8

The Relation of Hearing to Other Senses

Books on perception usually concentrate on a single modality (such as vision or hearing) or a subdivision of a modality (for example, color vision or speech perception). Even when an introductory book on perception deals with several modalities, it is generally subdivided into sections with little overlap. The few books treating the senses together as a single topic generally emphasize philosophy or epistemology (but see Gibson, 1966; Marks, 1978).

Yet the senses are not independent. Events in nature are often multidimensional in character, and stimulate more than one sensory system. An organism which optimizes its ability to interact appropriately with the environment is one which integrates relevant information across sensory systems.

MULTIMODAL PERCEPTION

While speech perception may appear at first to be strictly an auditory task, this is not quite true. Lip reading (or speech reading) can play an important subsidiary role. But before dealing further with this interaction of modalities in speech perception, let us consider cross-modality interactions involving senses which do not include hearing.

Interaction of Vision with Senses Other than Hearing

Depth perception in vision is based on a number of cues, including disparity of the images at the two retinae, and motion parallax. But, in addition to these cues transmitted by the optic nerve, there are proprioceptive ocular cues from the muscles producing accommodation (changes in the curvature of the lens necessary to produce a sharp image) and convergence (adjustment of the ocular axes so that the fixated object is imaged on each fovea). These cues originating in muscles are integrated with, and are indistinguishable from, the purely optical cues to depth. For a person perceiving an object at a particular distance, these muscle-sense cues are as fully visual as those providing information via the optic nerve.

Another example of visual interaction is afforded by the vestibulo-ocular reflex, which helps the viewer maintain visual fixation that might otherwise be lost during head movements. Receptors in the semicircular canals found in the inner ear signal rotary acceleration of the head, and this information is transmitted by the vestibular branch of the auditory nerve. Reflex eye movements are produced by this stimulation in a direction that compensates for the effects of head movements. When the vestibular and ocular systems are operating normally in a coordinated fashion, we are unaware of the contribution of vestibular input to the stability of our visual world.

Graybiel and his associates have studied the effect of vestibular information on visual perception in a quantitative fashion using the "oculogyral illusion." This illusion can be observed when a person is accelerated in a rotating chair while viewing a dim star-like pattern in an otherwise dark room, with the position of the visual pattern being fixed relative to the head of the subject. At very low acceleration rates which would be imperceptible with the eyes closed ($0.12°/\text{sec}^2$) the visual target appears to be displaced in the direction of acceleration (Graybiel, Kerr, & Bartley, 1948), so that the only way of detecting acceleration is through the effect of vestibular input on vision. It appears as if the information from the vestibular branch of the VIIIth nerve and the IInd (optic) nerve are integrated into a single percept, with vestibular contributions leading to effects interpreted as strictly visual.

Interesting effects are observed when conflicts are generated between information furnished by the vestibulo-auditory nerve and the optic nerve. If a person with closed eyes is spun rapidly (say at 180°/sec) in a rotating chair for a few minutes and then brought to rest in a few seconds, disruption of visual perception by vestibular after-effects takes place when the eyes are opened. Involuntary eye movements which occur following the vestibular stimulation cause images of stationary objects to move continuously over the retina. This vestibulo-ocular reflex leads to illusory movement of the visual world, which is coupled sometimes with rather unpleasant feelings of vertigo and nausea. Yet, ice-skaters do not experience such after-effects following very rapid rotation and a sudden stop from a practiced spin. The vestibulo-ocular reflex and illusory motion can not only be suppressed by appropriate training, but it also can be changed in direction. It has been reported that, after several weeks of continuous wearing of goggles which were fitted with prisms reversing right and left, the direction of the vestibulo-ocular reflex was reversed, so that rotary acceleration produced eye movements which could compensate for the effect of head movement upon the retinal image as seen through the prisms (Gonshor & Jones, 1973).

While vision appears to be dominant in cross-modality integration with vestibular information, vision does not exhibit dominance when it interacts with auditory information in the perception of speech.

