PART TWO

SENSORY PROCESSES
CHAPTER 4

Foundations of Visual Perception

MICHAEL KUBOVY, WILLIAM EPSTEIN, AND SERGEI GEPShteIN

This chapter contains three tutorial overviews of theoretical and methodological ideas that are important to students of visual perception. From the vast scope of the material we could have covered, we have chosen a small set of topics that form the foundations of vision research. To help fill the inevitable gaps, we have provided pointers to the literature, giving preference to works written at a level accessible to a beginning graduate student.

First, we provide a sketch of the theoretical foundations of our field. We lay out four major research programs (in the past they might have been called “schools”) and then discuss how they address eight foundational questions that promise to occupy our discipline for many years to come.

Second, we discuss psychophysics, which offers indispensable tools for the researcher. Here we lead the reader from the idea of threshold to the tools of signal detection theory. To illustrate our presentation of methodology we have not focused on the classics that appear in much of the secondary literature. Rather, we have chosen recent research that showcases the current practice in the field and the applicability of these methods to a wide range of problems.

The contemporary view of perception maintains that perceptual theory requires an understanding of our environment as well as the perceiver. That is why in the third section we ask what the regularities of the environment are, how may they be discovered, and to what extent perceivers use them. Here too we use recent research to exemplify this approach.

Reviews of the research on higher visual processes are available in this volume in the chapters by Palmer and by Proffitt and Caudek.

THEORIES AND FOUNDATIONAL QUESTIONS

Four Theories 87
Eight Foundational Questions 89

PSYCHOPHYSICAL METHODS 96
Threshold Theories 97
Signal Detection Theory 103
The ROC Curve; Estimating d’ 105
Energy Thresholds and Observer Thresholds 105
Some Methods for Threshold Determination 105

THE “STRUCTURE” OF THE VISUAL ENVIRONMENT AND PERCEPTION 110
Regularities of the Environment 110
Redundancy and Covariation 111
Co-occurrence Statistics of Natural Contours 112

REFERENCES 116

The writing of this chapter was supported by NEI grant R01 EY 12926-06.

87
nature and effects of these cognitive operations may be profitably studied in any setting that activates them. It is neither necessary nor desirable to reinstate the typical conditions of ordinary seeing.


**Gestalt Theory**

Gestalt theory proposes that the process of perception is an executive-free expression of the global properties of the brain. The organization and orderliness of the perceptual world is an emergent property of the brain as a dynamical system. Gestalt theory intends to distance itself from any position that posits an executive (a homuncular agency) that oversees the work of the perceptual system. The Gestalt theory thus recognizes regulation but will not allow a regulator. A dynamical system which instantiates a massively parallel self-organizing process satisfies is regulated but does not have a regulator. As such, the perceptual world is different from the sum of its parts and cannot be understood by an analytic investigative strategy that adopts a purely local focus. To understand perception we need to discover the principles that govern global perception. The most familiar application of this notion involves the Gestalt principles of grouping that govern perceived form (see chapter by Palmer in this volume).

Gestalt theory emerged in the early decades of the century in the writings of Wertheimer (1912), Köhler (1929, 1940), and Koffka (1935). Although Gestalt theory fell from favor after that period, its influence on modern thought is considerable. Moreover, although ardent advocacy of the original Gestalt theory may have come to an end with the death of Köhler in 1967, a new appreciation for and extension of Gestalt theory or metatheory (Epstein, 1988) has developed among contemporary students (e.g., Kubovy & Gepshtein, in press).

**Ecological Realism**

The ecological approach has also been called the theory of direct perception: The process of perception is nothing more than the unmediated detection of information. According to this approach, if we describe the environment and stimulation at the appropriate level, we will find that stimulation is unambiguous. In other words, stimulation carries all the information needed for perception. The appropriate level of description can be discovered by understanding the successful behavior of the whole organism in its ecological niche.

This approach appeared in embryonic form in 1950 in Gibson’s Perception of the Visual World and in mature form in Gibson’s last book (1979), in which he explicitly denied the fundamental premises of his rivals. Despite this, a significant segment of the contemporary scientific community is sympathetic to his views (Bruce, Green, & Georgeson, 1996; Nakayama, 1994).

**Computational Constructivism**

According to computational constructivism, the perceptual process consists of a fixed sequence of separable processing stages. The initial stage operates on the retinal image to generate a symbolic recoding of the image. Subsequent stages transform the earlier outputs so that when the full sequence has been executed the result is an environment-centered description. Computational constructivism bears a family resemblance to cognitive constructivism. Nevertheless, the computationalist is distinguished in at least three respects: (a) The canonical computationalist approach resists notions of cognitive operations in modeling perception, preferring to emphasize the contributions of biologically grounded mechanisms; (b) the computationalist approach involves stored knowledge only in the last stage of processing; (c) the computationalist aspires to a degree of explicitness in modeling the operations at each stage sufficient to support computer implementation.

Computational constructivism is the most recent entry into the field. The modern origins of computational constructivism are to be found in the efforts of computer scientists to implement machine vision (see Barrow & Tenenbaum, 1986). The first mature theoretical exercise in computational constructivism appeared in 1982 in Marr’s Vision.

The preceding may create the impression that the vision community can be neatly segregated into four camps. In fact, many students of perception would resist such compartmentalization, holding a pragmatic or eclectic stance. In the view of the eclectic theorists, the visual system exploits a variety
of processes to fulfill its functions. Ramachandran (1990a, 1990b) gives the most explicit expression of this standpoint in his utilitarian theory.

**Eight Foundational Questions**

The commonalities and differences among the four theories under consideration are shaped by their approaches, implicit or explicit, toward a number of basic questions.

**What Is Vision For?**

What is the visual system for? The answer to the question can shape both the goals of experimentation and the procedures of investigation. For most of the twentieth century one answer has been paramount: The function of the visual system is to generate or compute representations or descriptions of the world. Of course, a representation is not to be considered a picture in the mind. Nevertheless, representations serve a useful function by mirroring, even if symbolically, the organization and content of the world to be perceived.

Acceptance of the preeminence of the representational function is apparent in the Gestalt insistence that the first step in the scientific analysis of visual perception is application of the phenomenological method (Kubovy, 1999). This same endorsement is not as wholehearted in cognitive constructivist approaches (Kubovy & Gepshtein, in press). Nevertheless, a review of two of the major documents of cognitive constructivism, Rock’s (1983) *The Logic of Perception* and the edited collection *Indirect Perception* (Rock, 1997), shows that in every one of the dozens of investigations reported, the dependent variables were direct or indirect measures of perceptual experience. Marr (1982) was also explicit in aligning himself with the representational view. For Marr, the function of vision is “discovering from images what is present in the world.” The task for the vision scientist is to discover the algorithms that are deployed by the visual system to take the raw input of sensory stimulation to the ultimate object-centered representation of the world. Given this conception of a disembodied visual system and the task for the visual system, the ideal preparation for the investigation of vision is the artificial (nonbiological) vision system realized by the computer.

The ecological realists do not join the broad consensus concerning the representational function of the visual system. For Gibson, the primary function of the visual system is to detect the information in optical structures that specifies the actions afforded by the environment (e.g., that a surface affords support, that an object affords grasping). The function of the visual system is to perceive possible action, that is, actions that may be successfully executed in particular environmental circumstances.

The representationalists also recognize that perception is frequently in the service of action. Nonetheless, the difference between the representationalists and the ecological realists is significant. For the representationalists the primary function of the visual system is description of the world. The products of the visual system may then be transmitted to the action system. The perceptual system and the action system are separate. Gibson, by contrast, dilutes the distinction between the perceptual system and the action system. The shaping of action does not await perception; action possibilities are perceived directly.

We might expect that following on the ecological realist redefinition of the function of the visual system there would be a redirection of experimental focus to emphasize action and action measures. However, a redirection along these lines is not obvious in the ecological realist literature. Although there are several notable examples of focus on action in the studies of affordances (e.g., Warren, 1984; Warren & Whang, 1987), overall, in practice it is reformulation of input that has distinguished the ecological approach. The tasks set for the subjects and the dependent measures in ecologically motivated studies are usually in the tradition established by the representationalists.

The last two decades of the twentieth century have witnessed a third answer to the question of function. According to this new view, which owes much to the work of Milner and Goodale (1995), the visual system is composed of two major subsystems supported by different biological structures and serving different functions. The proposal that there is a functional distinction between the two major projections from primary visual cortex is found in earlier writing by Schneider (1969) and Ungerleider and Mishkin (1982). These writers proposed that there were two visual systems: the “what” system designed to process information for object identification and the “where” system specialized for processing information for spatial location. The newer proposal differs from the older ones in two respects: (a) The functions attributed to the subsystems are to support object identification (the what function) and action (the how function), and (b) these functions are implemented not by processing different inputs but by processing the same input differently in accordance with the function of the system. As Milner and Goodale (1995, p. 24) noted, “we propose that the anatomical distinction between the ventral and dorsal streams corresponds to the distinction . . . between perceptual representation and visuomotor control. . . . The reason there are two cortical pathways
is that each must transform incoming visual information for different purposes.” The principal support for this two-vision hypothesis has been provided by findings of double dissociations between action and perception—that is, between assessments of effective action and measures of perceptual experience—in brain-damaged and intact individuals. These findings (summarized by Milner & Goodale, 1995, and by Goodale & Humphrey, 1998) imply that it will be profitable to adopt dual parallel investigative approaches to the study of vision, one deploying action-based measures, the other more traditional “perceptual” measures.

Goodale and Humphrey (1998) and Norman (in press) proposed that the two-vision model provides a framework for reconciling the ecological and computational approaches: “Marrian or ‘reconstructive’ approaches and Gibsonian or ‘purposive animate-behaviorist’ approaches need not be seen as mutually exclusive, but rather as complementary in their emphasis on different aspects of visual function” (Goodale & Humphrey, 1998, p. 181). We suspect that neither Gibson nor Marr would have endorsed this proposal. (Chapters by Heuer and by Proffitt and Caudek in this volume also discuss the distinction between the perceptual system and the action system.)

Percepts and Neurons

Perceptual processes are realized by a biological vision system that evolved under circumstances that have favored organisms (or genetic structures) that sustain contact with the environment. No one doubts that a description and understanding of the hardware of the visual system will eventually be part of an account of perception. Nevertheless, there are important differences among theories in their uses of neurophysiology.

One of the tenets of first-generation information-processing theory (e.g., Johnson-Laird, 1988; Neisser, 1967) is that the mind is an operating system that runs on the brain and that the proper business of the psychology of cognition and perception is study of the program and not the computer—the algorithm and not the hardware. Furthermore, inasmuch as an algorithm can be implemented by diverse computational architectures, there is no reason to look to hardware for constraints on algorithms. Another way of expressing this position is that the aim of information-processing theory, as a theory of perception, is to identify functional algorithms above the level of neurophysiology.

