account of the ecological theory in her seminal book on perceptual development in 1969. In particular, two general principles of perceptual development (reviewed below) provide a foundation and guiding principles for the development of intersensory perception.

### Increasing Specificity of Information Pick-up Characterizes Perceptual Development

Eleanor Gibson proposed that perceptual development is characterized by a process of progressive differentiation of finer and finer levels of stimulation. That is, across development, infants detect global, abstract levels of stimulation and progress to increasingly more specific information about objects and events. Thus, our early perception of events, such as a person walking, might begin with the general perception of a person walking, and progress to more detail regarding perhaps the gender and age of the person, how they walked, how they talked, and finally to their identity and specific appearance. This pattern, Gibson proposed, is a general characteristic of perceptual learning. Thus, it also characterizes child and adult learning in new domains, such as mastering a new language or developing expertise in new areas such as bird-identification or wine-tasting. This principle goes hand in hand with the early salience of amodal, redundant information. Amodal information is considered abstract, and global, in the sense that it is not specific to one sense modality, but is common to several. It is therefore detected early in the process of perceptual differentiation, and thus can serve as a general framework for later perception of specific details, as illustrated in the research described in the section titled "The development of audiovisual perception."

### Increasing Economy of Information Pick-up Characterizes Perceptual Development

Across development we become more and more experienced perceivers. With experience, according to Eleanor Gibson, we not only differentiate finer levels of stimulation, but we also become more efficient perceivers. Thus, the development of perception entails increasing economy of information pick-up. More recently, it has been found that this may be characterized by two opposing processes. With experience, children not only improve their perceptual discrimination of the information they attend to, but they may also show a progressive decline in perceptual discrimination ("perceptual narrowing") for information to which they receive little exposure. Perceptual narrowing has been known to characterize auditory perception of speech sounds since the later quarter of the twentieth century, but it was only in the first few years of 2000s that it was demonstrated for perception of visual and multimodal information.

In speech perception, it was found that infants younger than 6 months of age were able to discriminate phonemes (speech sounds) that occurred in a variety of languages, but by 10–12 months, they were no longer able to discriminate some phonemes that did not occur in their native language. Perception of native phonemes improved as a result of experience, but perception of nonnative phonemes declined as a result of lack of experience. Similarly, in the area of face perception, infants of 6 months

Author's personal copy



were able to discriminate among a variety of different human faces and monkey faces; however, by 9 months, discrimination was limited to human faces. Intermodal perception of face–voice relations may also undergo perceptual narrowing. Infants of 4, 6, 8, and 10 months were shown pairs of monkey faces producing two different vocalizations, a coo and a grunt, along with a soundtrack synchronized with both but matching only one of them. Only the two youngest groups showed intersensory perception and looked at the face that matched the soundtrack. This suggests that intersensory tuning to intermodal face–voice relations is initially broad across the first year and it may narrow as infants acquire more experience with human faces than monkey faces.

Thus, increasing economy of information pick up, like increasing specificity of information pick up, characterizes intermodal perceptual development. As we become more expert perceivers, perceptual narrowing may occur as a natural consequence of lack of input and economy of information pick up. Given limited resources for attention and information pick up, perceptual narrowing allows perceivers to focus attentional resources on aspects of stimulation that are relevant to their particular environment and ignore those that are irrelevant. This process is consistent with recent advances in our understanding of neural development, discussed next in greater detail. Research on perceptual narrowing provides some of the only examples of perceptual abilities possessed by infants that are superior to those of children and adults!

## Neural Bases of Intermodal Perception

Current neuroscience research is consistent with and provides a biological basis for the principles reviewed above and the research findings reviewed in the sections ahead. Virtually every major area of the brain, including early sensory cortex, receives and sends input from different sensory modalities. Recent work has identified especially high concentrations of these multimodal neurons, neurons that respond to input from multiple modalities, in the superior colliculus (a nerve bundle early in sensory processing that is linked to attention and orienting). These neurons respond in a manner that is often described as super-additive. That is, while they may respond to input from one modality alone (e.g., visual), they respond most strongly when inputs from two or more modalities are combined (e.g., visual plus auditory). In addition, there are other neurons that are tuned to particular locations in space. Although they respond primarily to stimulation in one modality, prior experience with multimodal events allows the brain to relate auditory and visual input from the same spatial location, for example.

Research in the neurosciences is also consistent with the principles of increasing economy of information pick up, perceptual narrowing, and increasing specificity that characterize intermodal perception at the behavioral level. The developing brain can be understood as a mass of interconnected neurons, whose connections are strengthened (through a process called long-term potentiation) or gradually weakened (through a process called long-term depression) in accord with experience. As connection strengths change, these neural networks become ever more efficient and economical at dealing with stimuli and patterns of information previously experienced (something called "sparse coding"). For example, if input to the visual modality is somehow modified (by wearing glasses with prisms that distort one's vision), the receptive field of cells in the superior colliculus can compensate and realign with those of the other modality to maintain a coherent spatial mapping, by adjusting the strength of neural connections. This constant process of reorganization and tuning allows the brain to be extremely plastic not only to new experiences but to injury as well. Following amputation, areas of the brain once devoted to the missing limb become responsive to sensations from other areas of the body. Unfortunately, the brain occasionally interprets this information as having come from the missing limb, leading to phantom limb sensations.