Interaction of Vision and Hearing in Speech Perception

Research in the last few years has indicated that people with normal hearing use speech reading (lip reading) to a greater extent than had been realized previously. Visual information concerning what is being said seems to be heard rather than seen; that is, it enters into determining what we believe we hear the speaker say.

Comprehension based on speech reading alone is very difficult, although some deaf people can understand what a speaker is saying through close visual observation. While lip movements are considered the most important cues, some articulatory information may be furnished by movements of the jaw and Adam's apple. In addition, there are a variety of facial expressions and other gestures correlated with meaning but not related directly to articulation. Nevertheless, speech reading, at best, provides incomplete and ambiguous information concerning articulation. Some speech sounds are produced by articulatory movements which cannot be seen (for example, /h/ and /k/), and some sounds with articulatory movements which can be seen have "homophenes" with the same appearance (/m/, /p/, and /b/ involve similar lip movements, so that "may," "pay," and "bay" all look alike).

Recent experiments have indicated that speech reading can function not only as an alternative mode for perceiving speech by the deaf, but that visual cues to articulation provide a supplementary source of information used by people with normal hearing to enchance intelligibility under noisy environmental conditions (for example, see Dodd, 1977, 1980). Speech reading can be facilitated even by an auditory signal which is itself completely unintelligible. Rosen, Fourcin, and Moore (1981) used listeners with normal hearing and compared their ability to perceive what a speaker said using speech reading alone and speech reading supplemented with acoustic input corresponding to the fundamental frequency pattern of the speaker's voice. While the voice-pitch information by itself was unintelligible, it was found to produce a dramatic increase in the ability to understand what the speaker was saying when used in conjunction with speech reading.

Perceptual Resolution of Conflicting Visual and Auditory Information Concerning Speech

An interesting interaction occurs when visual and auditory information to speech are placed in conflict. McGurk and MacDonald (1976) in a paper entitled "Hearing Lips and Seeing Voices," used a video recording of a talker producing a consonant-vowel syllable while a dubbed sound tract showed production of a different consonant-vowel syllable. For example, when the sound of *ba-ba* was dubbed onto a video recording of *ga-ga*, 98 percent of adults and 80 percent of preschool children reported hearing *da-da*. It should be noted that there was not any awareness of a visual contribution to perception—listeners believed that the illusory *da-da* was heard and was solely auditory. The illusion occurred even with the knowledge of the nature of the visual and auditory input, and by closing their eyes, subjects could hear *ba-ba* which reverted to *da-da* when their eyes were opened. It is as if the modality serving as the source of the information were irrelevant to the perceptual task of determining the nature of the acoustic utterance, and so is not perceived directly.

A conflict between vision and hearing in speech perception leads to the pooling of cues and the "hearing" of visual information, but this type of conflict resolution does not take place with other tasks in which vision is pitted against hearing. When visual and auditory cues to the localization of a source are in disagreement, vision tends to dominate, and the sound usually is heard to come from the position close to that indicated by vision (see Bertelson and Radeau, 1981).

AUDITORY INPUT PERCEIVED AS TOUCH

An interesting example of an inability to determine the modality furnishing perceptual information is provided by "facial vision" of the blind. Obstacles can be avoided by some sightless people through echoes reflected from surfaces, yet they usually are unaware that they are using hearing, and frequently attribute the detection of obstacles to sensitivity of their face and forehead. Acoustic cues are perceived by them as "pressure waves" stimulating their skin, producing a sensation which gets stronger and assumes an unpleasant quality if they continue on a collision course (see Supa, Cotzin, & Dallenbach, 1944; Worchel & Dallenbach, 1947). Indeed, if the blind do not heed the auditory information indicating their approach to an obstacle, they might suffer actual injury to the head.

The inability of the blind to appreciate that hearing serves as the basis for their obstacle sense and their false perception of auditory input as tactile stimulation seem anomalous only if we consider that people can appreciate directly the nature of sensory input used for perceptual evaluation. However, this misattribution is consistent with the hypothesis that we are aware of events correlated with sensory input rather than sensation per se.