The cognitive constructivist shares many of the basic assumptions of standard information-processing theory and has adopted the independence stance toward physiology. Of course, perceptual processes are implemented by biological hardware. Nevertheless, perceptual theorizing is not closely constrained by the facts or assumptions of sensory physiology. The use of physiology is notably sparse in the principal documents of cognitive constructivism (e.g., Rock, 1983, 1997). Helmholtz may seem to be an important exception to this characterization; but, in fact, he was careful to keep his physiology and psychology separate (e.g., Helmholtz, 1866/2000, Vol. 3).

Physiological talk is also absent in the canonical works of the ecological theorists, but for different reasons. The ecologists contend that the questions that have been addressed by sensory physiologists have been motivated by tacit acceptance of a metatheory of perception that is seriously flawed: the metatheory of the cognitive constructivist. As a consequence, whereas the answers discovered by investigations of sensory physiologists may be correct, they are not very useful. For example, the many efforts to identify the properties of the neuronal structures underlying perception by recording the responses of single cells to single points of light seem to reflect the tacit belief that the perceptual system is designed to detect single points. If the specialization of the visual system is different, such as detecting spatiotemporal optical structures, the results of such studies are not likely to contribute significantly to a theory of perception. In the ecological view what is needed is a new sensory physiology informed by an ecological stance toward stimulation and the tasks of perception.

The chief integrative statement of the computational approach, Marr’s (1982) *Vision*, is laced with sensory physiology. This is particularly true for the exposition of the computations of early vision. Nevertheless, in the spirit of functionalism Marr insists that the chief constraints are lodged in an analysis of the goals of perceptual computation. In theorizing about perceptual process (i.e., the study of algorithms) we should be guided by its computational goal, not by the computational capabilities of the hardware. When an algorithm can satisfy the requirements of the task, we may look for biological mechanisms that might implement it.

The Gestalt theorists (e.g., Köhler, 1929, 1940) were forthright in their embrace of physiology. For them, a plausible theory must postulate processes that are characteristic of the physical substrate, that is, the brain. Although it is in principle possible to implement algorithms in diverse ways, it is perverse to ignore the fit between the properties of the computer and the properties of the program. This view is in sharp contrast to the hardware-neutral view of the cognitive constructivist: For the Gestalt theorist, the program must be reconciled with the nature of the machine (Epstein, 1988; Epstein & Hatfield, 1994). In this respect, Gestalt theory anticipated current trends in cognitive neuroscience, such as the connectionist approaches (Epstein, 1988).
The consensus among contemporary investigators of perception favors a bimodal approach that makes a place for both the neurophysiological and the algorithmic approaches. The consensus is that the coevolution of a neurophysiology that keeps in mind the computational problems of vision and of a computational theory that keeps in mind the competencies of the biological vision system is most likely to promote good theory.

Although this bimodal approach might seem to be unexceptionable, important theoretical disagreements persist concerning its implementation. Consider, as an example, Barlow’s (1972, 1995) bold proposal called the single-neuron doctrine: “Active high level neurons directly and simply cause the elements of our perception” (Barlow, 1972, §6.4, Fourth Dogma). In a later formulation, “Whenever two stimuli can be distinguished reliably, then some analysis of the neurological messages they cause in some single neuron would enable them to be distinguished with equal or greater reliability” (Barlow, 1995, p. 428). The status of the single-neuron doctrine has been reviewed by Lee (1999) and by Parker and Newsome (1998). The general experimental paradigm assesses covariation between neural activity in single cortical neurons and detection or discrimination at threshold. The single-neuron doctrine proposes that psychophysical functions should be comparable to functions describing neural activity and that decisions made near threshold should be correlated with trial-to-trial fluctuations of single cortical neurons (e.g., Britten, Shadlen, Newsome, & Movshon, 1992).

The available data do not allow a clear-cut decision concerning this fundamental prediction. However, whatever the final outcome may be, disagreements about the significance of the findings will arise from differences concerning the appropriate unit of analysis. Consider first the perceptual side that was elected for analysis. From the standpoint of the ecological realist (e.g., Gibson, 1979), the election of simple detection and discrimination at threshold is misguided. The ecological realist holds that the basic function of the visual system is to detect information in spatiotemporal optical structure that is specific to the affordances of the environment. Examining relations between neuronal activity and psychophysical functions at threshold is at the wrong level of behavior. As noted before, it is for this reason that the canonical documents of the ecological approach (Gibson, 1950, 1966, 1979) made no use of psychophysiology.

Similar reservations arise in the Gestalt approach. Since its inception, Gestalt theory (Hatfield & Epstein, 1985; Köhler, 1940) has held that only a model of underlying brain processes can stand as an explanation. In searching for the brain model, Gestalt theorists were guided by a heuristic: The brain processes and the perceptual experiences that they support have common characteristics. Consequently, a careful and epistemically honest exploration of perceptual experience should yield important clues to the correct model of the brain. According to Gestalt theory, phenomenological exploration reveals that global organization is the most salient property of the perceptual world, and it is a search for the psychophysiological correlates of global experience that will bring understanding of perception.

There are analogous differences concerning the choice of stimulation. If there is to be an examination of the neurophysiological correlates of the apprehension of affordances and global experience, then the stimulus displays must support such perceptions. Proponents of this prescription suspect that the promise of the pioneering work of Hubel and Wiesel (1962) has not been realized because investigators have opted for the wrong level of stimulation.

Concerning Information

The term information has many uses within psychology (Dretske, 1986). Here the term refers to putative properties of optical stimulation that could specify the environmental state of affairs (i.e., environmental properties), structures, or events that are the distal source of the optical input. To specify an environmental state is to pick out the actual state of affairs from the family of candidate states that are compatible with the given optical stimulation.

Cognitive constructivists have asserted that no properties of optical stimulation can be found to satisfy the requirements of information in this sense because optical stimulation is intractably equivocal. At best optical stimulation may provide clues—but never unequivocal information—concerning the state of the world. This assessment was already entrenched when Berkeley wrote his influential Essay Towards a New Theory of Vision (Berkeley, 1709/2000), and the assessment has been preserved over the ensuing three centuries. The assumption of intractable equivocality is one of the foundational premises of constructivism; it serves as a basic motivation of the enterprise. For example, the transactionalists (Itelson, 1960; Kilpatrick, 1950, chap. 2) lay the foundation for their neo-Helmholtzian approach by showing that for any proximal retinal state there is an infinite class of distal “equivalent configurations” that are compatible with a given state of the retina. In the same vein, computational research routinely opens with a mention of the “inverse projection problem.” If optical stimulation does not carry information that can specify the environment, we must look elsewhere for an account of perception.
The view of the theory of direct perception concerning information is radically different. Proponents of this theory (Rogers, 2000) vigorously reject the assumption of intractable equivocality. Following Gibson, they contend that the tenet of equivocality is false, that it is mistakenly derived from premises about the nature of the stimulation that enters into the perceptual process. The cognitive constructivist who mistakenly uses static displays of points or objects isolated from their optical context (e.g., a point of light or an illuminated object in the dark or a display presented briefly) mistakenly concludes that stimulation is informationally impoverished. But direct perception argues that this paradigm does not represent the optical environment that has shaped the visual system. Even worse, the paradigm serves to create experiments with informationally impoverished displays. Thus equivocality is only an artifact of the constructivist’s favored paradigm and not a characteristic of all optical stimulation. The stimulation that the perceptual system typically encounters and to which it has been attuned by evolution is spatially and temporally distributed. These spatiotemporal optical structures, which are configurations of optical motion, can specify the environment. There is sufficient information in stimulation to support adaptive perception. And when pickup of information suffices to explain perception, cognitive operations that construct the perceptual world are superfluous.

The stance of the computational constructivist regarding the question of information cannot be characterized easily. If by information is meant a unique relationship between optical input and a distal state that is unconditional and not contingent on circumstances, then the computational constructivist must be counted among the skeptics. Optical structures cannot specify distal states noncontingently. Other conditions must be satisfied. The other conditions, which may be called constraints, are the regularities, covariances, and uniformities of the environment. Accordingly, assertions about the informational status of optical stimulation must include two conjoint claims: One is about properties of optical stimulation, and the other is about properties of the environment.

Moreover, from a computational constructivist stance, still more is needed to make information-talk coherent. Consideration must be given to the processes and algorithms that make explicit the relationships that are latent in the raw optical input. Whereas the advocates of the theory of direct perception talk of spatiotemporal optical structures, the computationalist sees the structure as the product of processes that operate on unstructured optical input. It is only in the tripartite context of optical input, constraints, and processing algorithms that the computationalist talks about information for perception.

The Gestalt psychologists, writing well before the foregoing theorists, also subscribed to the view that optical stimulation does not carry information. Two considerations led them to this conclusion. First, like the later computationalists, they were convinced that it was a serious error to attribute organization or structure to raw optical input. The perceptual world displays organization, and by Gestalt hypothesis the brain processes underlying perception are organized; but retinal stimulation is not organized. Second, even were it permissible to treat optical input as organized, little would be gained because optical input underdetermines the distal state of affairs. For example, even granting the status of an optical motion configuration to an aggregate of points that displace across the retina by different directions, amplitudes, and velocities (i.e., granting organization to stimulation), there are infinitely many three-dimensional structures consistent with a given configuration of optical motion. For Gestalt theory, structure and organization are the product of spontaneous dynamic interactions in the brain. Optical input is a source of constraints in determining the solution into which the brain process settles.

Concerning Representation

A representation is something that stands for something else. To stand for a represented domain the representation does not have to be a re-presentation. The representations that are active in theoretical formulations of the perceptual process are not iconic images of the represented domain. Rather, a representation is taken to be a symbolic recoding that preserves the information about objects and relations in the represented domain (Palmer, 1976).

Representations play a prominent role in cognitive and computational constructivism. Positing representations is a way of reconciling a sharp disparity between the phenomenology of everyday seeing and the scientific analysis of the possibilities of seeing. The experience of ordinary seeing is one of direct contact with the world. But as the argument goes, even cursory analysis shows that all that is directly available to the perciipient is the light reflected from surfaces in the world onto receptive surfaces of the eye. How can this fundamental fact be reconciled with the nature of the experience of seeing? Moreover, how can the fact that only light gets in be reconciled with the fact that it is the world that we see, not light? (Indeed, what could it mean to say that we see light?) Both questions are resolved by the introduction of representations. It is representations that are experienced directly, and because the representations preserve the features, relationships, and events in the represented world,
the experience of perception is one of direct contact with the world. In this way, representations get the outside inside (Epstein, 1993).

According to constructivist theory, the perceptual world is constructed or assembled from the raw material of sensory input and stored knowledge. The process of construction has been likened to inference or problem solving, and more recently the process has been characterized as computational. The representational framework serves as a superstructure for support of this conception of the perceptual process. Proponents of the computational/representational approach (e.g., Fodor, 1983; Fodor & Pylyshyn, 1981) argue that the only plausible story of perception is computational and that the only plausible computational story must assume a representational system in which the computations are executed.