In early development, unused and weakened synapses atrophy or are pruned altogether. Consistent with the principle of increasing economy of information pick up, the neural connectivity of the brain also becomes increasingly more economical across development. Pruning is a process by which the neural connections that are not used disappear. Early in development, the brain massively overproduces neurons and their connections rapidly increase reaching a peak near the end of the first year of life. Connectivity is continuously pruned back as a function of experiences such as reaching, looking, and listening, particularly across the first years of life. In other words, the structure of the brain is a "history of its use." Coordinated visual, auditory, and motor behavior may thus lead to more intermodal neural connectivity, in turn supporting more intermodal coordination. Although this plasticity is most apparent in early development, the brain continues to exhibit some plasticity throughout the lifespan. The early plasticity of the brain, its sensitivity to multimodal inputs based on coincidence detection, and its reliance on experience in the multimodal world to guide neuronal development provides a neural basis for the behavioral research findings (in the sections ahead) on the early development of intermodal perception.

## Fetal Development

Currently, little is known about the development of intermodal perception during fetal development. However, it is known that the senses develop in an invariant sequence (common across birds and mammals) during fetal development, with tactile/vestibular development, followed by the chemical senses (taste and olfaction), followed by audition, and finally visual functioning. Auditory stimulation from external sounds and the mother's voice is detected by the fetus during the third trimester.

In the fetal environment, there is a wide range of sensory stimulation and ample opportunity to experience synchrony across the senses. For example, the fetus experiences sounds, both external to and within the fetal environment, vestibular changes as the mother moves about, tactile sensations from touching it's own body as well as the uterine wall, as well as proprioceptive

Author's personal copy



feedback from self movement. This stimulation is patterned over time and synchronized across the senses, likely promoting the development of neural architecture for synchrony detection in utero. For example, the mother's speech is accompanied by synchronous movements of her body, vibrations of her spinal column, and movements of her diaphragm. This synchrony would create a great deal of amodal, redundant information that may make the mother's voice and its prosody particularly salient to infants when they encounter it later, providing a basis for early maternal preferences after birth. Research has yet to directly test these hypotheses. However, research has found that newborns are able to detect both audiovisual and visual-tactile synchrony. Further, using a nonnutritive sucking procedure, research has demonstrated that newborns prefer their mother's voice over that of a female stranger. Neonates, aged 1–3 days, show a visual preference for the face of their mother over that of a female stranger. A recent study demonstrated that this preference disappeared if neonates were prevented from hearing their mother's voice from birth until the time of testing. Preference for the mother's face was shown only by neonates who had postnatal experience with the mother's face–voice combination. This suggests that the early preference for the mother's face is the result of intermodal perception: neonates are attracted to their mother's face shortly after birth because it occurs in synchrony and concurrent with her voice, a familiar stimulus with which they have had several months experience as a fetus. Thus, it is possible that the preference for the mother's voice is not only a result of familiarity, but also a result of intermodal perception during prenatal development.

## The Development of Audiovisual Perception

Connecting sight with sound is fundamental to a host of activities: from lip-reading and learning words to matching a face to a voice or locating the person who is calling your name in a crowded room. Discovering the connections between what one sees and what one hears is essential to making sense of our environment and vital to our survival.

Even in the first weeks of life, infants have developed a primitive coordination of audiovisual space. They turn their eyes toward the locus of a sound. This coordination is important, because there is typically visual information about an audible event in the direction of the sound. This ability ensures that infants will focus their eyes and ears in the same general location, ready to pick up intersensory redundancy, should any be available. Localization is imprecise at first but improves with age. Experience helps infants calibrate or align their auditory–visual spatial maps.

Infants appear to be quite skilled at detecting amodal, redundant information uniting visual and acoustic stimulation. For example, they can detect the temporal synchrony between sights and sounds of an object's impact, between faces and voices during speech, and between showing and naming an object. They are sensitive to more fine-grained amodal information in faces and voices for emotion and for the gender and age of a person. Infants perceive auditory and visual information for the changing position of an object moving through space, and the substance and composition of an object striking a surface. These remarkable capabilities and what we know about their development are discussed in more detail in the sections that follow.

In contrast to the large research focus on infants' detection of amodal information and its importance in linking information across the senses, fewer studies have focused on infants' detection of arbitrary relations across the senses and their ability to perceive modality-specific properties of stimulation. However, perception of objects and events relies on the development of both amodal and arbitrary audiovisual relations. In particular, the development of language and social perception rely to a large extent on mastering knowledge of arbitrary relations. In order to relate objects with speech sounds, one must learn the arbitrary relation between the speech sound and the particular object to which it refers. Social perception, for example, requires that we learn to differentiate among individuals by discriminating among hundreds of different faces, and learn to relate the appearance of specific faces with specific voices.

### Development of Object and Event Perception

#### *Increasing Specificity in Event Perception*

Research has shown that infants' detection of amodal redundant information, such as temporal synchrony, typically precedes and provides a basis for learning arbitrary relations such as those between speech sounds and objects, between specific faces and voices, or objects and their particular impact sounds. Thus, research focusing on the development of object and event perception illustrates the principle of increasing specificity.

