that sounds can arrive with a delay of around 0.5 ms between the two ears when the sound is presented off to one side. Similar processing mechanisms that make use of small timing differences could also apply in other sensory pathways.

In general, it is clear that the speed with which the brain can process stimuli is absolutely critical for survival, whether you are predator or prey. It is remarkable that the brain, with neurons that fire, at most, a few hundred times per second and with conduction velocities for nervous transmission that are so slow, can compete with and even outperform the most sophisticated artificial vision systems.

Simon J. Thorpe

See also Consciousness; Neural Representation/Coding; Priming; Rapid Serial Visual Presentation

Further Readings

Hegde, J. (2008). Time course of visual perception: Coarse-to-fine processing and beyond. *Progress in Neurobiology, 84*(4), 405–439.

Kiani, R., Esteky, H., & Tanaka, K. (2005). Differences in onset latency of macaque inferotemporal neural responses to primate and non-primate faces. *Journal of Neurophysiology, 94*(2), 1587–1596.

Kirchner, H., & Thorpe, S. J. (2006). Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. *Vision Research, 46*(11), 1762–1776.

Mormann, F., Kornblith, S., Quiroga, R. Q., Kraskov, A., Cerf, M., Fried, I., et al. (2008). Latency and selectivity of single neurons indicate hierarchical processing in the human medial temporal lobe. *Journal of Neuroscience, 28*(36), 8865–8872.

Quian Quiroga, R., Reddy, L., Kreiman, G., Koch, C., & Fried, I. (2005). Invariant visual representation by single neurons in the human brain. *Nature, 435*, 1102–1107.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381*(6582), 520–522.

STATISTICAL LEARNING

It is natural to think of perception in terms of the processing of individual features (such as color and shape) and how they are combined into discrete objects (such as animals and bicycles). This simple characterization underestimates the information that is available in perceptual input, though, because there are also massive amounts of information about how these features and objects are *distributed* in space and time. In time, for example, eating food at a restaurant is more likely to be followed by paying a bill than by climbing a tree—just as (in English) the syllable /sci/ is more likely to be followed by /ence/ than by /on/. And in space, for example, a car is more likely to be next to a bicycle than to a stapler. Discovering such regularities is difficult because they are embedded within complex and continuous environments where not all information is relevant. But the mind is nevertheless sensitive to such regularities, uncovering them in part by means of *statistical learning:* an automatic and unconscious perceptual process that encodes statistical regularities across space and time. We are often unaware of the operation of statistical learning, yet it may play a crucial role in segmenting the continuous perceptual world into discrete manageable units, such as words, events, objects, and scenes. This entry describes statistical learning, how and when it operates, and how it may support online perception.

Segmenting the World via Statistical Regularities

Statistical learning (henceforth abbreviated SL) is a type of *implicit* learning, in that it can occur without intent or awareness. Indeed, like many other aspects of perception, it occurs beneath the level of consciousness, and we are not even normally aware that our minds are engaging in such learning at all. Implicit learning has a long history in psychology, but the study of SL in perception research has arisen only in the last 15 years, as researchers have focused on how such processing may serve to segment continuous perceptual input into discrete units.

Auditory Statistical Learning

In its modern incarnation, the study of SL began in the domain of language acquisition. When you first hear a foreign language, you do not understand the words, but you also may find it difficult to tell where the words start and stop in the first place. This is because most natural speech involves a largely *continuous* pattern of sound: What seem like pauses between individual words do not exist in the sounds entering our ears, but are rather constructed by our minds. This means that a first challenge for language learners is simply to find word boundaries in natural speech. Developmental researchers including Richard Aslin, Elissa Newport, and Jenny Saffran proposed that this task might be accomplished in part by the processing of statistical regularities in speech streams. Using a simple artificial language, they showed that eight-month-old infants were able to use statistical regularities over time to segment words out of continuous syllable streams.