It seems undeniable that if a variant of the standard constructivist/computational approach is adopted, the representational framework is needed to allow the approach to proceed smoothly. Any theory that postulates a process resembling nondemonstrative inference (Gregory, 1970; Rock, 1983, 1997; or the Bayesian approaches, e.g., Hoffman, 1998; Knill & Richards, 1996) or a process of representational transformation (e.g., Marr, 1982) must postulate a representational medium for the display of “premises” or the display of representations, that is, the output of processes (algorithms) that operate over mappings. No one has been more straightforward and exacting in promoting this approach than Marr in his Vision.

In contrast, the theory of direct perception makes no use of representations. Advocates of direct theory argue that the flaws of representationalism are insurmountable. Some of these flaws are logical, such as the familiar troubles with the representational theory of mind, the philosophical progenitor of the contemporary representational framework. As one example, if direct perception were only of representations, how do we come to know what external objects are like, or which representations they resemble? By hypothesis, we can only perceive representations, so that whenever we may think that we are observing external objects to compare them with representations or to discover their intrinsic nature, we are only observing other representations. In general, it is difficult to escape from the world of representations.

In addition to pointing to logical difficulties, proponents of the theory of direct perception see no need to invoke representations in the first place. According to the ecological realists, representationalism is parasitic on constructivism. If constructivism is accepted, then representationalism is compelling; but if it is rejected, then representationalism is unmotivated.

Gestalt theory developed before the age of self-conscious representationalism. There is no explicit treatment of representations in the writings of the Gestalt theorists. Nevertheless, we can infer that the Gestalt theorists would have sided with the advocates of direct perception in this matter. Considerations that support this inference emerge in the next two sections.

**Representational Transformation**

As a general rule, perceiving is automatic and seamless. Compare, for instance, the effortlessness of seeing with the trouble and toil of learning and reasoning. Although the characterization is unlikely to be questioned as a description of the experience of ordinary seeing, when we consider the process that underlies perceiving, important differences among theories emerge with respect to decomposability. Ignoring theoretical nuances for the present, we find that constructivist theories, both cognitive and computational, hold a common view, whereas Gestalt theory and the theory of direct perception adopt a contrasting position.

The constructivist view is that the process of perception may be decomposed into a series of operations whose function is to take the raw input to the sensory surface and by a series of transformations generate a distally correlated representation of the environment. The process of perception is a process of representational transformation. The constructivists are drawn to this position by an a priori belief that only a model of representational transformation will be sufficient as a description of the perceptual process. One form of empirical support for this belief is found in the requirements of successful algorithms for the attainment of the objectives of perception, such as generating three-dimensional structure from stereopsis. Evidence of the psychological reality of the putative intermediate representations is provided by experimental procedures that ostensibly segregate the component representations.

Neither Gestalt theory nor the theory of direct perception makes use of the model of representational transformation. They do not agree that postulation of a sequential multistage process is necessary, and they question the interpretation of the experimental data. For Gestalt theory, the perceptual process is a noncognitive, highly interactive process that automatically settles into the best fitting state (Epstein & Hatfield, 1994; Hatfield & Epstein, 1985). Any effort to parse the process into intermediate states is arbitrary. On no account should such contrived states be assigned a role in the causal story of perception. Proponents (e.g., Gibson, 1966, 1979; Turvey, Shaw, Reid, & Mace, 1981) of the theory of
direct perception have been equally adamant in rejecting the model of representational transformation. They maintain that the model results from questionable premises. Once these are abandoned, the apparent need for positing intervening representational states vanishes.

**Perception and Cognition**

What is the relationship between perceptual processes and cognitive processes? The answers to this question have ranged widely over the theoretical landscape. The cognitive constructivists consider perception to be perfused by cognition. In the view of the cognitive constructivist, the perceptual process is a cognitive process. The principal distinction between perceptual processes and cognitive processes is that in the former case mental operations are applied to the transformation of representations originating in occurrent optical input, whereas in the latter case mental operations are applied to the transformation or representations drawn from the pre-existing knowledge base. This attribution is clear-cut for contemporary constructivists, such as Rock (1983, 1997), who characterize perception as a process of intelligent problem solving, as it was in the classical description (Helmholtz, 1866/2000) of perception as a process of unconscious inference and in the New Look movement in North American psychology (Bruner, 1957). The assumption that perception and cognition are continuous is also commonly found in applying standard information theory to problems of perception (e.g., Lindsay & Norman, 1977; Rumelhart, 1977).

The continuity claim is central to the cognitive constructivist position. The claim rests on a diverse set of experimental observations that are said to imply the interpenetration of perception and cognition. Many of the parade cases emerged from the laboratory of Rock (1983, 1997). Despite the compelling character of some of these cases, they have not been decisive. Pylyshyn (1999) has presented a thorough airing of the controversy. In his assessment the cases featured by the cognitive constructivists do not support the claim of cognitive penetrability of perception; “rather, they show that certain natural constraints on interpretation, concerned primarily with optical and geometrical properties of the world, have been compiled into the visual system” (p. 341).

The computational constructivist takes a more restrained position. The aim of the computational approach is to advance the explanation of perception without invoking cognitive factors. Nevertheless, the full explanation of perception requires cognitive operations. In the model of representational transformation adopted by the computational approach, the sequence of operations is divided into early and late vision. The former is supposed to be free of cognitive influence. The operations are executed by modular components of the visual system that are cognitively impenetrable; that is, the modules are encapsulated, sealed off from the store of general knowledge. These operations of early vision perform vital work but do not deliver a representation sufficient to sustain adaptive behavior. A full-bodied, environment-centered representation requires activation of stored mental models and interpretation of the representations of early vision in this context. An exemplar of this stance toward cognition and perception is Marr’s (1982) computational theory.

The attitudes of Gestalt theory and the theory of direct perception are opposed to the constructivist stance. Indeed, in the case of Gestalt theory the difference is particularly striking. Whereas the constructivist proposes that perception has significant cognitive components, the larger program of Gestalt theory proposes that much of cognition, such as thinking and problem solving, is best understood as an expression of fundamental principles of perception. The theory of direct perception considers the entire perceptual system to be encapsulated, and therefore uninvolved, in interaction with other information-processing operations. This position does not carry with it a rejection of influences of past experience or learning in perception, but it does require a different construal of the mechanism that supports these influences.

**Modularity**

Is the visual system a general-purpose processor serving all of the diverse perceptual needs of the organism, or is it a collection of independent perceptual modules that have evolved to compute domain-specific solutions, such as depth from shading, shape, or motion? The answer to this question depends on how modularity is construed. Consider three construals that vary the conditions they impose on the postulation of modularity (the terms weak, moderate, and strong modularity are ours).

**Weak Modularity.** Weak modularity stipulates only two conditions: (a) that a segregated bit of the biological hardware be shown to be exclusively dedicated to representation of a specific environmental feature, such as solidity; and (b) that the designated hardware be specialized for the processing of a particular form of stimulation, such as retinal disparity. Under this construal, when these two conditions are satisfied, postulation of a stereoscopic depth module is warranted. If this minimal set of features for modularity is adopted, there probably will be little disagreement that the visual system is modular.
Moderate Modularity. Moderate modularity is defined by a list of features made explicit by Fodor (1983) in his *Modularity of Mind*. To the two criteria given above, Fodor adds several others: that modules are informationally encapsulated, that modular processes are unconscious, that modular processing is very fast and obligatory, that modules have shallow outputs, that modules emerge in a characteristic ontogenetic sequence, and that following insult modules exhibit characteristic disruption.

Among cognitive psychologists, claims for modularity tend to be measured against Fodor’s expanded list. Unsurprisingly, when the expanded list of criterial features is adopted, agreement on modularity is harder to reach. Much of the controversy involves encapsulation. By this test, a dedicated biological device that is uniquely sensitive to an eccentric form of stimulation will be considered to be a modular component only if under normal conditions of its operation its processes run their course uninfluenced by factors that are extraneous to the module. Neither concurrent activity in other modules nor reference to stored knowledge of past events or anticipations of future events affects the module. The module is an impenetrable encapsulated system (Fodor, 1983; Pylyshyn, 1984).

Two kinds of problems recur in assessments of encapsulation. First, it is universally accepted that performance of almost any task may be affected by a host of cognitive factors. Accordingly, the claim for encapsulation says that however these cognitive factors influence performance, they do not do so by influencing the computations of the module. Consequently, an experimental demonstration that performance is affected by cognitive factors or by the output of parallel computations does not necessarily negate modularity unless it can be shown that the effects are located in the computations that are endogenous to the putative module. This latter assertion is hard to establish (e.g., Pylyshyn, 1999).

Second, there is the problem of the practice effect. That is, performance of a task that seems unlikely to be supported by a dedicated biological device or to be dependent on access to special stimulation will exhibit many of the features of modularity when the task is highly practiced. For example, performance may become very fast, mandatory, and inaccessible to conscious monitoring. Consequently, evidence that the process underlying performance exhibits these features does not necessarily implicate modularity.

Strong Modularity. Strong modularity adds to the composite list just given the added requirement that the candidate module exhibit a distinctive style of processing. Although no one has advanced this claim explicitly, it is implicit in the writings of modularists that modules work by implementing the same process. As two examples, in Marr’s (1982) approach all the modules are noncognitive computational devices, and in Fodor’s (1983) canonical analysis of modularity all of the modules are inferential engines. Because the modularity stance does not seem tied to views of process, a stance on modularity does not exert strong constraints on the characterization of the perceptual process. Thus an ecological realist might also adopt modularity, holding that the modules are independent devices for detection (pickup) of information in spatiotemporal optical structure.

Although the postulation of modularity is compatible with a variety of positions regarding perceptual process, an exception to the rule must be made for cognitive constructivism. On the face of it, modularity and cognitive constructivism cannot be linked except in the weak sense of modularity (the first construal). The cognitive constructivist takes the perceptual process to be a cognitive process that ranges freely over the knowledge domain. The principal arguments for the claim that the perceptual process is a form of “hypotheses testing” or “intelligent problem solving” very often take the form of demonstrations that perception is cognitively penetrable (e.g., Rock, 1997). Certainly this is the way that the cognitive constructivist wishes to be understood.

On Illusion and Veridicality

Generally, perception is a reliable guide to action. Occasionally, however, perception misrepresents and action predicated on the implications of perception fails. Perceptual misrepresentations arise under a variety of conditions: (a) The normal link between the environmental state of affairs and optical input is severed: For example, the spatial arrangement of points on the retina and the spatial arrangement of points comprising an environmental object are normally in alignment. A straight stick will have a correspondingly straight retinal contour associated with it. However, if the stick is half immersed in water, the different refractive indices of water and air will result in a “bent” retinal contour. Under these circumstances the straight stick will look bent. (b) The normal pattern of neuronal activation engendered by exposure to a distal arrangement is modified: For example, continuous visual inspection of a line tilted in the frontal plane will modify the pattern of activity of neuronal orientation detectors. The resultant perceptual effect is an alteration of perceived orientation; a test line tilted in the same direction as the inspection line will look upright, and an upright line will look tilted in the direction opposite to the tilt of the inspection line. (c) Rules of perceptual inference are overgeneralized; that is, the rules are applied under conditions for which they are inappropriate. A widely held view (e.g., Gregory, 1970, 1997)
attributes many geometric illusions of size to this sort of misapplication of rules. For example, the illusions of size in perspective renderings of a scene, such as the Ponzo illusion, are attributed to the irrepressible but inappropriate application of the putative standard rule for computing size on the basis of visual angle and perceived distance.