Temporal synchrony is thought to be the most global type of amodal information, serving as the glue that binds information across auditory and visual stimulation. Research shows that for events such as a bouncing ball, the temporal synchrony window for infants is approximately 350 ms and decreases with age, to approximately 80 ms for adults. This window may be larger for detecting synchrony between faces and voices. A more fine-grained level of temporal information called "temporal microstructure" serves as the basis for perceiving the relation between the type of sound and the type of movement an object makes. For example, object substance (e.g., hard vs. soft) and composition (e.g., single vs. multiple components) are perceived by detecting temporal microstructure when the object contacts a surface. Research indicates that very young infants (3 weeks of age or younger) perceive the temporal synchrony (global information) between the movement of an object and the natural impact sounds it produces, and a few weeks later (by 6 weeks of age), they perceive the substance and composition (microstructure) of the same object. A few months later infants detect arbitrary, modality-specific relations such as the relation between the pitch of the impact sound and the color and shape of the object. This illustrates the principle of increasing specificity where detection of global amodal relations

Author's personal copy



(synchrony) promotes detection of more specific, nested, amodal relations (temporal microstructure) and finally, further exploration promotes detection of arbitrary, modality-specific relations (pitch-color/shape).

In this manner, detection of amodal information can guide and constrain perceptual learning about increasingly more specific aspects of multimodal stimulation. Amodal relations such as temporal synchrony, common rhythm, and tempo provide a basis for selecting unitary sights and sounds. This promotes continued processing of the unified event and, in turn, differentiation of temporal microstructure and modality-specific properties such as color, pattern, pitch, and timbre. Detection of amodal information thus provides a natural avenue for connecting the correct sights and sounds and a buffer against connecting the wrong sights and sounds.

A training and transfer study with young infants illustrated this principle. Following training with films and soundtracks of objects striking a surface, 3-month-old infants were able to learn the correct sight-sound connections but unable to learn the incorrect connections. Infants received training with films depicting a single, large marble and a cluster of small marbles striking a surface along with the natural soundtracks to each (see Fig. 3). Training occurred under one of four conditions; the films were paired with the correct vs. the incorrect (mismatched) soundtracks and the films were presented either in synchrony or out of synchrony with the soundtracks. Results indicated that only infants who were trained with synchronized and correctly paired films and soundtracks learned to match the film with the soundtrack they had received during training.

Learning of sight-sound relations occurred only when both temporal synchrony and temporal microstructure specifying object composition were correct (e.g., single objects make single impact sounds and compound objects make compound impact sounds). Infants did not learn to relate objects with inappropriate impact sounds (e.g. single objects with compound impact sounds), even when their movements were synchronized with the onsets of the sounds, at least following a single training session. This illustrates that detection of redundant, amodal information can serve as a gatekeeper, promoting processing of sights and sounds that belong together and buffering against processing of unrelated ones.

### Social Development

The world of social interaction provides a great deal of multimodal information, making intermodal perception critically important for learning and functioning in this domain. Understanding emotion and communicative intent, relating the faces and voices of individuals, identifying who is speaking in a crowd, and sharing experience with another, all rely on connecting information across the senses.

At birth, infants look longer at faces accompanied by voices. Around 4–6 months (but not at 2 months) infants learn to relate the specific face and voice of a novel individual on the basis of a few minutes of familiarization. At this same age, infants can also match unfamiliar faces and voices on the basis of age and gender of the speaker. They appear to know that males and adults have larger faces and features and deeper voices than females and children. By 3 months of age, infants can match facial and vocal expressions of

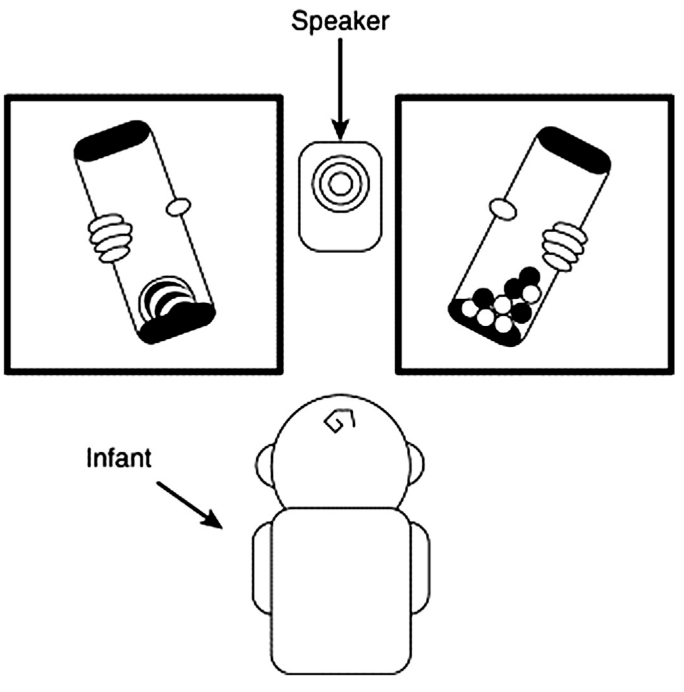


**Figure 3** Illustration of the experimental set-up and videotaped events used. Adapted from Bahrick, L.E., 1988. Intermodal learning in infancy: Learning on the basis of two kinds of invariant relations in audible and visible events. Child Dev. 59, 197–209.

