

References

- Goldsmith, J. (1990). *Autosegmental and Metrical Phonology*. Oxford: Blackwell.
- Goldsmith, J. (1994). Tone languages. *Encyclopedia of Language and Linguistics*. Pergamon Press.
- Goldsmith, J., Ed. (1995). *The Handbook of Phonological Theory*. Oxford: Blackwell.
- Hyman, L. (1978). Tone and/or accent. In D. J. Napoli, Ed., *Elements of Tone, Stress, and Intonation*. Washington, DC: Georgetown University Press, pp. 1–20.
- McCawley, J. (1978). What is a tone language? In V. A. Fromkin, Ed., *Tone: A Linguistic Survey*. New York: Academic Press, pp. 113–131.
- Odden, D. (1995). Tone: African languages. In J. Goldsmith, Ed., *The Handbook of Phonological Theory*. Oxford: Blackwell, pp. 444–475.
- Pike, K. (1948). *Tone Languages: A Technique for Determining the Number and Type of Pitch Contrasts in a Language, with Studies in Tonemic Substitution and Fusion*. University of Michigan Publications in Linguistics, no. 4. Ann Arbor: University of Michigan Press.
- Yip, M. (1995). Tone in East Asian languages. In J. Goldsmith, Ed., *The Handbook of Phonological Theory*. Oxford: Blackwell, pp. 476–494.

Further Readings

- Beckman, M., and J. Pierrehumbert. (1986). Intonation structure in Japanese and English. *Phonology Yearbook* 3: 255–309.
- Bolinger, D. (1985). Two views of accent. *Journal of Linguistics* 21: 79–123.
- Duanmu, S. (1996). Tone: An overview. *Glott International* 2: 3–10.
- Fromkin, V., Ed. (1978). *Tone: A Linguistic Survey*. New York: Academic Press.
- Hyman, L., and R. Schuh. (1974). Universals of tone rules: Evidence from West Africa. *Linguistic Inquiry* 5: 81–115.
- Pulleyblank, D. (1986). *Tone in Lexical Phonology*. Dordrecht: Reidel.
- van der Hulst, H., and N. Smith, Eds. (1988). *Autosegmental Studies on Pitch Accent*. Dordrecht: Foris.

Top-Down Processing in Vision

Perception represents the immediate present, what is happening around us as conveyed by the pattern of light falling on our RETINA. And yet the current pattern of light alone cannot explain the stable, rich experience we have of our surroundings. The problem is that each retinal image could have arisen from any of a vast number of possible 3-D scenes. That we rapidly perceive only one interpretation tells us that we see far more than the immediate information falling on our retina. The highly accurate guesses and inferences that we make rapidly and unconsciously are based on a wealth of knowledge of the world and our expectations for the particular scene we are seeing. The influences of these sources beyond the images on the retina are collectively known as top-down influences.

Both top-down analyses and the complementary bottom-up processes use local cues to assign depth to the regions of an image. They differ in the manner in which they resolve the ambiguity of the local cues. A bottom-up analysis, part of MID-LEVEL VISION and SURFACE PERCEPTION, makes direct



Figure 1.

links between local geometrical features and depth. For example, whenever one object partially covers another, the visible contours of the more distant object terminate at the outer boundary of the nearer one, forming what are called T-junctions. When a T-junction is encountered in an image, this logic can be reversed: the stem of the T is designated a contour of a more distant, partially hidden object and the top of the T is assigned to the outer boundary of a nearer object.

A top-down process, on the other hand, depends on the content of the image and its analysis by processes of HIGH-LEVEL VISION. Cues operate by suggesting objects—a nose contour might suggest a face, for example—and then stored information about that object's structure can be applied to the assignment of depth in the image. Other features in the image are then examined to verify or reject the postulated object. The cues used for the initial selection of potential objects are not limited to the current images but include preceding images as well as nonvisual sources which affect our expectations for the scene. The sources of object knowledge which are called upon may be built up over both evolutionary or individual time scales.

Our guesses for appropriate internal models are best when we know what to expect in a scene. Upon opening a door to a classroom, for example, we expect to see desks and a black or white board. If these elements are present in the scene, they are rapidly interpreted. Incongruent elements are seen less reliably as Biederman (1981) showed when he reported increased errors in identifying fire hydrants presented in kitchens or sofas floating over city streets than when they were presented in their usual contexts. As Biederman's example demonstrates, top-down analyses work because there is a great deal of semantic redundancy in the content of a scene—noses are expected to be seen along with mouths, cars with roads, classrooms with desks, and sofas with coffee tables; moreover, noses, cars, and sofas have typical shapes so that once a few distinctive features have implied the presence of say, a car, the other expected features of a car can be verified or even just assumed to be present.

Textbook examples of top-down processing typically make use of images with two or more equally likely interpretations which are sometimes referred to as ILLUSIONS. A hint as to which interpretation to see may then trigger one or the other, as in the examples shown here. (a) Two faces, or one vase, or one face behind a vase (Costall 1980); (b) a man playing a saxophone seen in silhouette, or a woman's face in sharp shadow (Shepard 1990); and (c) a sphere in a four-point setting or a white angel (Tse 1998). In these instances,

the 2-D positions of light and dark values are unchanged as we alternate our percepts, but new positions in depth are assigned to each point, some areas change from being dark shadow to dark pigment, and some regions change from being disconnected surfaces to continuous pieces.

