CHAPTER 6

Summary

Throughout, I have made the argument that we feel that perceiving is effortless and accurate regardless of context. But, it is difficult to understand how "getting it right" is accomplished. Perceiving things depends on the simultaneous integration of many physical attributes and the transformation of the resulting neural signals in the central nervous system. In fact, I often wonder how it is possible to perceive at all. In the end, it is a "best estimate" based on the input sensory signals that are modified by top-down neural control processes, context, and past experience. Instead of a static word like perception, we should use a dynamic word like perceiving to emphasize that we are constantly adapting our actions and knowledge to a changing world.

In this short summary, we return to the basic problem of constructing objects and sources in extended space and time. To perceive those objects and sources, we first need to identify what parts of visual, auditory, and tactual sensations go with each object or source, that is, which are the connected points. Edges and gaps, both physically present and inferred, bound objects and sources and separate them; edges belong to the objects and sources, and place figures in front of grounds. Temporal synchrony and rhythmic grouping bind sounds into sources and allow for the hearing of simultaneous sources even when the sensations from each are mashed together. But as the apparent movement and Ternus displays show, changes in either spatial arrangements or in the timing between the displays alter the other dimension. Local processes must be integrated into global ones.

Not only do we have to construct the objects, surfaces, and sources, we have to track them over time as they transform and emerge in different forms and backgrounds. In some cases we can solve the correspondence problem by picking up the predictable changes arising from physical properties. On the whole, physical changes are correlated with one another, allowing us to distinguish among alternative possibilities. If the vertical and horizontal dimensions of a square change equally, for example, that suggests back or forth movement. But

© The Author(s) 2019
S. Handel, *Perceptual Organization*,
https://doi.org/10.1007/978-3-319-96337-2_6

if only one dimension changes, that suggests rotation. Correlated or redundant changes reduce errors, since one change can lead to the correct percept even in the absence of all others.

In other cases, there is no simple solution to the correspondence problem. Consider face recognition, all faces have the same parts (e.g., foreheads, eyes, noses, lips, etc.) and the configuration is the same, eyes above noses, noses above lips, and so on. The problem then, is to identify specific individuals amidst many similar others (Behrmann, Richler, Avidan, & Kimchi, 2015). The prevalent naïve view is that people are very good at it in spite of the numerous cases of false eyewitness identifications. What is actually true is more complex, and it bears on all the issues discussed here. The task is simple: people view one photograph of an individual (even for extended periods of time), and then are asked to identify that individual in a different pose (i.e., different haircut, shadows, facial expression, gaze direction) among a set of photographs of different individuals. People are very good at identifying familiar faces, but are quite poor at identifying unfamiliar faces. In fact, for most people even extensive training does not improve performance for novel faces. Young and Burton (2017) suggest that having seen a familiar face in many different poses, we have built up a model of how the face will look in still other poses. In other words, for that face we have learned to separate the stable characteristics from the variable ones and have learned the transformations that connect all the possible poses. But the variability and resulting transformations will differ for each face, so that without extensive experience with each new face, people are unable to match those poses.

Since the variability associated with each face is idiosyncratic it is impossible to create a general transformation that will allow us to identify any face in a new pose. We cannot solve the correspondence problem without observing the poses of each face separately. This outcome is not unique to faces but is true for the discrimination of tonal and atonal melodies discussed in Chap. 2, the discrimination of instruments at different pitches discussed in Chap. 5, the identification of singers at different notes discussed in Chap. 5, and the detection of human movements discussed in Chap. 2. A good parallel is the identification of instruments at different pitches taken up in Chap. 5. Musicians who are experienced with one instrument are able to identify that instrument at different pitches, but they are not able to do so for unfamiliar instruments. Each instrument changes timbre in idiosyncratic ways across their playing range, so experience with one instrument does not transfer to other instruments.

Perceiving occurs at the same time across multiple temporal and spatial levels that are interlocked. Small visual spatial regions are grouped into larger more encompassing areas, small uniform surface regions become part of larger varying surfaces, and individual sounds and beats turn into rhythmic meters that organize time. This ability to see, touch, and hear a sequence at different time intervals and spatial distances yields a hierarchical representation in which the levels are not independent, but are constrained by each other. They are not

like Russian Dolls in which independent, fully formed smaller dolls fit inside independent, fully formed larger ones.

Rhythms create multilayered figures in time and space. One can zoom in and out. When listening it is possible to react to every beat or to beats widely separated in time. Without the faster beats, the slower beats are simply recurring accents, and without the slower beats, the faster ones also are simply recurring accents. Meter is the mental construct that fuses the beats at different levels together and organizes time. In similar fashion, when looking it is possible to view a narrow spatial field or a wide one. Viewers construct layers to group the visual regions into larger, more encompassing areas containing overlapping objects. But, the auditory meter and the visual space do not simply match the auditory or visual sensory features. The meter and space emerge from the interaction of those sensory features and the perceptual acts.

Another compelling example of how these interacting multiple layers create our perceptual world is demonstrated by our ability to abstract the movements of parts of objects moving in different trajectories and at differing speeds as shown in the videos by Johansson (1973) in Chap. 2 and Toiviainen, Luck, and Thompson (2010) in Chap. 4. The slower components act as frames of reference for the faster ones, so that it is possible to partition the movement of the faster components into two parts: One part common to the slower component that "stays" with the slower one, and one part that is unique to the faster component and is seen separately. The perception of auditory meter based on beats at differing rates, and the perception of visual motion based on movements at different speeds, is therefore based on the same concept of hierarchical layers. I am strongly drawn to this conceptualization. It makes theoretical questions such as whether perception is bottom-up or top-down largely irrelevant, and it also seems to make distinctions about brain localization unlikely to be useful in explaining what we see, feel, or hear.

Ultimately, there is no single answer to the basic questions of "why do we see, hear, and touch what we see, hear, and touch" or "why do objects and sources look, sound, and feel the way they do." Furthermore, the sensory and perceptual processes are not linear or unidirectional: it is not the case the A leads to B, B leads to C, and so on. The processes always interact, the feedback from later stages modify the outputs from earlier ones. The receptors at the eye, ear, and hand transform the sensory energy into neural signals. Those signals are further modified at higher brain regions before reaching the cortex. The neural information is underdetermined; many external objects could have generated the same sensory energy and the transformed neural signal. The cortical firings are interpreted in terms of their coherence and organization, and the person's expectations (the prior probabilities in Bayesian terms), to give rise to the percept. None of this is simple or straightforward; that's what makes perceiving so surprising and interesting.

References

Behrmann, M., Richler, J., Avidan, G., & Kimchi, R. (2015). Holistic face perception. In J. Wagemans (Ed.), *The Oxford handbook of perceptual organization* (pp. 758–774). Oxford, UK: Oxford University Press.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics, 14*, 201–211.

Toiviainen, P., Luck, G., & Thompson, M. R. (2010). Embodied meter: Hierarchical eigenmodes in music induced movement. *Music Perception, 28*(1), 59–70. https://doi.org/10.1525/mp.2010.28.1.59

Young, A. W., & Burton, A. M. (2017). Recognizing faces. *Psychological Science, 26*, 212–217. https://doi.org/10.1177/0963721416688114