Author's personal copy



emotion, such as happy, sad, and angry, in their mothers, and by 5–7 months, they can do so with unfamiliar adults too. As noted earlier, research indicates that perception of emotion, like perception of the rhythm and tempo of events, is based on detecting intersensory redundancy across sight and sound. It develops in a sequence consistent with intersensory facilitation and developmental improvement described by the IRH. Emotion perception emerges first in multimodal stimulation (intersensory facilitation) and is later extended to unimodal auditory and finally, to unimodal visual stimulation. One study showed that at 4 months of age, infants discriminated between happy, sad, and angry expressions in unfamiliar adults when they were seen and heard together; by 5 months they discriminated these expressions in vocal stimulation alone, and by 7 months they were able to distinguish the facial expressions alone.

Infants not only recognize the emotional expressions of others, but they use others' expressions as information about external events. This skill, called social referencing, involves discriminating an adult's emotional expression and connecting it to an ambiguous object or event, such as a novel toy or situation, and allowing the emotional expression of the adult to influence exploratory behavior toward the novel event. One study showed that after watching a video of a woman responding with pleasure and interest to the movements of one toy, and with fear to the movements of another toy, 7-month-old infants touched the toy more often that had been paired with expressions of interest and joy. Infants, however, showed this effect only when they could both see and hear the videos, and not when they could see them alone. This demonstrates intersensory facilitation and highlights the importance of intersensory redundancy for perceiving these relations when infants are first learning about them. Thus, social referencing likely emerges first in a multimodal context, where the caregiver provides coordinated auditory and visual emotional information, and then is later extended to unimodal contexts. A fearful facial and vocal expression might inhibit exploration at first, and later a visual expression alone might suffice.

Infants not only detect amodal information across the senses, but they participate in temporally coordinated, interactions with adults. In close face-to-face interaction, infants coordinate their movements and sounds with those of the adult in a burst-pause, turn-taking pattern. This "protoconversation" involving sensitivity to the emotional dynamics and temporal characteristics of social interaction promotes emotional attunement and a sense of shared experience between the infant and caretaker. It provides an important foundation for social development and communication.

In contrast to perception of emotion, perception of individual faces and voices is promoted in unimodal visual and unimodal auditory stimulation, respectively. Recall the dual role of intersensory redundancy, facilitating attention to amodal information but impairing attention to modality-specific information in multimodal events, at least for unfamiliar events. Faces are discriminated primarily based on modality-specific information such as the appearance of individual features and their configuration and voices are primarily discriminated on the basis of pitch and timbre. Thus, perception of individual faces and voices is not optimal when the individual can be both seen and heard speaking. Consistent with the principle of unimodal facilitation of the IRH, detection of this modality-specific information is selectively promoted in unimodal stimulation in early development (see Fig. 2). Research has supported this view and shows that discrimination among faces of unfamiliar women emerges by the age of 2 months in unimodal visual stimulation following a single training session, but is not present in multimodal, audiovisual stimulation until somewhat later in development. Similarly, discrimination among voices of unfamiliar women emerges first in unimodal auditory stimulation and is later extended to bimodal, audiovisual speech.

Thus, as a result of experiencing the social world of faces, voices, speech and emotions, young infants rapidly learn to differentiate the complex and dynamically changing information provided by multimodal social events. Detection of amodal properties of stimulation appears to develop first in redundant, multimodal stimulation, whereas detection of modality-specific properties such as that underlying recognition of specific faces and voices develops first in unimodal, nonredundant stimulation. It is not yet known to what extent these principles characterize adult perception, particularly when tasks are difficult or adults are learning new information.

## Intermodal Development of Speech Perception and Language

Another important domain of audiovisual perception is language. Language, by its very nature relies on uniting visual with auditory stimulation: From localizing a particular speaker, to lip-reading and learning that sounds stand for particular objects, combining sights and sounds is fundamental to typical language acquisition. Language development can also be seen as a special case of social development; however, because of the large and varied research on this topic, and because of the important role of arbitrary relations between speech sounds and their referents, we include it as a separate topic.

### *Localizing Speakers Through Audiovisual Synchrony*

We are often exposed to multiple sources of sound and motion when observing someone speaking. For example, a television may be heard in the background at the same time as the caregiver is speaking to the infant. Despite the potential perceptual processing difficulty for an infant, detecting audiovisual synchrony can assist in localizing the speaker. Infants' abilities to localize sound are poor when they have only auditory information. For example, 6-month-olds can only locate objects within about 19° using auditory information alone. Although adults are more accurate, they also make extensive use of visual information in localizing talkers. This is illustrated by the "ventriloquist effect." By moving the dummy's mouth in synchrony with the ventriloquist's own speech sounds, a ventriloquist is able to fool observers into thinking the dummy is speaking. Infants especially rely on this visual information to localize a speaker and are sensitive to the temporal synchrony between movements of the mouth and speech sounds. Thus,

Author's personal copy



10–16-week-old infants prefer to look at faces that are synchronized with speech sounds as opposed to faces that are out of synchrony with speech sounds. There is now some evidence that even newborns show this preference.

These findings serve as the basis for a new method for assessing individual differences in infants' abilities to locate a sounding event amidst other events. In this method, the intermodal processing efficacy protocol (IPEP), the child views an array of six dynamic faces or objects, while hearing the synchronous soundtrack to one of them at a time. For example, six women are shown, each reciting a different story, and the natural, synchronized voice from one of them is heard, and then another, and so forth, across a large number of short trials. The infant's visual fixations are monitored by an eye tracking device to see how well they pick out the speaker based on detecting amodal audiovisual synchrony. Research indicates that infants learn to find the speaker in the crowd with greater and greater speed and accuracy across development. Most research discussed thus far has used methods designed to assess intermodal perception characterizing groups of infants at particular ages and these methods have not been fine-grained or reliable enough to characterize the differences in abilities of individual infants within an age group. This focus of the IPEP is particularly important if one wishes to study how intermodal perception serves as a foundation for other skills such as language, social, and cognitive development. Accordingly, recent research findings suggest that the infant's accuracy in finding the speaker in the crowd in the IPEP may be related to their early language skills (see Word Learning, ahead, for more on this topic).