All cases of misrepresentation do heavy work, but the contrasting theories depend on different forms of misrepresentation to promote their aims. Consider the theories of direct perception and cognitive constructivism. The ecological realist turns for support to examples of the first kind, which break the link between the environment and optical input. By decoupling the distal state from the optical state while preserving the spatiotemporal optical structure, as in the optical tunnel (Gibson, 1979, Figure 9.2, p. 154), the ecological realist means to demonstrate that information is a property of optical structure and that perception is the pickup of information in optical structure. The advocates of the theory of direct perception are of course deeply distrustful of misrepresentations of the third kind (rules are applied under conditions for which they are inappropriate). They contend that these cases are artifacts of special situations and cannot illuminate the workings of ordinary seeing. Occasionally, advocates of direct theory have suggested that a special theory, a theory of judgment and decision making under uncertainty, is needed for perceptual misrepresentations of the third kind.

The cognitive constructivist, on the other hand, relies heavily on misrepresentations of the third kind. Indeed, these instances of misrepresentations form the core of the empirical case for cognitive constructivism. It is supposed that these perceptual misrepresentations disclose the workings of the hidden processes that govern perception. According to the cognitive constructivist, the processes that underlie veridical and illusory perception are the same, and these processes are revealed by misrepresentations of the third kind. The cognitive constructivist doubts that the demonstrations of the first kind can, in fact, be interpreted in the manner urged by the advocate of direct theory.

Although it has been often suggested that investigations of misrepresentation can be decisive for theories of perception, only infrequently do analyses of misrepresentation test competing hypotheses originating in rival general theoretical orientations. The more common practice is to offer examples of misrepresentation as elements in a confirmation strategy (Nickerson, 1998). The misrepresentations of choice serve as existence proofs of some fundamental postulate of the theoretical approach. The confirmation strategy acts as a directive influence in selecting the class of misrepresentation for investigation and a disincentive that discourages consideration of contrasting accounts generated by rival theories.

PSYCHOPHYSICAL METHODS

The purpose of the remainder of this chapter is to introduce the reader to a selection of experimental techniques and tools for theory construction. Wherever possible we do this by referring new concepts to contemporary experiments and theories. In this way, the reader will understand the ideas in context. We recommend the following textbooks and chapters for the reader who wishes to pursue topics introduced in this section: Gescheider (1997), Hartmann (1998), Link (1992), Luce and Krumhansl (1988), Macmillan and Creelman (1991), and Swets (1996).

Psychophysical methods are indispensable to the advancement of perceptual research. Baird and Noma (1978, p. 1) put it well:

Psychophysics is commonly defined as the quantitative branch of the study of perception, examining the relations between observed stimuli and responses and the reasons for those relations. This is, however, a very narrow view of the influence it has had on much of psychology. Since its inception, psychophysics has been based on the assumption that the human perceptual system is a measuring instrument yielding results (experiences, judgments, responses) that may be systematically analyzed. Because of its long history (over 100 years [in 1978]), its experimental methods, data analyses, and models of underlying perceptual and cognitive processes have reached a high level of refinement. For this reason, many techniques originally developed in psychophysics have been used to unravel problems in learning, memory, attitude measurement, and social psychology. In addition, scaling and measurement theory have adapted these methods and models to analyze decision making in contexts entirely divorced from perception.

After Fechner (1860/1996) developed psychophysics, two kinds of questions were asked: (a) How sensitive are observers to intensities of stimulation? and (b) How intense do certain amounts of stimulation appear to observers? The first question is about thresholds, the second about scaling. Given the magnitude of these two fields of research and the number of research tools each has spawned, we have chosen to focus on the more fundamental problem of observer sensitivity.

The notion of threshold comes from Leibniz’s New Essays on Human Understanding (1765/1981):

I would prefer to distinguish between perception and being aware. For instance, a perception of light or colour of which we are aware is made up of many minute perceptions of which we are unaware; and a noise which we perceive but do not attend to is brought within reach of our awareness by a tiny increase or addition. If the previous noise had no effect on the soul, this minute addition would have none either, nor would the total. (book 2, chap. 9, p. 134; see also Leibniz, 1989, p. 295.)
Herbart (1824/1890) may have been the first to use the Latin term for threshold, *limen*, and its German equivalent, *schwelle*, to refer to the limit below which a given stimulus ceases to be perceptible. Although the idea of threshold appears straightforward, it turns out to be complex. We now explore the original idea of threshold and show its limitations. After that we present it in its current form: signal detection theory.

**Threshold Theories**

We begin with the simplest threshold theory, to which we will gradually add elements until it can be confronted with data. By that point we will have introduced two fundamental ideas: the receiver operating characteristic (ROC) and the possibility of disentangling the sensitivity of observers from their response bias.

**Fixed Energy Threshold—Naive Observer**

The simplest threshold theory is depicted in Figure 4.1. We look at the panels from left to right.

**Panel 1: Threshold Location.** This panel represents two fundamental ideas: (a) The observer can be in one of two
states $D$ or $\bar{D}$, and (b) on the scale of stimulus energy there is a fixed value below which observers can never detect a stimulus. Above this threshold, observers always detect it.

**Panel 2: Detection Probability.** This panel represents the same idea in more modern terms. This graph represents the probability of an observer being above the observer threshold—in state $D$—as a function of stimulus energy. As you would expect, this probability is 0 below the energy threshold and 1 above it. The multiple arrows represent stimuli; they are unfilled if they are at an energy that is below the energy threshold and filled if they are above it.

**Panel 3: False Alarm Rate (Catch Trials).** Suppose that the psychophysical experiment we are doing requires the observers to respond “Yes” or “No” depending on whether they detected a stimulus or not. This is called a yes-no task. This panel represents what would happen if on some trials, called catch trials, we withheld the stimulus without informing the observers. In Table 4.1 we show the nomenclature of the four possible outcomes in such an experiment, generated by two possible responses to two types of stimuli.

According to a naive conception of detection, observers are “honest”; they would never respond “Yes” on catch trials. In other words, they would never produce false alarms. That is why $p(\text{fa}) = 0$ for all values of stimulus energy.

**Panel 4: Hit Rate (Signal Trials).** This panel shows the probability that the observer says “Yes” when the signal was presented (a hit) depends on whether the stimulus energy is above or below threshold. If it is above threshold, then $p(h) = 1$, otherwise $p(h) = 0$. The function that relates $p(\text{Yes})$, or hit rate, to stimulus energy is called the psychometric function.

**Panel 5: The ROC Space.** This panel is a plot of the hit rate as a function of the false-alarm rate. Here, where observers always respond “No” when in a $\bar{D}$ state, there is little point in such a diagram. Its value will become clear as we proceed.


**Fixed Energy Threshold—Guessing Observer**

Instead of assuming naive observers, we next assume sophisticated ones who know that some of the trials are catch trials, so some of the time they choose to guess. Let us compare this threshold theory, called high-threshold theory, with the most unrestricted form of two-state threshold theory (Figure 4.2). The general theory is unrestricted because it (a) allows observers to be in either a $D$ state or a $\bar{D}$ state on both signal and catch trials and (b) imposes no restrictions on when guessing occurs.

In contrast, high-threshold theory (Figure 4.3) has three constraints: (a) During a catch trial, the observer is always in a $\bar{D}$ state: $p(D|\text{catch}) = 0$. (b) When in a $D$ state, the observer always says “Yes”: $p(\text{Yes}|D) = 1$. (c) When in a $\bar{D}$ state, the observer guesses “Yes” (emits a false alarm) at a rate $p(\text{fa}) = g$, and “No” at a rate $1 - g$. So $p(\text{Yes}|\bar{D}) = g$. The theory is represented in Figure 4.1B, whose panels we discuss one by one.

**Panel 1: Threshold Location.** Unchanged from the corresponding panel in Figure 4.1A.

**Panel 2: Detection Probability.** Unchanged from the corresponding panel in Figure 4.1A.

**Panel 3: Catch Trials.** In this panel we show that observers, realizing that some stimuli are below threshold and wishing to be right as often as possible, may guess when in a $\bar{D}$ state. This increases the false-alarm rate.

**Panel 4: Signal Trials.** The strategy depicted here does not involve guessing when the observers are in a $D$ state. This panel shows that this strategy increases the observers’ hit rates when they are in a $\bar{D}$ state, that is, below the energy threshold. Note that the psychometric function does not rise from 0 but from $g$.

*Note:* With the help of Figure 4.3 we can see that when signal energy is 0, the hit rate is equal to the false-alarm rate, $g$. We begin by writing down the hit rate as a function of the probability of the observer being in a $D$ state when a signal is presented, $p(D|\text{signal})$, and the probability of the observer guessing, that is, saying “Yes” when in a $\bar{D}$ state, $p(\text{Yes}|\bar{D}) = g$:

$$p(h) = p(D|\text{signal}) + [1 - p(D|\text{signal})]g$$

$$= p(D|\text{signal})(1 - g) + g.$$  \hspace{1cm} (1)

When signal energy is 0, $p(D|\text{signal}) = 0$, and therefore $p(h) = g$. 

---

**TABLE 4.1** Outcomes in a Yes-No Experiment with Signal and Catch Trials (and their abbreviations)

<table>
<thead>
<tr>
<th>Stimulus Class</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signal</td>
<td>Hit (h)</td>
</tr>
<tr>
<td>Catch</td>
<td>False alarm (fa)</td>
</tr>
<tr>
<td></td>
<td>Miss (m)</td>
</tr>
<tr>
<td></td>
<td>Correct rejection (cr)</td>
</tr>
</tbody>
</table>
Figure 4.2  The general structure of a detection experiment, assuming two observer states.

Figure 4.3  The structure of a detection experiment, assuming high-threshold theory. This is a special case of the structure shown in Figure 4.2.
Panel 5: The ROC Diagram. This panel differs from the corresponding one in Figure 4.1A because the false-alarm rate has changed. When the observer is in a \( D \) state, \( p(\text{fa}) = p(\text{hit}) = g \) (represented by the unfilled dot); when the observer is in a \( D \) state, \( p(\text{fa}) = g \) and \( p(\text{hit}) = 1.0 \).

Variable Energy Threshold—Guessing Observer

The preceding versions of threshold theory are idealizations; step functions are nowhere to be found in psychophysical data. In response to this realization Jastrow (1888) assumed that the threshold varies from moment to moment. This conception is depicted in Figure 4.1C.

