

Trehub, S. E. (2003). The developmental origins of musicality. *Nature Neuroscience*, 6, 669–673.

---

## PERCEPTUAL DEVELOPMENT: INTERMODAL PERCEPTION

---

Speaking faces, baking bread, speeding cars—the world provides a richly structured, continuously changing stream of stimulation to all of our senses. *Intermodal perception* (also called intersensory or multimodal perception) refers to perception of information from objects or events available to multiple senses simultaneously. Because most objects and events can be seen, heard, and touched, everyday perception is primarily intermodal. Despite the fact that information about the world is carried through different sensory channels that each provide distinct forms of stimulation, we are able to perceive a stable world of unitary objects and events (people speaking, cars honking), rather than separate sights, sounds, and tactile impressions. The senses work together as a coordinated perceptual system, even in newborns, and intermodal perception develops rapidly and with increasing specificity across infancy. How we accomplish this “integration” is a puzzle that has fascinated philosophers and scientists for centuries. This entry covers the history and theory of intermodal perception, as well as aspects of the development of the senses as they relate to intermodal perception.

### History and Theory

Over 2,000 years ago, Aristotle proposed a *sensus communis*, or common sense, to explain how we perceive qualities common to different senses, such as number, form, and unity. Centuries later, philosophers such as John Locke and George Berkeley proposed that we needed to integrate information across separate sensory channels before perceiving a unified object or event, such as a bell ringing, rather than perceiving separate streams of light and sound. This posed a “binding” problem for perception, in that the brain would need to somehow unify separate channels of sensory information to “construct” a coherent world. This constructivist view dominated thinking throughout

most of the 20th century, including Jean Piaget’s well-known theory of cognitive development. Piaget proposed that integration took place gradually over the first year of life through interacting with objects and coordinating information across the senses. Prior to integration, infants were thought to perceive a world of unrelated sights, sounds, smells, and tactile impressions, much like the “blooming, buzzing confusion” described by William James in 1890.

Not until the psychologist James J. Gibson published his work on ecological perception in 1966, *The Senses Considered as Perceptual Systems*, was this constructivist view seriously questioned. Gibson argued that our senses work together as a unified perceptual system, and that the existence of different forms of sensory stimulation was an advantage rather than a disadvantage for perception of unified objects and events. This is because we perceive amodal (not tied to a particular sense modality) information that is redundant or identical across the senses. This includes temporal synchrony, rhythm, duration, tempo, changes in intensity, and spatial location common to audiovisual events, and shape, substance, size, and texture common to visual–tactile events. Because all events occur over time and space, all events provide amodal information.

Gibson’s theory is considered a differentiation view of development. In contrast to the integration view, which holds that the senses are separate at birth, the differentiation view, articulated by Eleanor J. Gibson, proposes that the senses are unified at birth and that perceptual development is characterized as a gradual process of differentiating increasingly finer levels of sensory stimulation. From this view, now widely accepted among developmental psychologists, there is no “blooming, buzzing confusion” in early development. Rather, infants detect many general features of unified multimodal events from birth and learn to perceive increasingly more subtle differences and more complex objects and events through looking, listening, and interacting with their environment.

### Importance of Amodal Information and Intersensory Redundancy for Development

How do infants learn to detect unified multimodal events such as a person speaking? Evidence indicates

that amodal information, particularly temporal synchrony between sights and sounds, provides the glue that binds information across the senses and thus serves as a solution to the age-old binding problem. For example, by detecting the synchrony, rhythm, and tempo common to the sights, sounds, and tactile impressions of one's hands clapping, we attend to the event as a whole, rather than separate sights, sounds, and tactile stimulation. When the same amodal information is concurrently available and synchronized across more than one sense modality, this is called *intersensory redundancy*. Research has shown that intersensory redundancy is highly salient to humans and animals. For example, the special attention shown to social stimulation by infants of many species is likely a result of the high degree of intersensory redundancy social partners provide. Further, we know that redundancy highlights unitary multimodal events and their amodal properties, at the expense of other properties, much like a figure-ground effect. For example, an infant will notice the intersensory redundancy between the face and voice of a person speaking (synchrony, rate, rhythm, and co-location of audiovisual speech) at the expense of other (modality-specific) information, such as the appearance of the face or the pitch and timbre of the voice. In contrast, when redundancy is not available, such as when a person is still and silent, infants preferentially explore nonredundant (modality-specific) aspects, such as the appearance of the face.