### Face–Voice Connection

Infants not only can find the person who is talking using audiovisual correspondence, but they also rapidly develop the ability to perceive faces and match faces to appropriate voices. Not only can they link faces and voices on the basis of gender, age, and identity, but infants also remember and predict the kinds of speech sounds that people are likely to make. They link the shape of a persons' mouth with the sounds that they produce. For example, 2-month-olds can determine which of two women is producing an/i/ sound and which is producing an/a/ sound by simply looking at their mouth movements.

Perceptual illusions also highlight the powerful interactions of auditory and visual speech perception. The McGurk Effect is an auditory-visual illusion that illustrates how perceivers merge information for speech sounds across the senses. For example, when we hear the sound "ba" while seeing the face of a person articulate "ga," many adults perceive the sound "da," a third sound which is a blend of the two. Similarly, infants also show evidence of this effect in the first half-year of life.

In addition, adults typically speak to infants using exaggerated intonation, called "motherese" or infant-directed speech. Natural infant-directed speech contains a great deal of amodal, redundant information such as exaggerated rhythm and tempo changes, longer pauses, more repetition, wider pitch excursions, and shorter utterances that can be experienced in the sounds of speech, facial movements, and gestures together. Research indicates that infant-directed speech and the accompanying facial expressions are highly salient and preferred over adult-directed speech by infants, regardless of culture or language spoken, in part because of the high degree of multimodal information and intersensory redundancy. However, even in unimodal auditory speech, the preference for infant-directed speech is already apparent in the first months of life. Given that infants have prenatal exposure to the sounds of speech (see prior section Fetal Development), learning about the sounds of speech in one's native language is particularly rapid.

### Hearing Better Through Sight

Another way in which intermodal perception is useful for language is in helping us to separate a speech stream from a background of noise. It appears that infants can actually use what they see to hear better and to selectively attend to a particular speech stream. When it is noisy, adults can use the information in the speaker's face to help them figure out what is being said. This is one reason why it is often more difficult to understand a person on the telephone than in person. Infants, as well, appear to gain an advantage by seeing the face of a speaker. When a face is synchronized with the voice, infants are better at picking out individual words (speech segmentation) in the context of greater background noise than for auditory speech alone. Interestingly, infants also are able to segment words from speech better if, instead of a face, an oscilloscope pattern (a wiggly line) is synchronized with the voice, demonstrating the importance of audiovisual redundancy for speech segmentation. So infants likely can use any synchronous movement to help them hear better – even gestures, or head movements.

### Word Learning

Perhaps the most dramatic evidence of relating what infants hear with what they see comes in the form of early "word learning." Part of the earliest language acquisition involves learning associations between the sound patterns of words and meaning as, for example, when one learns to connect the visual image of a rose to the sound combination /roz/. Recent investigations have shown that by 6 months of age, infants are already starting to learn the meanings of some very common words such as "mommy" and "daddy." Using the intermodal preferential-looking paradigm (see Fig. 4), 6-month-olds were presented with two videos, one of their mother and the other of their father, played side-by-side simultaneously. Alternating recordings of a voice, saying "mommy" or "daddy," were presented. Infants looked more often at the correct video in response to hearing the names. Of course, it is not known whether 6-month-olds truly understand that the speech sounds "mommy" and "daddy" actually refer to their own mother and father or if they simply have noticed that those speech sounds and sights are associated (e.g., "mommy" is heard when mother is present). In either case, it is clear that infants have made some connection between the acoustic words /mami/ "mahmee" and /dædi/ "dadee" and their visual referents.

In keeping with an increasing specificity view of perceptual development and the role of intersensory redundancy in uniting sights and sounds, this learning may be driven, in part, by amodal synchrony. For example, 7-month-olds can learn the links between two speech sounds and two objects if the speech sounds occur in synchrony with the motions of the object, but show

Author's personal copy



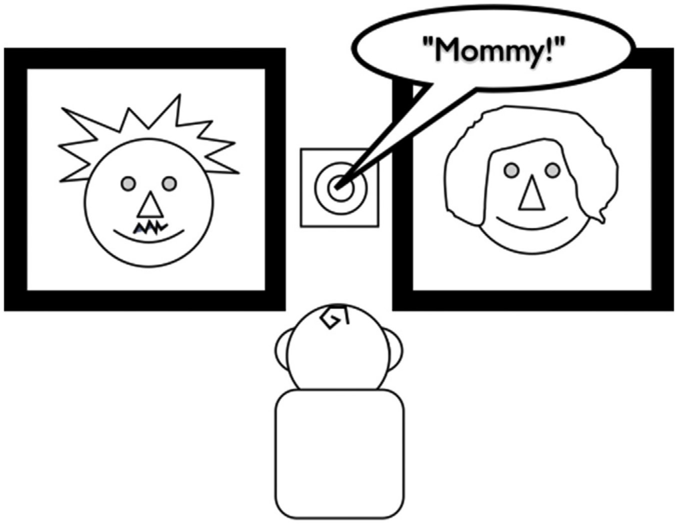


**Figure 4** The intermodal preferential-looking procedure. Infants are expected to look longer at the screen that matches the auditory information. In this case, they would be looking longer at the picture of their mommy.