Infants in such experiments might initially hear a continuous stream of syllables consisting of repeated sequences of four different triplet "words," each composed of three syllables (e.g., *bidaku, golabu, padoti, turopi*). These four triplets are repeated in a randomized order in a continuous stream, with no acoustic cues (such as longer pauses or prosody) to indicate where each triplet starts or stops (e.g. " . . . *bidakupadotigolabubidakuturopi* . . ."). Nevertheless, this stream contains robust statistical structure: The first syllable in a triplet perfectly predicts the second syllable, but the third syllable could be followed by several other syllables (corresponding to the first syllable of the following word). For example, if we denote the four triplets as *ABC, DEF, GHI,* and *JKL*—with each syllable denoted by a capital letter—then *A* will always be followed by *B,* but *C* will be followed by either *D, G,* or *J.* After only a few minutes of exposure to such a stream, a test phase then begins in which the infants are presented with three-syllable sequences in isolation—either triplets (e.g. *ABC*) or "nonword" foils composed from the same syllables, but in an order they have never heard before (e.g. *AEI*). Despite the fact that all of the individual syllables are equally familiar, infants can reliably discriminate the words and nonwords (as evaluated by a procedure that measures how long they attended to each type). Moreover, they can even discriminate triplets (e.g. *ABC*) from "partword" foils that they did hear, but less often (e.g. *BCG,* from when triplet *ABC* happened to be followed by *GHI*)—and later studies demonstrated a sensitivity to even more subtle statistical patterns. This is an example of SL because (1) the relevant regularities existed only in the distribution of syllables in time, because each individual syllable was heard equally often, and (2) the triplet "words" were obscured in an otherwise continuous stream of perceptual input.

This ability has led some researchers to speculate that SL may be a means by which children come to learn where the words are in continuous speech, but the precise relationship of SL to language acquisition remains uncertain. On one hand, this connection is supported by demonstrations that SL occurs more readily for linguistically relevant auditory information, and that it directly facilitates subsequent word learning. On the other hand, SL operates in subjects who have already mastered language (i.e., human adults), and even in nonlinguistic subjects (including nonhuman primates and even rodents)—and it also operates over many types of nonlinguistic input, such as musical tones. Moreover, some computational modeling studies suggest that SL is not sufficient to identify words in actual natural-language speech streams.

Visual Statistical Learning

Though SL had its origins in studies of language acquisition, it also operates pervasively in visual perception, where it has recently become a phenomenon of considerable interest. Visual statistical learning (VSL) can be measured in adult observers, for example, with continuous sequences of nonsense-shapes, when they appear one after another in an exact analogue to the auditory studies with infants—with each auditory syllable now replaced by a visual shape. After passively observing such a shape sequence (always with a constant between-shape delay) for several minutes, observers are able to reliably distinguish triplet "words" from three-shape foils when they are pitted directly against each other (now explicitly segmented) in a surprise "Which is more familiar?" test. Recent functional magnetic resonance imaging (fMRI) studies using similar designs have been able to explore the time course of VSL, and suggest that it is highly efficient—with the brain already discovering the presence of statistical structure after only the third or fourth repetition of each triplet.

VSL can also extract spatial regularities when multiple objects are viewed simultaneously. In some such studies, for example, observers see several shapes that are placed into a grid in reliable patterns (unbeknownst to them). For example, shape A might always appear immediately above and to the right of shape B. Because many such patterns are spatially interleaved in each grid, however, the pairs can only be spatially segmented on the basis of their statistically reliable spatial relations as observers see grid after grid. At test, observers are presented with the original spatial pairs pitted against foils consisting of two equally familiar shapes that had appeared together less frequently. Again, observers are able to judge the actual pairs as being more familiar.

Note that although SL may operate over spatial or temporal regularities, some of the mental representations that result from SL may be abstracted from such contexts. Indeed, VSL of spatial layouts can later be expressed in purely temporal contexts, and VSL of temporal sequences can later be expressed in purely spatial layouts.

What Type of Process Is Statistical Learning?

Since SL was first studied in these ways in the mid-1990s, a great deal has been learned about its underlying nature.

Automaticity and Implicitness

In adults, SL has most commonly been measured by explicitly separating sequences into their respective triplet "words" in a separate test phase, and then testing to see whether observers judge them to be more familiar than various types of foils, in forced choices. Such familiarity judgments could indicate conscious recognition of the statistical structure, but in fact such judgments are often reliable even when subjects were completing a separate "cover" task during the initial exposure, and even when they think they are merely guessing during the test phase. VSL has also been demonstrated with several implicit measures, however, including behavioral measures showing speeded response times for statistically predictable targets, and fMRI measures that have shown robust neural sensitivity to statistical structure even in subjects who show no SL by conventional familiarity tests. At the same time, additional studies have shown that SL in both the auditory and visual domains will only operate when the relevant stimuli are attended. If attention is instead directed away to other stimuli during the initial exposure, SL will not occur. Thus, it seems that SL is automatic in some senses, but not in others: It is gated by attention, but it nevertheless operates without intent or awareness.