Panel 1: Threshold Location. The idea that the location of a momentary threshold follows a normal density function comes from Boring (1917). Subsequently, other density functions were proposed: the lognormal (Gaddum, Allen, & Pearce, 1945; Thurstone, 1928), the logistic (which is a particularly useful approximation to the normal; Bush, 1963; Jeffress, 1973), and the Weibull (Marley, 1989a, 1989b; Pearce, 1945; Thurstone, 1928), the logistic (which is a particular useful approximation to the normal; Bush, 1963; Jeffress, 1973), and the Weibull (Marley, 1989a, 1989b; Quick, 1974), to mention only three.

Panel 2: Detection Probability. This panel shows that when the threshold is normally distributed, the probability of being in a \( D \) state follows the cumulative distribution that corresponds to the density function of the momentary threshold. When that density is normal, this cumulative is sometimes called a normal ogive.

Panel 3: Catch Trials. Unchanged from the corresponding panel in Figure 4.1C.

Panel 2: Detection Probability. In this panel the energy of the stimulus is indicated by a downward-pointing arrow. The corresponding \( p(D) \) is indicated on the ordinate. For reasons we explain in a moment, this value is connected to a point on the ordinate of the ROC diagram.

Panel 3: Catch Trials. In this panel we assume that we have persuaded the observer to adopt four different guessing rates \( (g_1, \ldots, g_4) \) during different blocks of the experiment. The corresponding values are marked on the abscissa (the false-alarm rate) of the ROC diagram (Figure 4.1D, fifth panel).

Panel 4: Signal Trials. The general structure of a detection experiment, assuming two observer states, detect and \( \overline{D} \), is shown in Figure 4.2. In this figure (which is an augmented version of Figure 4.3) we show how to calculate the hit rate and the false-alarm rate, as well as which parts of this model are observable and which are hidden.

Panel 5: The ROC Diagram. In this panel hit rate and false-alarm rate covary and follow a linear function. In the note that follows we give the equation of this line and show that it allows us to estimate \( p(D|\text{signal}) \), which is the measure of the signal’s detectability.

Note: If in Equation 1 we let \( b = p(D|\text{signal}) \) and \( m = 1 - p(D|\text{signal}) \), and we recall that \( g = p(\text{“Yes”}|\overline{D}) = p(\text{fa}) \), then the equation of the ROC is \( p(\text{hit}) = b + mp(\text{fa}) \), a straight line. The intercept \( b \) gives the measure of the signal’s detectability: \( p(D|\text{signal}) \).

We can now understand the importance of the ROC diagram. Regardless of the detection theory we hold, it allows us to separate two aspects of the observer’s performance: stimulus detectability (or equivalently observer sensitivity) and observer bias. In high-threshold theory these measures are \( p(D|\text{signal}) \) and \( g \).
TABLE 4.2 The Observer’s Decision Problem in High-Threshold Theory

| Stimulus | $p(D|\text{signal})$ | $p(D|\text{catch})$ | $p(D|\text{signal})$ | $p(D|\text{catch})$ |
|----------|----------------------|----------------------|----------------------|----------------------|
| Signal   |                      |                      |                      |                      |
| Catch    |                      |                      |                      |                      |

There are two ways to manipulate an observer’s guessing rate: (a) Manipulate the probability of a catch trial or (b) use a payoff matrix. The observers’ goal is to guess whether a signal or a catch trial occurred; according to high-threshold theory all they know is they were in $D$ state or $\bar{D}$ state (see Table 4.2).

Observers do not know which type of trial ($t_1$ or $t_2$) caused the state they are presently experiencing (which we denote $e$). Assuming that they know the probabilities of the types of trial and the probabilities of the states they could be in, they can use Bayes’s rule for the probability of causes (Feller, 1968, p. 124) to determine the conditional probability of the cause (type of trial) given the evidence (their state), which is called the posterior probability of the cause. For example,

$$p(t_1|e) = \frac{p(e|t_1)p(t_1)}{p(e|t_1)p(t_1) + p(e|t_2)p(t_2)},$$

(2)

where $p(t_1|e)$ is the posterior probability of $t_1$, $p(e|t_1)$ and $p(e|t_2)$ are likelihoods (of their state given the type of trial), and $p(t_1)$ and $p(t_2)$ are the prior probabilities of the types of trial. In a different form,

$$\frac{p(t_1|e)}{p(t_2|e)} = \frac{p(e|t_1)p(t_1)}{p(e|t_2)p(t_2)},$$

where $\frac{p(e|t_1)}{p(e|t_2)}$ is the likelihood ratio, $\frac{p(t_1)}{p(t_2)}$ is the prior odds, and $\frac{p(t_1|e)}{p(t_2|e)}$ is the posterior odds in favor of $t_1$.

In high-threshold theory, the posterior odds in favor of signal is

$$\frac{p(\text{signal}|D)}{p(\text{catch}|D)} = \frac{p(D|\text{signal})p(\text{signal})}{p(D|\text{catch})p(\text{catch})}.$$

Suppose that the observers are in a $D$ state, and they believe that half the trials are catch trials [$p(\text{signal}) = p(\text{catch}) = .5$], and that the threshold happens to be at the median of the distribution of energies [$p(D|\text{signal}) = p(\bar{D}|\text{signal}) = .5$], then

$$\frac{p(\text{signal}|D)}{p(\text{catch}|D)} = \frac{.5}{.5} = 1.$$ Because the posterior odds in favor of signal are infinite, observers have no reason to guess. But if they are in a $\bar{D}$ state and hold the same beliefs, then the posterior odds are

$$\frac{p(\text{signal}|\bar{D})}{p(\text{catch}|\bar{D})} = \frac{.5}{.5} = 1,$$

or $p(\text{signal}|\bar{D}) = \frac{1}{3}$, that is, they should believe that one third of the $\bar{D}$ trials will be signal trials, and they will increase the number of correct responses by guessing.

We can use two methods to induce observers to change their guessing rate.

Prior Probabilities. In the example above, the prior odds were 1. If we change these odds, that is, increase or decrease the frequency of signal trials, the posterior probability of signal in $\bar{D}$ state will increase or decrease correspondingly. As a result the observer’s guessing rate will increase or decrease.

Payoff Matrix. We can also award our observers points (which may correspond to tangible rewards) for each of the outcomes of a trial (Table 4.1), as the examples in Table 4.3 show.

We could reward them for correct responses by giving them $B(h)$ or $B(cr)$ points for hits or correct rejections, and punish them for errors by subtracting $C(fa)$ or $C(m)$ points for false alarms or misses. To simplify Table 4.3 we set $C(fa) = C(m) = 0$. It is easy to see that when we bias the observer toward “Yes,” the guessing rate will increase, and when we bias the observer toward “No,” it will decrease.

The ROC curve is a particular case of a general framework for thinking about perception—the Bayesian approach to perception. It is summarized in Figure 4.4 (Mamassian, Landy, & Maloney, in press).

This diagram represents a prescriptive framework: how one should make decisions. Bayes’s rule is the correct way to combine background information with present data. Furthermore, there is a considerable body of work on the correct way to combine the resulting posterior distribution with information about costs and benefits of the possible decisions (the

TABLE 4.3 Payoff Tables for Responses to Signal and Catch Trials (in points)

<table>
<thead>
<tr>
<th>General Case</th>
<th>Bias Toward “Yes”</th>
<th>Bias Toward “No”</th>
<th>No Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>Response</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>$B(h)$</th>
<th>$C(fa)$</th>
<th>$C(m)$</th>
<th>$B(h)$</th>
<th>$C(fa)$</th>
<th>$C(m)$</th>
<th>$B(h)$</th>
<th>$C(fa)$</th>
<th>$C(m)$</th>
<th>$B(h)$</th>
<th>$C(fa)$</th>
<th>$C(m)$</th>
<th>$B(h)$</th>
<th>$C(fa)$</th>
<th>$C(m)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signal</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Catch</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
gain function in Figure 4.4). When it is used as a prescriptive framework, it is called an *ideal observer*.

Observers in the laboratory, or parts of the visual system, are not subject to prescriptions. What they actually do is shown in Figure 4.5, which is a descriptive framework: how observers (or, more generally, systems) actually make decisions (Kubovy & Healy, 1980; Tanner & Sorkin, 1972).

The diagram identifies four opportunities for the observer to deviate from the normative model:

1. Observers do not know the likelihood function or the prior probabilities unless they learned them. They are unlikely to have learned them perfectly; that is why we have replaced the “likelihood function” and the “prior distributions” of Figure 4.4 with their subjective counterparts.

2. Instead of combining the “likelihood function” and the “prior distributions” by using Bayes’s rule, we assume that the observer has a computer that combines the subjective counterparts of these two sources of information. This computer may not follow Bayes’s rule.

3. The subjective gain function may not simply reflect the payoffs. Participants in an experiment may not only desire to maximize gain; they may also be interested in exploring the effect of various response strategies.

4. Instead of combining the “posterior distribution” with the “gain function” in a way that will maximize gain, we assume that the observer has a biaser that combines the subjective counterparts of these two sources of information.

**Problems with Threshold Theories**

We have seen that the ROC curve for high-threshold theory is linear. Such ROC curves are never observed. Let us consider an example. In the animal behavior literature, a widely accepted theory of discrimination was equivalent to high-threshold theory. Cook and Wixted (1997) put this theory to a test in a study of six pigeons performing a texture discrimination. On each trial the pigeons were show one of many potential texture patterns on a computer screen (Figure 4.6).

In some of these patterns all the texture elements were identical in shape and color. Such patterns were called Same (Figure 4.6D). In the other patterns some of the texture elements
differed in color (Figure 4.6A), shape (Figure 4.6B), or both (Figure 4.6C); they were called Different. In the test chamber two food hoppers were available; one of them delivered food when the texture was Same, the other when the texture was Different. Choosing the Different hopper can be taken to be analogous to a “Yes” response, and choosing the Same hopper analogous to a “No” response. To produce ROC curves, Cook and Wixted (1997) manipulated the prior probabilities of Same and Different patterns. The ROC curves were nonlinear, as Figure 4.7 shows.

Signal Detection Theory

Nonlinear ROC curves require a different approach to the problem of detection, called signal detection theory, summarized in Figure 4.8. The key innovation of signal detection theory is to assume that (a) all detection involves the detection of a signal added to background noise and (b) there is no observer threshold (as we will see, this does not mean that there is no energy threshold).
Variable Criterion

The observers’ task is to decide on every trial whether it was a signal trial or a catch trial. The only evidence they have is the stimulus, \(x\), which could have been caused by \(N\) or \(SN\). As with high-threshold theory, they could use Bayes’s rule to calculate the posterior probability of \(SN\),

\[
p(SN|x) = \frac{\ell(x|SN)p(SN)}{\ell(x|SN)p(SN) + \ell(x|N)p(N)}.
\]

The expressions \(\ell(x|SN)\) and \(\ell(x|N)\), explained in Figure 4.8E, are called *likelihoods*. (We use the notation \(\ell(\cdot)\) rather than \(p(\cdot)\), because it represent a density, not a probability.) They could also calculate the posterior odds in favor of \(SN\),

\[
\frac{p(SN|x)}{p(N|x)} = \frac{\ell(x|SN)p(SN)}{\ell(x|N)p(N)}.
\]
(We need not assume that observers actually use Bayes’s rule, only that they have a sense of the prior odds and the likelihood ratios, and that they do something akin to multiplying them.)