no learning if they are out of synchrony. In one study, infants saw a hand move one of two unfamiliar objects (a toy crab and a porcupine, or a lamb chop and a star), in synchrony with the vowel sound /a/ "ahhh" (e.g., for the crab) and /i/ "eee" (e.g., for the porcupine). Each time the object was moved, the sound was uttered, simulating showing and naming an object. In the unsynchronized condition the same movements were made; however, the vowels were uttered between the object movements. We are not suggesting that this is all it takes to learn a word. Word-learning is a complicated task involving multiple cues, and including numerous social-pragmatic factors. However, when infants begin to learn words, audiovisual synchrony is one way they can jumpstart the process, and it may provide a perceptual gateway into language learning.

In fact, parents appear to have an implicit understanding of this principle. When mothers were asked to teach their infants novel names for objects, they used a great deal of temporal synchrony between movements of the object and the verbal label, particularly for younger or preterm infants. This helps young infants learn that objects have names. For older infants who understand this idea, mothers adapted their style to include more naming and pointing, and somewhat less synchrony.

## The Development of Visual-Tactile Perception

Amodal information also unites perception across vision and touch. Information for shape, texture, substance, and size are invariant across visual and tactile stimulation. When we feel a ball in our hands, we perceive the same shape, size, texture, and substance that we see. Typically, we see and feel an object concurrently and during visual-tactile exploration, we detect redundant information for these properties through touch and sight. Research using the "cross-modal transfer" method has been conducted to investigate perception of visual-tactile correspondence in infants and adults. In this method, an object is presented to one sense modality alone and a preference test is then given in another sense modality to determine whether the information transfers across modalities. Using this method, research has shown that even 1-month-old infants can perceive the correspondence between an object they experienced tactually, through oral exploration of the object on the back of a pacifier, and a three-dimensional version of the object available visually. Infants looked more to the object of the shape and texture they had previously experienced orally than to a novel object of a different shape. Infants are also able to transfer information about the substance of an object (rigid vs. deforming) across touch and sight.

Evidence also shows that infants can transfer information obtained through manual exploration to vision, and this develops across the first year. One factor determining the extent to which manual information is perceived is whether exploration is active or passive. Tactile exploration develops over the first year. Young infants tend to grasp objects, whereas older infants become more adept at obtaining tactile feedback by moving their hand relative to the object's surface. By 4 months, infants can perceive whether two parts of an object are connected or separate by the type of motion they produce during haptic exploration. Not until 5–6 months of age, as infants learn to sit independently and gain better control of their trunk and head, do they begin to spontaneously couple tactile exploration with visual examination. This facilitates their perception of the amodal, 3-dimensional shape of objects. By 6 months, infants can recognize the shape of an object visually that they have manually explored, as long as exploration is active. One study found that even newborns could transfer information about the shape of an object from touch to vision for two shapes that could be easily discriminated by grasping (objects with straight vs. round contours). Following tactile exploration of a cylinder or a cube, newborns showed a visual preference for the object they had touched.

Author's personal copy



Recent research also indicates that newborns can perceive visual-tactile synchrony related to their own body. Newborns received rhythmic stimulation from a brush stroking their cheek along with videos that were either in synchrony with the stroking or out of synchrony. Newborns preferred to look to the video that was synchronized with the brush stroking their cheek. This finding leaves ripe for investigation questions about the fetal experiences that led to this early sensitivity. Recall that the fetal environment provides a wide range of synchrony experiences related to the infant's own body, ranging from the mother's voice and her movements that can be felt by synchronous tactile, vestibular, and proprioceptive (see next section) stimulation to the infant, to the infant's own movements that can be simultaneously experienced both proprioceptively and tactually, as one part of the infant's body comes in contact with another.

## The Development of Visual-Motor Correspondence and the Self

As described above, amodal information serves as a fundamental basis for infants' ability to perceive the self. Research has shown that young infants, including newborns, can imitate facial expressions. After viewing an adult model a tongue protrusion or a mouth opening, young infants show an increase in their production of the gesture they have just seen. In order to do this, they must relate their own production of the facial expression with the visual appearance of the adult's facial expression. This is likely guided by proprioceptive information. Proprioception is information about self-motion provided by feedback from the muscles, joints, and vestibular system. Facial imitation reveals evidence of an early intermodal coordination between motor behavior and visual information, and this coordination continues to develop across the first year.

Another procedure which reveals evidence that infants are able to perceive their own body motion by detecting amodal information is one in which they receive live video feedback from their own limb movements (see Fig. 5). In this method, infants view two video images side-by-side: one, a live visual display of their own motion, and the other a prerecorded image or an image of another infant's motion. Research reveals that by 3–5 months of age, infants can distinguish between a live video of their own legs kicking and a video of another infant's legs kicking, a spatially incongruent video of the legs, or a prerecorded video of their own legs. Further, infants consistently look more to the prerecorded image or the image of the peer, indicating a preference for social stimulation over stimulation from the self by the age of 5 months. Infants apparently use redundant, amodal information including temporal synchrony and spatial congruence between the proprioceptive experience of their own limb motions and the visual image of the motion in the video display to make this discrimination between self and other.