Thus, redundant, amodal information plays a significant role in guiding and constraining which aspects of events we selectively attend, particularly in early development when attention is least flexible and capacity is most limited. Later, when attention is more flexible and efficient, we can attend to amodal properties as well as more specific properties of the same events in a shorter time. Furthermore, because selective attention provides the basis for what is perceived, learned, and later remembered, intersensory redundancy has a powerful organizing influence on early perceptual, cognitive, social, and emotional development.

### Development of Auditory-Visual Perception

Much research has focused on the ability of young infants to perceive multimodal events, and in particular, information that is amodal and redundant across audition and vision. Scientists have discovered

that even young infants are skilled at perceiving amodal information, that intermodal perception improves across the first year of life, and that it develops in order of increasing specificity, with global information detected developmentally prior to more specific levels of stimulation, consistent with the pattern of differentiation and "increasing specificity" proposed by Gibson.

### *Audiovisual Space*

As early as the first weeks of life, infants reliably move their eyes in the direction of a sound. This early coordination of auditory and visual space is important because it enables infants to discover visual information at the source of the sound and thus promotes detection of intersensory redundancy. At first, localization is rather imprecise, but it improves rapidly across infancy. Through auditory and visual experience with objects and events, infants calibrate and align their sense of audiovisual space.

### *Object and Event Perception*

In the first months following birth, infants can detect the temporal synchrony and spatial co-location uniting the sights and sounds of an object moving and striking a surface. By two to five months, infants match soundtracks with object motions on the basis of finer levels of amodal information, such as substance (elastic vs. rigid), composition, tempo, and rhythm. Soon thereafter, infants detect arbitrary, modality-specific information, such as the relation between the color and shape of an object and the pitch of its sound. Detection of global, sight-sound relations (e.g., synchrony) promotes detection of more specific amodal relations (e.g., tempo or rhythm), and this in turn promotes detection of arbitrary, modality-specific relations (e.g., pitch and color). These remarkable abilities illustrate the principle of increasing specificity and how processing of global information such as intersensory redundancy scaffolds (guides and supports) the development of more specific processing.

### *Social Development*

People provide a great deal of multimodal stimulation for infants, including talking, laughing,

singing, and touching. Adults and infants also engage in richly structured multimodal interactions, called *protoconversation*. This highly inter-coordinated mutual exchange of sounds, movements, and touch is a foundation for communication and social development. These early experiences contribute to and shape infants' fascination with social stimulation. Research has shown that newborns look longer to faces that are paired with a voice and prefer their mother's face hours after birth if they have previously seen her face together with hearing her familiar voice. By three months, infants can match facial and vocal expressions of emotion, such as happy, sad, and angry, in familiar people, and by five months, in unfamiliar people. Discrimination of emotional expressions, like perception of the rhythm and tempo of events, emerges first in multimodal (audiovisual) stimulation (by four months) and then is later extended to unimodal auditory (around five months) and unimodal visual (seven months) stimulation.

### *Speech Perception and Language*

Speech is inherently multimodal, involving coordinated facial, vocal, and gestural information, and audiovisual redundancy promotes learning in this domain as well. For example, adults typically speak to infants using a distinctive speech pattern with slower tempo, larger pitch changes, and more repetition. This style of speaking is called *motherese*, and research has found that infants across cultures and languages are particularly attracted to this type of speech. Adults also produce exaggerated facial expressions and gestures coordinated with speech during motherese. This multimodal motherese exaggerates amodal, redundant information (intensity, rhythm, and tempo changes produced by synchronous facial and vocal stimulation), making it easier for infants to maintain attention to and learn speech. Thus, speech is first learned and experienced in the context of multimodal, redundant stimulation. Even during the last trimester of prenatal development, fetuses likely receive multimodal stimulation from the sound of the mother's voice, coordinated with synchronous body movements, vibrations of her spinal column, and movements of her diaphragm. This redundancy may contribute to the salience of the mother's voice observed soon after birth.

By the age of two months, infants can detect voice–lip synchrony during speech. By five to seven months, they can use face–voice synchrony to selectively attend to a stream of speech, separating it from a background of other sounds, and pick out individual words (speech segmentation) much better than in auditory stimulation alone. Audiovisual synchrony is also important for learning the arbitrary relation between speech sounds and the objects to which they refer, a cornerstone of language development. For example, by seven months, infants can learn to pair a speech sound with an object, provided the sound is uttered in synchrony with moving and showing the object. However, it is not until many months later that infants can learn to pair a speech sound with an object in the absence of synchrony. Parents are intuitively aware of this need and synchronize naming and showing objects to their younger infants.