Once the observers have calculated the posterior probability or odds, they need a rule for saying “Yes” or “No.” For example, they could choose to say “Yes” if \( p(SN/H) \geq 0.5 \). This strategy is by and large equivalent to choosing a value of \( e \) below which they would say “No,” and otherwise they would say “Yes.” This value of \( e, e_c \), is called the criterion.

We have already seen how we can generate an ROC curve by inducing observers to vary their guessing rates. These procedures—manipulating prior probabilities and payoffs—induce the observers to vary their criteria (Figures 4.8C and 4.8D) from lax (\( e_c \) is low, hit rate and false-alarm rate are high) to strict (\( e_c \) is high, hit rate and false-alarm rate are low), and produce the ROC curve shown in Figure 4.8F. Different signal energies (Figure 4.8G) produce different ROC curves. The higher \( d^\prime \), the further the ROC curve is from the positive diagonal.

The ROC Curve; Estimating \( d^\prime \)

The easiest way to look at signal detection theory data is to transform the hit rate and false-alarm rate into log odds. To do this, we calculate

\[
H = k \ln \frac{p(h)}{1 - p(h)} \quad \text{and} \quad F = k \ln \frac{p(fa)}{1 - p(fa)},
\]

where \( k = \frac{\pi}{\sqrt{3}} \approx 0.55133 \) (which is based on a logistic approximation to the normal). The ROC curve will often be linear after this transformation. We have done this transformation with the data of Cook and Wixted (1997; see Figure 4.9).

If we fit a linear function, \( H = b + mF \), to the data, we can estimate \( d^\prime = \frac{b}{m} \) and \( \sigma_{SN} = \frac{1}{m} \), the standard deviation of the \( SN \) distribution (assuming \( \sigma_N = 1 \)). Figure 4.9 shows these computations. (This analysis is not a substitute for more detailed and precise ones, such as Eng, 2001; Kestler, 2001; Metz, 1998; Stanislaw & Todorov, 1999.)

Energy Thresholds and Observer Thresholds

It is easy to misinterpret the signal detection theory’s assumption that there are no observer thresholds (a potential misunderstanding detected and dispelled by Krantz, 1969). The assumption that there are no observer thresholds means that observers base their decisions on evidence (the likelihood ratio) that can vary continuously from 0 to infinity. It need not imply that observers are sensitive to all signal energies. To see how such a misunderstanding may arise, consider Figures 4.8A and 4.8B. Because the abscissas are labeled “energy,” the panels appear to be representations of the input to a sensory system. Under such an interpretation, any signal whatsoever would give rise to a signal + noise density that differs from the noise density, and therefore to an ROC curve that rises above the positive diagonal.

To avoid the misunderstanding, we must add another layer to the theory, which is shown in Figure 4.10. Rows (a) and (c) are the same as rows (a) and (b) in Figure 4.8. The abscissas in rows (b) and (d) in Figure 4.10 are labeled “phenomenal evidence” because we have added the important but plausible assumption that the distribution of the evidence experienced by an observer may not be the same as the distribution of the signals presented to the observer’s sensory system (e.g., because sensory systems add noise to the input, as Gorea & Sagi, 2001, showed). Thus in row (b) we show a case where the signal is not strong enough to cause a response in the observer: the signal is below this observer’s energy threshold. In row (d) we show a case of a signal that is above the energy threshold.

Some Methods for Threshold Determination

Method of Limits

Terman and Terman (1999) wanted to find out whether retinal sensitivity has an effect on seasonal affective disorder (SAD;
reviewed by Mersch, Middendorp, Bouhuys, Beersma, & Hoofdakker, 1999). To determine an individual’s retinal sensitivity, they used a psychophysical technique called the method of limits and studied the course of their dark adaptation (for a good introduction, see Hood & Finkelstein, 1986, §4).

Terman and Terman (1999) first adapted the participants to a large field of bright light for 5 min. Then they darkened the room and turned on a dim red spot upon which the participants were asked to fix their gaze (Figure 4.11). Because they wanted to test dark adaptation of the retina at a region that contained both rods and cones, they tested the ability of the participants to detect a dim, intermittently flashing white disk below that fixation point. Every 30 s, the experimenter gradually adjusted the target intensity upward or downward and then asked the participant whether the target was visible. When target intensity was below threshold (i.e., the participant responded “no”) the experimenter increased the intensity until the response became “yes.” The experimenter then reversed the progression until the subject reported “no.” Figure 4.12 shows the data for one patient with winter depression. The graph shows that the transition from “no” to “yes” occurs at a higher intensity than the transition from “yes” to “no.” This is a general feature of the method of limits, and it is a manifestation of a phenomenon commonly seen in perceptual processes called hysteresis.

Figure 4.10 Revision of Figure 4.8 to show that energy thresholds are compatible with the absence of an observer threshold.

Figure 4.11 Display for the seasonal affective disorder experiment (Terman & Terman, 1999). Rules of thumb: 20° of visual angle is the width of a hand at arm’s length; 2° is the width of your index finger at arm’s length.
Terman and Terman (1999) overcame the problem of hysteresis by taking the mean of these two values to characterize the sensitivity of the participants. The cone and rod thresholds of all the participants were lower in the summer than in the winter. However, in winter the 24 depressed participants were more sensitive than were the 12 control participants. Thus the supersensitivity of the patients in winter may be one of the causes of winter depression.

**Method of Constant Stimuli**

Barraza and Colombo (2001) wanted to discover conditions under which glare hindered the detection of motion. Their stimulus is one commonly used to explore motion thresholds: a drifting sinusoidal grating, illustrated in Figure 4.13 (Graham, 1989, §2.1.1, defines such gratings).

The lowest velocity at which such a grating appears to be drifting consistently is called the lower threshold of motion.

---

**Figure 4.12** Visual detection threshold during dark adaptation for a patient with winter depression. The curves are exponential functions for photopic (cone) and scotopic (rod) segments of dark adaptation. Source: From “Photopic and scotopic light detection in patients with seasonal affective disorder and control subjects,” by J. S. Terman and M. Terman, 1999, *Biological Psychiatry*, 46, Figure 1. Copyright 1999 by Society of Biological Psychiatry. Reprinted with permission.

---

**Figure 4.13** (A) The sinusoidal grating used by Barraza and Colombo (2001) drifted to the right or to the left at a rate that ranged from about one cycle per minute (0.0065 cycles per second, or Hz) to about one cycle every 3.75 s (0.0104 Hz). The grating was faded in and out, as shown in Figure 4.14. It is shown here with approximately its peak contrast. (B) The luminance profile of a sinusoidal grating, and its principal parameters.
To determine the LTM, Barraza and Colombo (2001) showed the observers two gratings in succession. One was drifting to the right, and the other was drifting to the left. The observer had to report whether the first or the second interval contained the leftward-drifting grating. Such tasks are called forced-choice tasks. More specifically, this is an instance of a temporal two-alternative forced-choice task (2AFC; to learn more about forced-choice designs, see Macmillan & Creelman, 1991, chap. 5, and Hartmann, 1998, chap. 24).

To simulate the effect of glare, Barraza and Colombo (2001) used an incandescent lamp located 10° away from the observer’s line of sight. On each trial, they first turned on the glare stimulus, and then after a predetermined interval of time, they showed the drifting grating. Because neither the glare stimulus nor the grating had an abrupt onset, they defined the effective onset of each as the moment at which the stimulus reached a certain proportion of its maximum effectiveness (as shown in Figure 4.14). The time interval between the onset of two stimuli is called stimulus-onset asynchrony (SOA). In this experiment the SOA between the glare stimulus and the drifting grating took on one of five values: 50, 150, 250, 350, or 450 ms.

Barraza and Colombo (2001) were particularly interested in determining whether the moments just after the glare stimulus was turned on were the ones at which the glare was the most detrimental to the detection of motion (i.e., it caused the LTM to rise). To measure the LTM for each condition, they used the method of constant stimuli: They presented the gratings repeatedly at a given drift velocity so that they could estimate the probability that the observer could discriminate between left- and right-drifting gratings.

To calculate the LTM, they plotted the proportion of correct responses for a given SOA as a function of the rate at which the grating drifted (Figure 4.15, top panel). They then fitted a Weibull function to these data and determined the LTM by finding the grating velocity that corresponded to 80% correct responses (dashed lines). Although there is no substitute for publishing the best-fitting normal, logistic, or Weibull distribution function to such data (using logistic regression for a logistic distribution or a probit model for the normal; Agresti, 1996), the easiest way to look at such data is to transform the percentage of correct data into log odds. Let us denote motion frequency by \( f \) and the corresponding proportion of correct responses by \( \pi(f) \). We plot the log-odds of being right (using the natural logarithm, denoted by \( \ln \)) as a function of \( f \). In other words, we fit a linear function, \( \ln \frac{\pi(f)}{1-\pi(f)} = \alpha + \beta f \), to the data obtained. Figure 4.15, bottom panel, shows the results. Fitting the linear regression does not require specialized software, and the results are usually close to estimates obtained with more complex fitting routines.

**Adaptive Methods**

Adaptive methods combine the best features of the method of limits and forced-choice procedures. Instead of exploring the response to many levels of the independent variable, as in the method of constant stimuli, adaptive methods quickly
converge onto the region around the threshold. In this they resemble the method of limits. But adaptive methods do not suffer from hysteresis, which is characteristic of the method of limits.

For example, Näsänen, Ojanpää, and Kojo (2001) used a staircase procedure (Wetherill & Levitt, 1965) to study the effect of stimulus contrast on observers’ ability to find a letter in an array of numerals (Figure 4.16). The display was first presented at a duration of 4 s. After three consecutive correct responses, its duration was reduced by a factor of 1.26 \((\log_{10} 1.26 / 0.1)\), and after each incorrect response the duration was increased by the same factor. As a result, the duration was halved in three steps (4, 3.17, 2.52, 2.00, ..., 0.10, ..., s), or doubled (4, 5, 6.4, 8, ..., s). When the sequence reversed from ascending to descending (because of consecutive correct responses) or from descending to ascending (because of an error), a reversal was recorded. The procedure was stopped after eight reversals. The length of the procedure ranged from 30 to 74 trials. Since the durations were on a logarithmic scale, the threshold was computed by taking the geometric mean of the eight reversal durations.

What does this staircase procedure estimate? It estimates the array duration for which the observer can correctly identify the letter among the digits 79% of the time \((p_c = .79)\). Let us see why. Suppose that we are presenting the array at an observer’s threshold duration. At this level, the procedure has the same chance of (a) going down after three correct responses as it has of (b) going up after one error. So \(p_c^3 = 1 - p_c = .5\), which gives \(p_c = \sqrt[3]{.5} = .79\) (for further study: Hartmann, 1998; Macmillan & Creelman, 1991).