Coordination between vision and motor development is also evident by the rapid development of "visually guided reaching" during the first year. Visually guided reaching entails continuous adjustments in the reaching and manual behavior of infants as a function of visual input about the size, shape, and position of objects. Infants show evidence of this coordination in the first months of life. They are even able to contact a moving object by aiming their reach ahead of it and taking into account the speed and direction of their own arm motion in relation to that of the object. Later, infants show an ability to adapt their crawling and exploratory behavior as a function of visual information about the slant and solidity of the surface. These examples illustrate a rapidly developing relationship between vision and motor behavior and an understanding of self in relation to objects.

Posture-control is a critical factor in the infant's development, allowing the child to maintain a stable relation to the environment and is a prerequisite for exploratory behavior such as reaching, grasping, and locomoting. Posture-control is

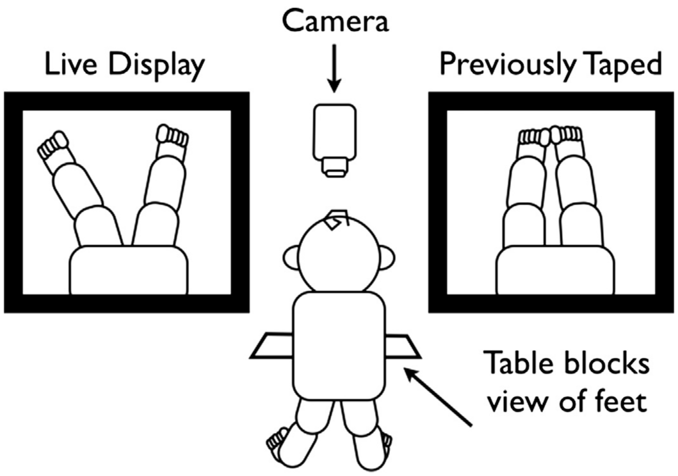


**Figure 5** Infants can distinguish live video feedback from their own limbs from that of a prerecorded display of their limb motion. Adapted from Bahrick, L.E., Watson, J.S., 1985. Detection of intermodal proprioceptive-visual contingency as a potential basis of self-perception in infancy. Dev. Psychol. 21, 963–973.

Author's personal copy



considered an example of intermodal perception because vision plays a critical role in the maintenance and control of posture. We rely on a visual frame of reference for orienting in space and maintaining an upright posture. Studies using a moving-room paradigm, where a conflict between visual and proprioceptive information is created, have demonstrated that vision can override proprioceptive information, particularly in inexperienced walkers. In this method, the observer is placed in a room with a stable floor and walls that can either move or display visual information characteristic of motion (see Fig. 6). When visual information (optic flow) specifies that we are moving forward, we compensate by moving backward, and this compensatory body sway can be observed and measured. Infants as young as 5 months show postural sway while sitting. Postural sway increases up to 9–13 months of age and then levels off or decreases. Infants show greater sway in response to optic flow in the early stages of learning a posture (sitting up or standing) and with experience they are better able to resolve the conflict between vision and proprioception. Even adults show some degree of body sway in the moving-room paradigm.

## Other Sensory Combinations

Infants show evidence of an early developing interaction between auditory and vestibular/proprioceptive information. Young infants can feel the beat of music. They show intersensory perception of rhythm by coordinating what they hear with what they feel. When they listened to a musical passage and at the same time were bounced in time to it, with an emphasis on every second beat (like a march) or every third beat (like a waltz), they perceived the musical passages with the same emphasis.

Young infants are also able to link visual and olfactory information. They show recognition of their mother on the basis of her odor in the first weeks of life. Breastfed infants orient preferentially to the odor of their mother over that of another lactating female. This illustrates their sensitivity to natural object–odor combinations. Research suggests that this learning may originate in fetal development. The chemical make-up of the mother's amniotic fluid is similar to that of her scent. A few months later, infants can also detect arbitrary object–odor relations. In one study, infants were presented with two objects side-by-side for a baseline visual preference test. Then one of the objects was paired with a distinctive cherry odor. During the test trials that followed, the two objects were presented side-by-side, with the cherry odor (experimental trial), and without the cherry odor (control trial). Four-month-old females showed greater looking to the object paired with the odor during the experimental than the control trial, demonstrating that they learned the arbitrary relation between the visual appearance of the object and the cherry odor. In another study, 6-month-olds demonstrated the ability to relate the color of a food with its temperature.

Similarly, the link between taste and vision is amply illustrated by the toddler who turns her head away long before the broccoli comes in contact with her mouth, and who struggles to reach for the box of cookies. Research indicates that children have a "sweet tooth" from birth and by the time they are fed solid foods learn to quickly recognize which colors and foods are tasty and which are not, much to the dismay of parents.