### *Audiovisual Interactions and Illusions*

Several intriguing audiovisual illusions illustrate the powerful interaction of auditory and visual perception. For example, when we view the face of a person speaking one speech sound (e.g., “ga”), while hearing a different speech sound (e.g., “ba”), we perceive another sound altogether (“da,” a blend between the two). Infants show this *McGurk effect* within the first six months following birth, illustrating that infants, like adults, merge information for speech across the senses. This is also illustrated by the *ventriloquist effect*. By moving a dummy's mouth in synchrony with the ventriloquist's own speech sounds, the ventriloquist creates amodal information and fools us into perceiving that the dummy is actually speaking. In other words, although sound is coming from one location, it is “heard” as coming from another location.

### **Development of Visual-Tactile and Visual-Motor Perception**

Amodal information also guides and constrains perception across vision and touch. For example, when we feel an object with our hands, we can perceive the same shape, size, texture, and substance that we see. One-month-old infants (and in one study, even newborns) perceive a correspondence between an object they have explored tactually (by

holding or mouthing) and a visual display of the same object and prefer to look at the familiar shape. Tactile exploration and active touch are coordinated with visual exploration and develop rapidly across the first year, providing increasingly more detail about objects and their surfaces.

Proprioception is information about self-movement based on feedback from the muscles, joints, and vestibular (balance) system. This important information about the self is coordinated with motor behavior and visual feedback from motion and develops rapidly over the first year of life. By three to five months, infants can distinguish between self and other on the basis of this information. They discriminate between a live video of their own legs kicking versus those of another infant by detecting temporal synchrony and spatial correspondence between proprioceptive feedback and the visual display of their own leg motion. Infants also show *visually guided reaching*, even in the first months of life. This entails continuous adjustments in manual and reaching behavior as a function of the changing position of objects, and this intercoordination becomes more precise with experience. A little later in development, infants can adapt their crawling and early walking attempts based on visual information about the slope and solidity of the surface. Visual feedback is essential for crawling, walking, and maintaining an upright posture at all stages of development, illustrating the essential link between vision, proprioception, and motor behavior.

### Odor and Vision

Young infants show recognition of their mother on the basis of her smell, and breastfed infants prefer the odor of their mother over that of another lactating woman, illustrating their sensitivity to object-odor combinations. In one study, four-month-olds looked more to an object that had previously been paired with a cherry odor when the odor was present, but not when it was absent, showing memory for the relation between the visual appearance of an object and its odor. In many species, recognition of the mother may be promoted by fetal sensitivity to the odor and taste of amniotic fluid and its similarity to other maternal odors.

*Lorraine E. Bahrnick and Robert Lickliter*

*See also* Amodal Perception; Attention: Cross-Modal; Binding Problem; Direct Perception; Infant Perception; Multimodal Interactions: Visual–Auditory; Perceptual Development: Face Perception; Perceptual Development: Hearing; Perceptual Development: Object Perception; Perceptual Development: Taste and Olfaction; Perceptual Development: Touch and Pain

### Further Readings

- Bahrnick, 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.
- Gibson, E. J., & Pick, A. D. (2000). *An ecological approach to perceptual learning and development*. New York: Oxford University Press.
- Gibson, J. J. (1966). *The senses considered as perceptual systems*. Boston: Houghton Mifflin.
- 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. (1994). *The development of intersensory perception: Comparative perspectives*. Hillsdale, NJ: Lawrence Erlbaum.
- Rosenblum, L. D. (2005). Primacy of multimodal speech perception. In D. Pisoni & R. Remez (Eds.), *Handbook of speech perception* (pp. 51–78). Malden, MA: Blackwell.

---

## PERCEPTUAL DEVELOPMENT: OBJECT PERCEPTION

---

When looking around the kitchen, most adults have no trouble determining that the toaster is separate from the blender, even if the surfaces of the two are touching. We can determine where one object ends and another begins without careful thought or effort. Is the same true for infants early in the first year of life? If so, which cues do they use to make such judgments and at what point in development are infants capable of making these judgments accurately?

Some of the first observations that provided insight into how infants perceive objects were published by Jean Piaget, who is widely regarded as one of the founders of the field of cognitive development.