MULTIMODAL SENSORY CONTROL OF SPEECH PRODUCTION

The level of our vocal output is monitored not only by hearing, but also by nonauditory information including "vocal effort" as signified by proprioceptive feedback from the muscles involved in controlling subglottal pressure and laryngeal pulsing (see chapter 7), and tactile cues to the magnitude of vibration along the vocal tract. Normally, these cues are in agreement, and each supplements and confirms the information provided by the others. Interesting effects occur when unusual conflicts are introduced (as will be discussed shortly). However, blocking, or at least partial blocking, of one of the monitoring systems (such as the masking of auditory feedback by noise) does not produce the interference associated with conflicting cues, since other monitoring systems can be used to maintain normal functioning.

This multimodal information concerning the level of our voice alone is insufficient to allow us to speak at an appropriate intensity. It is inappropriate to speak softly to someone far away, and a listener might show annoyance were we to shout in his ear. It would be unsuitable to use the same vocal level speaking to someone at a cocktail party that we would use when speaking to someone at the same distance in a quiet room. We take these factors into account automatically when speaking. Thus, it has been demonstrated that we are quite familiar with the effects of the inverse square law, and that we can adjust the level of our voice to compensate quite accurately for intensity changes at the listener's position due to the distance our voice must carry to reach its target (Warren, 1968a; see chapter 4). Visual cues play an important role in estimating this distance.

In addition to using hearing to monitor our own vocal level, we also use hearing to evaluate the ambient noise level. The increase in vocal intensity accompanying an increase in background noise is called the Lombard reflex, and it helps ensure that we can be heard by our targeted listeners. It is only after the distance to the target listener and the ambient noise level are determined, that the intensity required for comprehension by the listener can be reckoned and compared with the feedback from the several sensory systems monitoring our own voice.

GENERAL PERCEPTUAL RULES AND MODALITY-SPECIFIC RULES

The close interrelation of senses might lead us to believe that rules governing perception in one modality would apply to others as well. While it does seem that such general rules do exist, attempts to apply broad principles either across modalities or to different tasks within single modalities should be carried out with considerable caution.

The physical correlate theory is a case in point. As discussed in chapter 4, this theory considers that attempts to measure sensory intensity directly produce responses based on estimates of physical magnitudes. However, this general rule cannot tell us which physical magnitude serves as the correlate of intensity judgments involving a particular sensory system without detailed knowledge of that modality and the way it is used normally to evaluate stimulus relations.

Premature analogies across senses can inhibit the understanding of phenomena. An example is furnished by the perceptual transformations experienced in both hearing and vision with prolonged stimulation by a fixed pattern. During visual inspection of a two-dimensional ambiguous figure, such as an outline drawing of a cube, perception of the figure can flip from one perspective interpretation to another. It is tempting to consider that illusory changes in repeated words (see chapter 7) represent an auditory analog of visual reversible figures, and it appears that Warren and Gregory (1958) were led astray by this analogy. Evans (see Warren, 1981b) considered that these verbal transformations were an auditory analog of illusory changes experienced with stabilized retinal images (that is, images seen without the normal changes in retinal stimulation accompanying our ever-present eye movements), and he too may have been misled by an inappropriate analogy. Warren (1981b) has suggested that change in perceptual organization during continued stimulation represents a general perceptual phenomenon operating across modalities, but that the nature of perceptual reorganization reflects special strategies for perception which differ across modalities, with different strategies being used even for different classes of tasks within a single modality. Thus, it has been hypothesized that verbal transformations represent highly specialized reorganizational strategies employed in speech perception, and that repeating nonverbal auditory patterns are not subject to analogous changes (see chapter 7).