Where do these new assignments come from when the 2-D pattern is the same in all cases? We cannot invoke a bottom-up analysis of the depth cues in the image since they would be inconclusive (insufficient to unambiguously assign depth). For some of the examples above we have to be told what to see before the image becomes organized as the intended 3-D object. On the other hand, some of us see some of the interpretations spontaneously, implying that some characteristic features in the image have suggested a familiar object (a nose outline or eye-like shape could suggest a face) and our visual system then matched a possible 3-D version of such an object to the image. In both cases, our final perception is arrived at through the intermediate step of a guess or a suggestion of a possible object.

Once the presence of an object has been verified, our knowledge of that object can continue to constrain the interpretation of otherwise ambiguous dynamic changes to the object. For example, Chatterjee, Freyd, and Shiffrar (1996) have shown that the perception of ambiguous apparent motion involving human bodies usually avoids implausible paths where body parts would have to cross through each other.

Undoubtedly, the process of top-down matching of a candidate object to the image data occurs for natural images, not just the highly artificial ones shown in the figures above. Because of the extra information present in natural images, it is rare to have two alternative interpretations available. Nevertheless, the speed with which we organize and perceive the world around us arises to a great extent from the excellent (top-down), unconscious guesses we make based on sparse cues coming from either the actual or the expected content of the retinal image.

See also ATTENTION; DEPTH PERCEPTION; FACE RECOGNITION; FEATURE DETECTORS; GESTALT PERCEPTION

—Patrick Cavanagh

References

- Biederman, I. (1981). On the semantics of a glance at a scene. In M. Kubovy and J. Pomerantz, Eds., *Perceptual Organization*. Hillsdale, NJ: Erlbaum.
- Chatterjee, S. H., J. J. Freyd, and M. Shiffrar. (1996). Configural processing in the perception of apparent biological motion. *Journal of Experimental Psychology: Human Perception and Performance* 22: 916–929.
- Costall, A. (1980). The three faces of Edgar Rubin. *Perception* 9: 115.
- Shepard, R. (1990). *Mind Sights*. New York: W. H. Freeman.
- Tse, P. (1998). *Volume*. Unpublished ms., Harvard University.

Transparency

The light projecting to a given point in the RETINA possesses but a single value of color and intensity. When *transparency* is perceived, however, this light is interpreted as being

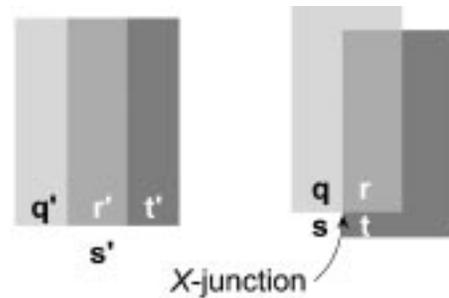


Figure 1.

reflected off of two (or sometimes more) surfaces lying in different depth planes. Perceptual transparency is a type of SURFACE PERCEPTION and illustrates the visual system's remarkable ability to reconstruct the three spatial dimensions of the environment given a stimulus (i.e., the retinal images) with only two. There are an infinite number of possible environmental causes of any particular pattern of retinal stimulation. The perception of transparency relies, as does visual perception in general, upon *context* to determine the most likely interpretation. For example, whereas region r' in figure 1 (*left*) is usually interpreted in terms of the color of a single surface, region r (*right*), an identical shade of gray, is seen to arise from light reflected off of two surfaces. This difference in perceptual interpretation is due to the presence of a contextual cue known as an *X-junction*.

X-junctions are the single most important monocular cue for transparency. They are defined by the presence of four contiguous regions (q, r, s, t ; see figure 1) of an image with a characteristic spatial arrangement. Psychophysical studies have shown that the intensity relationships between these four regions must lie within certain bounds for perceptual transparency to occur. When X-junctions elicit a perception of transparency, two regions (q and s in figure 1) are seen as differently colored parts of the unoccluded background and the other two regions (r and t in figure 1) appear to be viewed through a foreground transparent surface (the darker rectangle). Perceptual psychologists have developed several simple physical models to account for the perception of transparency (e.g., Beck et al. 1984; Matelli 1985). Though differing slightly in their details, the optical properties of *transmittance* and *reflectance* are generally invoked in these models.

Transmittance refers to the multiplicative attenuation of background intensity. One way to think of transmittance is to imagine that transparent surfaces are generally opaque but have holes (like a fine wire mesh) too small to resolve (Kersten 1991; Richards and Witkin 1979; Stoner and Albright 1996). Transmittance is then the proportion of the surface with holes. *Reflectance*, on the other hand, refers to the fraction of incident light reflected off of a surface. If the surface is a foreground transparent surface, this light adds to that reflected off of the background surface. X-junctions that elicit a sense of transparency are usually those in which the four sub-regions possess intensities consistent with physically realizable values of transmittance and reflectance, giving credence to the idea that the visual system possesses a tacit model of the physics of transparency. Given