Näsänen et al. (2001) varied the contrast of the letters and the size of the array. The measure of contrast they used is called the Michelson contrast: \(c = \frac{L_{\text{max}} - L_{\text{min}}}{L_{\text{max}}}\), where \(L_{\text{max}}\) is the maximum luminance (in this case the background luminance), and \(L_{\text{min}}\) is the minimum luminance (the luminance of the letters). In the notation of Figure 4.14, \(L_0 + mL_0 = L_{\text{max}}\) and \(L_0 - mL_0 = L_{\text{min}}\). Figure 4.17 shows that search time decreased when set size was decreased and when contrast was increased. Using an eye tracker, the authors also found that the number of fixations and their durations decreased with increasing contrast, from which they concluded that “visual span, that is, the area from which information can be collected in one fixation, increases with increasing contrast” (Näsänen et al., 2001, p. 1817).
THE “STRUCTURE” OF THE VISUAL ENVIRONMENT AND PERCEPTION

Regularities of the Environment

As we saw earlier, the contemporary view of perception maintains that perceptual theory requires that we understand both our environment and the perceiver. In the preceding section we reviewed some methods used to measure the perceptual capacity of perceivers. In this section we turn our attention to the environment and ask how one can determine (a) the regularities of the environment and (b) the extent to which perceivers use them.

The structure of the environment and the capacities of the perceiver are not independent. When researchers look for statistical regularities in the environment, they are guided by beliefs about the aspects of the environment that are relevant to perception. These beliefs are based on the phenomenology of perception as well as on psychophysical and neural evidence. We see that insights from the phenomenology and neuroscience of vision interact to establish a correspondence between the structure of the environment and the mechanisms of perception.

The phenomenology of perception, championed by Gestalt psychologists and their successors in the twentieth century (Ellis, 1936; Kanizsa, 1979; Koffka, 1935; Köhler, 1929; Kubovy, 1999; Kubovy & Gepshtein, in press; Wertheimer, 1923), is a prominent source of ideas about the kinds of information the visual system seeks in the environment. The Gestaltist program of research revealed many examples of correlation between the relational properties of visual stimulation and visual experience. The Gestalt psychologists believed that the regularities of experience arise in the brain by virtue of the intrinsic properties of the brain, independent of the regularities of the environment. On this view, the experience-environmental correlation occurs because the brain is a physical system, just as the environment is, and hence they operate along the same dynamic principles.

This Gestalt approach—known as psychophysical isomorphism—has been criticized by many, including Brunswik (1969), who nevertheless considered the factors of perceptual organization discovered by the Gestalt psychologists as “guides to the life-relevant properties of the remote environmental objects.” Brunswik and Kamiya (1953, pp. 20–21) argued that

the possibility of such an interpretation [of the factors of perceptual organization] hinges upon the “ecological validity” of these factors, that is, their objective trustworthiness as potential indicators of mechanical or other relatively essential or enduring characteristics of our manipulable surroundings.

Brunswik anticipated the modern interest in the statistical regularities of the environment by several decades; he was the first (Barlow, in press; Geisler, Perry, Super, & Gallogly, 2001) to propose ways of measuring these regularities (Brunswik & Kamiya, 1953).

Another prominent champion of environmental factors in perception was James J. Gibson, whose ecological realism we reviewed earlier. We will only add here that Gibson derived his ecological optics from an analysis of environment that is hard to classify as other than phenomenological. Epstein and Hatfield (1994, p. 174) put it clearly:

We cannot shake the impression that “the world of ecological reality” is largely coextensive with the world of phenomenal reality, and that the description of ecological reality, although couched in the language of “ecological physics,” nonetheless is an exercise in phenomenology. . . . Gibson’s distinction between ecological reality and physical reality parallels the Gestalt distinction between the behavioral environment and geographical environment.

Besides visual phenomenology, an important source of ideas about the information relevant for visual perception is visual neuroscience. The evidence of visual mechanisms selective to particular “features” of stimulation (such as the orientation, spatial frequency, or direction of motion of luminance edges) suggests the aspects of stimulation in which the brain is most interested. As we mentioned earlier, this line of thought can be challenged by the level of analysis argument: Particular features could be optimal stimuli for single cells...
not because the low-level features themselves are of interest for perception, but because these features make convenient stepping-stones for the detection of higher order features in the stimulation.

The view of a perceptual system as a collection of devices sensitive to low-level features of stimulation raises the difficult question of how such features are combined into the meaningful entities of our visual experience. This question, known as the binding problem, has two aspects: (a) How does the brain know which similar features (such as edges of a contour) belong to the same object in the environment? and (b) How does the brain know which different features (e.g., pertaining to the form and the color) should be bound into the representation of a single object? These questions could not be answered without understanding the statistics of optical covariation (MacKay, 1986), as we argue in the next section. That the visual system uses such statistical data is suggested by physiological evidence that visual cortical cells are concurrently selective for values on several perceptual dimensions rather than being selective to a single dimension (Zohary, 1992). We now briefly review the background against which the idea of optical covariation has emerged in order to prepare the ground for our discussion of contemporary research on the statistics of natural environment.

Redundancy and Covariation

Following the development of the mathematical theory of communication and the theory of information (Shannon & Weaver, 1949; Wiener, 1948; see also chapter by Proctor and Vu in this volume), mathematical ideas about information-handling systems began to influence the thinking of researchers of perception. Although the application of these ideas to perception required a good deal of creative effort and insight, the resulting theories of perception looked much like the theories of human-engineered devices, “receiving” from the environment packets of “signals” through separable “channels.” Whereas the hope of assigning precise mathematical meaning to such notions as information, feedback, and capacity was to some extent fulfilled with respect to low-level sensory processes (Graham, 1989; Watson, 1986), it gradually became clear that a rethinking of the ideas inspired by the theory of communication was in order (e.g., Nakayama, 1998).

An illuminating example of such rethinking is the evolution of the notion of redundancy reduction into the notion of redundancy exploitation (see Barlow, 2001, in press, for a firsthand account of this evolution). The notion of redundancy comes from Shannon’s information theory, where it was a measure of nonrandomness of messages (see Attnave, 1954, 1959, p. 9, for a definition). In a structureless distribution of luminances, such as the snow on the screen of an untuned TV set, there are no correlations between elements in different parts of the screen. In a structure-bearing distribution there exist correlations (or redundancy) between some aspects of the stimulation, so that we can to some extent predict one aspect of the stimulation from other aspects. As Barlow (2001) put it, “any form of regularity in the messages is a form of redundancy, and since information and capacity are quantitatively defined, so is redundancy, and we have a measure for the quantity of environmental regularities.”

On Attnave’s view, and on Barlow’s earlier view, a purpose of sensory processing was to reduce redundancy and code information into the sensory “channels of reduced capacity.” After this idea dominated the literature for several decades, it has become increasingly clear—from factual evidence (such as the number of neurons at different stages of visual processing) and from theoretical considerations (such as the inefficiency of the resulting code)—that the redundancy of sensory representations does not decrease in the brain from the retina to the higher levels in the visual pathways. Instead, it was proposed that the brain exploits, rather than reduces, the redundancy of optical stimulation.

According to this new conception of redundancy, the brain seeks redundancy in the optical stimulation and uses it for a variety of purposes. For example, the brain could look for a correlation between the values of local luminance and retinal distances across the scene (underwriting grouping by proximity; e.g., Ruderman, 1997), or it could look for correlations between local edge orientations at different retinal locations (underwriting grouping by continuation; e.g., Geisler et al., 2001). The idea of discovering such correlations between multiple variables is akin to performing covariational analysis on the stimulation. MacKay (1986, p. 367) explained the utility of covariational analysis:

The power of covariational analysis—asking “what else happened when this happened?”—may be illuminated by its use in the rather different context of military intelligence-gathering. It becomes effective and economical, despite its apparent crudity, when the range of possible states of affairs to be identified is relatively small, and when the categories in terms of which covariations are sought have been selected or adjusted according to the information already gathered. It is particularly efficacious where many coincidences or covariations can be detected cheaply in parallel, each eliminating a different fraction of the set of possible states of affairs. To take an idealized example, if each observation were so crude that it eliminated only half of the range of possibilities, but the categories used were suitably orthogonalized (as in the game of “Twenty questions”), only 100 parallel analyzers would be needed in principle to identify one out of $2^{100}$, or say $10^{50}$, states of affairs.
In the remainder of this chapter we explore an instance of covariational analysis applied by Geisler et al. (2001) to grouping by good continuation (Field, Hayes, & Hess, 1993; Wertheimer, 1923). We see how Geisler et al. used this analysis to ask whether the statistics of contour relationships in natural images correspond to the characteristics of the perceptual processes of contour grouping in human observers.

**Co-occurrence Statistics of Natural Contours**

Geisler et al. (2001) used the images shown in Figure 4.18 as a representative sample of visual scenes. In these images they measured the statistics of relations between contour segments. In every image they found contour segments, called *edge elements*, using an algorithm that simulated the properties of neurons in the primary visual cortex that are sensitive to edge orientations. This produced for every image a set of locations and orientations for each edge element. Figure 4.19A shows an example of an image with the selected edge elements (discussed later). Geisler et al. submitted these data to a statistical analysis of relative orientations and distances between every possible pair of edges within every image. We now consider what relations between the edge elements the authors measured and how they constructed the distributions of these relations.

The geometric relationship between a pair of edge elements is determined by three parameters explained in Figure 4.20. The relative position of element centers is specified by two parameters: distance between element centers, $d$, and the direction of the virtual line connecting elements centers, $\phi$. The third parameter, $\theta$, measures the relative orientation of the elements, called *orientation difference*. For every edge element in an image, Geisler et al. (2001) considered the pairs of this element with every other edge elements in the image and, within every pair, measured the three parameters: $d$, $\theta$, and $\phi$. The authors repeated this procedure for every edge element in the image and obtained the probability of every magnitude of the three parameters of edge relationships. They called the resulting quantity the *edge co-occurrence (EC) statistic*, which is a three-dimensional probability density function, $p(d, \theta, \phi)$, as we explain later. Geisler et al. used two methods to obtain edge co-occurrence statistics: One was independent of whether the elements belonged to the same contour or not, whereas the other took this information into account. The authors called the resulting statistics *absolute* and *Bayesian*, respectively. We now consider the two statistics.

**Absolute Edge Co-occurrence**

This EC statistic is called absolute because it does not depend on the layout of objects in the image. In other words, those edge elements that belonged to different contours in the image contributed to the absolute EC statistic to the same extent as did the edge elements that belonged to the same contour. As Geisler et al. (2001) put it, this statistic was measured “without reference to the physical world.”

Figures 4.19B and 4.19C show two properties of absolute EC statistic averaged across the images. Because the covariational analysis used by Geisler et al. (2001) concerns a relation between three variables, the results are easier to understand when we think of varying only one variable at a time, while keeping the two other variables constant.