## Conclusions

From combining sight with sound, touch, or motor movements, intermodal perception is fundamental to early development and perception of most objects and events in the environment. Early sensitivity to amodal information, the temporal, spatial, and intensity patterns of events that are redundant across stimulation to different senses (intersensory redundancy), guides infants' perceptual, cognitive, and social development. The senses work together as a coordinated perceptual system and intermodal perception develops rapidly across infancy. Even very young infants are sensitive to amodal information, allowing them to perceive unitary

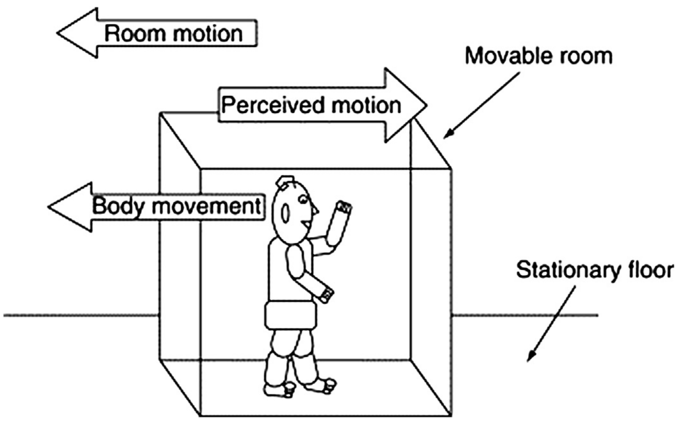


**Figure 6** The moving-room paradigm. Infants experience perceptual motion as a result of the optic flow created when the room walls move toward them. They then attempt to compensate for the perceptual motion and show backward body sway.

Author's personal copy



multimodal events by linking sights and sounds of speech, emotional expressions, and objects, as well as information across visual and tactile, olfactory, and proprioceptive stimulation. Intersensory redundancy is highly salient and captures attention, particularly in early development. This salience has a dual role, promoting attention to global, amodal information (such as tempo, rhythm, emotion, and prosody in speech) and impairing attention to modality-specific detail (such as color, pattern, pitch, and timber, including information for discriminating individual faces and voices), particularly in early development. This results in attentional salience hierarchies that guide early perceptual development in an organized, meaningful manner. In general, perceptual development proceeds along a path of differentiation of increasingly more specific levels of stimulation, from amodal to arbitrary relations, and exhibits increasing economy of information pick up with experience. The early plasticity of the brain, its multimodal nature, and its reliance on experience to guide neural development provide a neural basis for this developmental process.

## Acknowledgments

This article was supported by grants from the National Institutes of Health (NICHD RO1 HD053776 and K02 HD064943) to the first author.

## Further Reading

Bahrick, L.E., 1988. Intermodal learning in infancy: learning on the basis of two kinds of invariant relations in audible and visible events. Child. Dev. 59, 197–209.
Bahrick, L.E., 2010. Intermodal perception and selective attention to intersensory redundancy: implications for typical social development and autism. In: Bremner, G., Wachs, T.D. (Eds.), Blackwell Handbook of Infant Development. Oxford England, Blackwell, pp. 120–166.
Bahrick, L.E., Lickliter, R., 2002. Intersensory redundancy guides early perceptual and cognitive development. In: Kail, R. (Ed.), Advances in Child Development and Behavior, vol. 30. Academic Press, New York, pp. 153–187.
Bahrick, L.E., Lickliter, R., 2014. Learning to attend selectively: the dual role of intersensory redundancy. Curr. Dir. Psychol. Sci. 23, 414–420.
Bahrick, L.E., Watson, J.S., 1985. Detection of intermodal proprioceptive visual contingency as a potential basis of self-perception in infancy. Dev. Psychol. 21, 963–973.
Bahrick, L.E., Walker, A.S., Neisser, U., 1981. Selective looking by infants. Cogn. Psychol. 13, 377–390.
Bremner, A., Lewkowicz, D.J., Spence, C. (Eds.), 2012. Multisensory Development. Oxford University Press, Oxford England.
Gibson, E.J., Pick, A.D., 2000. An Ecological Approach to Perceptual Learning and Development. Oxford University Press, New York.
Gogate, L., Walker-Andrews, A., Bahrick, L.E., 2001. Intersensory origins of word comprehension: an ecological-dynamic systems view. Dev. Sci. 4, 1–37.
Gogate, L., Hollich, G., 2010. Invariance detection within an interactive system: a perceptual gateway to language development. Psychol. Rev. 171, 496–516.
Lewkowicz, D.J., 2000. The development of intersensory temporal perception: an epigenetic systems/limitations view. Psychol. Bull. 126, 281–308.
Lewkowicz, D.J., Lickliter, R., 1994. The Development of Intersensory Perception: Comparative Perspectives. Lawrence Erlbaum Associates, Hillsdale, NJ.
Rochat, P., 1995. The Self in Infancy: Theory and Research. Elsevier, New York.
Stein, B.F. (Ed.), 2012. The New Handbook of Multisensory Processes. MIT Press, Cambridge MA.
Walker-Andrews, A., 1997. Infants' perception of expressive behaviors: differentiation of multimodal information. Psychol. Bull. 121, 437–456.

## Relevant Websites

http://nuweb1.neu.edu/cdl/ – Communication Development Lab, David Lewkowicz, Northeastern University.
http://dpblab.fiu.edu – Developmental Psychobiology Lab, Robert Lickliter, Florida International University, Miami, FL.
http://infantlab.fiu.edu – Infant Development Lab, Lorraine E. Bahrick, Florida International University, Miami, FL.
http://hollich.psych.purdue.edu – Infant Lab, George Hollich, Purdue University, West Lafayette, IN.
http://infantstudies.psych.ubc.ca/ – Infant Studies Center, Dr. Janet F. Werker, Department of Psychology, University of British Columbia, Vancouver, BC.