PERCEPTUAL CALIBRATION OF SENSORY INPUT

Verbal transformations and illusory changes in unchanging visual displays occur *during* a continued exposure to the same stimulus. There are other perceptual changes which can be measured *following* exposure to a stimulus. These poststimulation changes follow a consistent rule applying to many types of judgments involving different sensory modalities. The rule can be stated as follows: the perceptual classification of a stimulus occupying a position along a continuum is shifted following exposure to an exemplar occupying a different position along the continuum, so that judgmental boundaries move toward the value represented by the exemplar. Let us clarify this "perceptual recalibration" rule by examples.

Perceptual recalibration can be observed for speech perception. As discussed in chapter 7, the time separating plosive release of air and the onset of voicing can determine whether a syllable is heard as /ta/ or /da/. The voice onset time corresponding to this category boundary shifts noticeably following listening to either /ta/ or /da/ restated several times. Thus, after listening to exemplars of /ta/, the category boundary moves into what previously was the /ta/ domain, and a sample which had been at the category boundary is now clearly heard as /da/. It is not necessary to use category boundaries separating phonemes to observe such perceptual shifts. Remez (1979) reported that the perceptual boundary for a vowel in a continuum extending from the vowel to a nonspeech buzz could be changed in an analogous fashion by exposure to an exemplar at one end of the continuum. Similar effects of prior stimulation were described for judgments along a variety of visual continua by Gibson, one such continuum being that of visual curvature. Gibson observed that if a curved line convex to the left was examined for some time, a line presented subsequently had to be curved in the same direction to appear straight, and an objectively straight line appeared convex to the right. Gibson called this phenomenon "adaptation with negative after-effect," and described it in terms of the following general rule: "If a sensory process which has an opposite is made to persist by a constant application of its appropriate stimulus-conditions, the quality will diminish in the direction of becoming neutral, and therewith the quality evoked by any stimulus for the dimension in question will be shifted temporarily toward the opposite or complementary quality" (Gibson, 1937). Similar errors in judgment were cataloged for a number of after-effects of seen motion by Wohlgemuth (1911), who reported that the visual input corresponding to perception of a stationary display was subject to temporary recalibration in the direction of previously perceived movement. Thus, after viewing a moving display, a new display must move slightly in the direction previously seen in order to appear stationary, so that an objectively stationary display is seen to move in the opposite direction. Related observations with other perceptual continua were described in 1910 by von Kries (1962, p. 239) who attributed these errors to what he called "the law of contrast." Cathcart and Dawson (1928, 1929) also discovered this rule governing perceptual shifts. They recognized its very broad applicability, naming it the "diabatic" effect. Helson (1964) tried to quantify this general principle with his "adaptation level theory," but this theory has had a rather limited success. There have been attempts to explain perceptual shifts produced by prior stimulation in terms of the adaptation of hypothetical neural feature detectors, but this approach, at least applied to phoneme boundary shifts, has come under considerable criticism (see chapter 7).

Perhaps these perceptual after-effects reflect a relativistic basis for judgments, with the criteria used for perceptual evaluation constructed from past exemplars in such a way that higher weighting is assigned to recent exemplars. Thus, it may not be necessary that neural adaptation or fatigue take place for perceptual judgments to change. A shift or recalibration of criteria used for evaluation could be responsible for judgmental shifts. In keeping with this model, it can be considered that a consistent change in sensory patterning maintained for brief periods leads to perceptual recalibration with temporary "after-effects" upon return to the previous norm as described above, while extended exposure to alteration in sensory input leads to the adoption of more permanent perceptual norms consistent with the new conditions. (See the discussion of adjustment to the spatial rearrangement produced by "pseudophones" in chapter 2.)

There is evidence that "sensory deprivation" or absence of patterned sensory input can lead to a breakdown in normal perceptual processing (see Riesen, 1975; Schultz, 1965; Vernon, 1963). When subjects see only a featureless white field through translucent goggles and hear only white noise, disorientation along with visual and auditory hallucinations occur within a short time. After a few hours of such sensory deprivation, subjects experience errors in perceptual evaluation when returned to a normal environment. Perhaps calibration of sensory input occurs continuously, and is required to maintain appropriate perceptual interpretation of environmental events.