The Input to Statistical Learning

One of the most critical steps in understanding any perceptual process is to determine the types of input over which that process operates. SL operates in multiple sensory modalities (including touch, in addition to vision and audition), and can accommodate many kinds of individual stimuli (such as musical tones in audition and dynamic events in vision). Research on VSL of temporal sequences also indicates that learning can operate at almost every level of the "hierarchy" of visual processing—from low-level visual features (such as shape), to discrete objects containing multiple features (such as color and shape), to high-level categories of visual scenes (such as "kitchens" and "forests").

Moreover, SL may help to determine what counts as an "object" in the first place. Statistical associations over time, for example, seem to help determine that a visual feature, such as color, is an intrinsic part of some objects (e.g., bananas) but not others (e.g., t-shirts). As a result, SL for colored-shape sequences can be expressed later for monochromatic shapes if the color-shape associations were loose during the initial exposure, but not if color and shape were reliably paired. For example, if shapes are assigned to unique fixed colors, then learning of colored-shape triplets can be expressed when the shapes are presented with their original colors at test, but not when the same shapes are presented in black during test (a context that is sufficient to express learning when the shapes have colors that randomly vary during familiarization). Similarly, studies of spatial VSL have demonstrated that learning does not operate efficiently over spatial patterns that are embedded within larger reliable patterns. For example, when spatial triplets are embedded in static grids, observers learn the triplets, but not their component pairs. And in audition, learning of *nonadjacent* statistical dependencies in temporal sequences is usually difficult, but it becomes easier when the interrupting stimuli are either highly variable or perfectly constant (as opposed to when they are only moderately variable). In each case, statistical variability seems to determine the "chunks" over which learning operates in the first place.

Controlling Statistical Learning

Recent research has also begun to determine the conditions under which SL will cease to operate. In most temporal SL studies, for example, the regularities are present in the sequences as soon as they begin, and they persist until the initial exposure phase ends. Learning is influenced in interesting ways, however, when the regularities appear only at the beginning or ending portions of the sequences—with the rest of the sequences filled with completely random orders of syllables or shapes. (Because of the implicit nature of SL, subjects in such experiments are not overtly aware of such transitions.) In this situation, learning will still occur when the structure precedes the structure-less "noise": the regularities that are initially discovered may remain robustly encoded, even when they later "fall apart." However, when the noise precedes the structure, learning is either weak or nonexistent: Perceptual processes may effectively learn that there is "nothing to learn" based on the initial noise and will stop trying (even though subjects still attend to the sequence).

What Is Statistical Learning Good For?

Statistical learning appears to a ubiquitous process in perception, but what is it good for? Recent research suggests at least three answers. First, by segmenting continuous streams of input, SL may effectively form the "chunks" that later processes depend on (e.g., the "words" in syntactic processing or the "objects" stored in visual working memory). This is supported by demonstrations that SL of linguistic stimuli can directly facilitate later word learning or relative-clause comprehension. Second, SL may speed later perceptual processing, making it more efficient—as is apparent in implicit response-time measures of VSL. Third, SL may be adaptive in that it yields a type of "future-oriented" processing, helping us to predict what we may be about to experience, so that we can adaptively tune our current behavior. Recent fMRI studies of VSL, for example, suggest that it yields a type of "implicit anticipation" of upcoming stimuli. Such implicit anticipation may be responsible for the facilitation previously described and may help the visual system cope with ambiguous or degraded stimuli. In sum, statistical learning is clearly a product of perception, but it may also in turn facilitate perception.

Brian J. Scholl and Nicholas B. Turk-Browne

See also Perceptual Learning; Unconscious Processes; Visual Scene Statistics

Further Readings

Fiser, J. Z., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science, 12,* 499–504.

Fiser, J. Z., & Aslin, R. N. (2002). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 28,* 458–467.

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences, 10,* 233–238.

Saffran, J. R. (2003). Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science, 12,* 110–114.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274,* 1926–1928.

Turk-Browne, N. B., Jungé, J. A., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology: General, 134,* 552–564.

Turk-Browne, N. B., Scholl, B. J., Chun, M. M., & Johnson, M. K. (in press). Neural evidence of statistical learning: Efficient detection of visual regularities without awareness. *Journal of Cognitive Neuroscience.*

Yang, C. D. (2004). Universal grammar, statistics, or both? *Trends in Cognitive Sciences, 8,* 451–456.

SURFACE AND MATERIAL PROPERTIES PERCEPTION

Objects in the environment can be described at many scales. For example, an orange has a shape