Consider first Figure 4.19B, which shows the most frequent orientation differences for a set of 6 distances and 36 directions of edge-element pairs. To understand the plot, imagine a short horizontal line segment, called a *reference element*, in the center of a polar coordinate system ($d$, $\phi$). Then imagine another line segment—a *test element*—at a radial distance $d_t$ and direction $\phi_t$ from the reference element. Now rotate the test element around its center until it is aligned with the most likely orientation difference $\theta$ at this location. Then color the segment, using the color scale shown in the figure, to indicate the magnitude of the relative probability of this most likely orientation difference. (The probability is called

![Figure 4.18](image_not_available)
The “Structure” of the Visual Environment and Perception 113

“relative” to indicate that it was normalized such that the highest probability in the plot was 1.0). Figure 4.19B, which shows such orientation differences, demonstrates that for 6 distances and 36 directions of the test element, the edge elements are likely to be roughly parallel to the reference element. Geisler et al. (2001, p. 713) concluded,

This result shows that there is a great deal of parallel structure in natural images, presumably due to the effects of growth and erosion (e.g., the parallel sides of a branch, parallel geological strata, etc.), perspective projection (e.g., the elongation of surface markings due to slant), shading (e.g., the illumination of a branch often produces a shading contour parallel to the side of the branch), and so on.

Now consider Figure 4.19C, which shows the most frequent directions for the same set of distances and directions of edge element pairs as in Figure 4.19B. To understand this plot, imagine you choose a test element under an orientation difference $\theta$ and a distance $d$ and rotate it around the center of polar coordinates (i.e., along a circumference with radius $d$) until it finds itself at the most likely direction $\phi$ for the given distance and orientation difference. Figure 4.19C shows that in the resulting pattern the test elements are approximately cocircular.

Figure 4.19  Statistics of edge co-occurrence in the sample images shown in Figure 4.18.  Source: Copyright 2001 by Elsevier Science Ltd. Reprinted with permission. See insert for color version of this figure.

Figure 4.20  Parameters of the relationship between two edge elements: distance $d$, direction $\phi$, and orientation difference $\theta$.  Source: Copyright 2001 by Elsevier Science Ltd. Reprinted with permission.
with the reference elements; that is, the most likely edge elements can be connected through the contours of minimal change of curvature. Geisler et al. (2001, p. 713) concluded that the absolute EC statistic “reflects the relatively smooth shapes of natural contours, and . . . provides direct evidence that the Gestalt principle of good continuation has a general physical basis in the statistics of the natural world.”

The authors reported that the same “basic pattern” as in Figures 4.19B and 4.19C occurred in the statistics obtained from all the images, as well as in the analysis of edges under different spatial scales. As a control, the authors ascertained that in the images containing random patterns (white noise), the absolute statistic of EC was random.

**Bayesian Edge Co-occurrence**

Before we explain this statistic, let us briefly recall the relevant ideas of Bayesian inference, which we have already encountered in the section on signal detection theory. In the context of a detection experiment, we saw that when observers generate two hypotheses about the state of affairs in the world (“noise trial” vs. “signal plus noise trial”) the relevant evidence can be measured by taking the ratio of the likelihoods of events associated with the two hypotheses (Figure 4.8E). The resulting quantity (the likelihood ratio) can be compared with another quantity (the criterion) to adjudicate between the hypotheses.

Similar to the conditions of a detection experiment, in measuring the EC statistics one can pit two hypotheses against each other with respect to every pair of edge elements: $C$, “the elements belong to the same contour” and $\sim C$, “the elements do not belong to the same contour.” The relevant evidence can be expressed in the form of a likelihood ratio:

$$\ell(d, \theta, \phi) = \frac{p(d, \theta, \phi|C)}{p(d, \theta, \phi|\sim C)},$$

where $p(d, \theta, \phi|C)$ and $p(d, \theta, \phi|\sim C)$ are the conditional probabilities of a particular relationship $\{d, \theta, \phi\}$ between edge elements to occur, when the elements belong or do not belong to the same contour, respectively. (We explain how to obtain the criterion in a moment.)

Geisler et al. (2001) measured the likelihood ratio for every available relationship $\{d, \theta, \phi\}$ as follows. In every image, observers were presented with a set of highlighted pixels (colored red in the example image in Figure 4.19A) corresponding to the centers of edge elements detected in the image. Using a computer mouse, observers assigned sets of highlighted pixels to the perceived contours in the image. Thus observers reported about the belongingness of edge elements to contours in every image. With this information Geisler et al. conditionalized the absolute probabilities of EC by whether the edge elements within every pair belonged to the same contour or not, that is, to obtain the likelihoods $p(d, \theta, \phi|C)$ and $p(d, \theta, \phi|\sim C)$.

The resulting distribution of $L(d, \theta, \phi)$ is shown in Figure 4.19D, again using a color scale, averaged across all the sample images and two observers. (The two observers largely agreed about the assignment of edges to contours, with the correlation coefficient between the two likelihood distributions equal to .98.) In contrast to the plots of absolute statistics in Figures 4.19B and 4.19C, the plot of conditional EC in Figure 4.19D shows all 36 orientations at every location in the system of coordinates $(d, \phi)$. The distribution of $L(d, \theta, \phi)$ shows that edge elements are more likely to belong to the same contour than not (when $L[d, \theta, \phi] > 1.0$, labeled from green to red in Figure 4.19D), within two symmetrical wedge-shaped regions on the sides of the reference edge element.

Why measure the Bayesian statistic of EC in addition to the absolute statistics? The Bayesian statistic allows one to construct a normative model (i.e., a prescriptive ideal observer model; Figure 4.4) of perceptual grouping of edge elements. Besides informing us on how the properties of element relations covary in natural images (which is already accomplished in absolute statistics), the Bayesian statistic tells us how the covariance of edge elements connected by contours differs from the covariance of edge elements that are not connected. As a result, the Bayesian statistic allows one to tell whether human performance in an arbitrary task of perceptual grouping by continuation is optimal or not. Human performance in such a task is classified as optimal if human observers assign edge elements to contours with the same likelihood as is prescribed by the Bayesian statistic. In the next section we see how Geisler et al. (2001) constructed the ideal observer model of grouping by continuation and how they compared its performance with the performance of human observers.

**Predicting Human Performance from the Statistics of Natural Images**

**Psychophysical Evidence of Grouping by Good Continuation**

To find out whether human performance in grouping by good continuation agrees with the statistics of EC in natural images, Geisler et al. (2001) conducted a psychophysical experiment. They used a stimulus pattern for which they could derive the predictions of grouping from their statistical data and pit the predictions against the performance of human observers. An example of the stimulus pattern is shown in Figure 4.21A.
The stimulus consists of a set of aligned line segments (arranged in this example in a nearly vertical path to the right of the vertical midline), embedded in the background of randomly oriented line segments. Observers viewed successive presentations of two patterns, one containing both the target path and the background noise and one containing only the background noise. Their order was randomized. The task was to tell which of the two presentations contained the target path.

Geisler et al. (2001) varied the length of the path, the amplitude of the path deviations from a straight line, and path noisiness (due to the range of random orientations of the line segments comprising the path) to generate up to 216 classes of random contour shape. The data from the psychophysical experiments provided the authors with a “detailed parametric measurement of human ability to detect naturalistic contours in noisy backgrounds” (p. 717).

To generate the predictions of contour grouping from the EC statistics, Geisler et al. (2001) needed a function that determines which pairs of edge elements group together. The authors derived two such local grouping functions (Figures 4.19E and 4.19F)—one based on the absolute statistic and one based on the Bayesian statistic—which we explore in detail in a moment. Because Geisler et al. measured EC for pairs of edge elements, they used a transitivity rule to construct contours consisting of more than two elements: “if edge element \(a\) binds to \(b\), and \(b\) binds to \(c\), then \(a\) becomes bound to \(c\).” Using this rule, Geisler et al. could predict which target paths are seen in their stimuli, using the local grouping functions derived from the statistics of natural images: An example of grouping by continuation from image statistics is shown in Figure 4.21B. We consider the Bayesian local grouping function first, because it requires fewer parameters than does the absolute local grouping function.

**Bayesian Local Grouping Function**

As we saw earlier, the likelihood ratio at every location in the \((d, \phi)\) space in Figure 4.19D tells, for 36 orientation differences, how likely it is that the edge elements belong to the same contour as the reference element. To decide whether two edge elements belong to the same contour, for any particular relationship \(d, \theta, \phi\) between the elements, the corresponding likelihood ratio can be compared with a criterion, which Geisler et al. (2001) called a binding criterion, \(\beta\). As the signal detection theory prescribes, the ideal binding criterion is equal to the ratio of prior probabilities (called prior odds, as discussed earlier):

\[
\beta = \frac{p(\neg C)}{p(C)} = \frac{1 - p(C)}{p(C)}
\]

where \(p(C)\) and \(p(\neg C)\) are the probabilities of two edge elements to belong or not to belong, respectively, to the same contour.

The prior odds \(\beta\) were available to Geisler et al. (2001) directly from the Bayesian EC statistic. In a true ideal observer model of grouping by good continuation, the local grouping function would have to completely determine which edge elements should group solely from the statistics of natural images (i.e., with no free parameters in the model). However, it turned out that Geisler et al. could not use this optimal strategy because they found that the magnitude of \(\beta\) varied as they varied the area of analysis in the image. In other words, the authors could not find a unique—an ideal—magnitude of \(\beta\). Instead, Geisler et al. decided to leave \(\beta\) as a (single) free parameter in their model, just as the observer criterion is a free parameter in modeling human image data obtained in a detection experiment. By fitting the single free-parameter model to human data, Geisler et al. found that the best results are achieved with \(\beta = 0.38\); the Bayesian local grouping function shown in Figure 4.19F was constructed using that best-fitting magnitude of \(\beta\). Thus, the local grouping function was not truly ideal.

**Absolute Local Grouping Function**

Because absolute EC statistics do not convey information about belongingness of edge elements to contours, Geisler et al. (2001) had to introduce a second parameter, in addition to binding criterion \(\beta\), in order to derive a local grouping function from the absolute EC statistics. This new parameter, called tolerance, determined how sharply the probabilities of element grouping fell off around the most likely parameters of EC shown in Figure 4.19C. For example, low tolerance implies that grouping occurs only when the parameters are close to the most common values evident in the absolute EC statistics. Different values of tolerance result in different absolute local grouping functions; one is shown in Figure 4.19E. When fitting the predictions of the two-parameter absolute local grouping functions to human data, Geisler et al. were able to obtain almost as good a correlation between the predicted and the observed accuracies \((r = .87)\) as they...
obtained in the one-parameter Bayesian local grouping function \( r = .89 \).

In the conclusion of this section, we wish to emphasize the profound theoretical repercussions of the kind of analysis undertaken by Geisler et al. (2001). These authors looked for a foundation of the principles of perceptual organization in the statistics of the natural world and discovered a covariational structure in these statistics. Furthermore, Geisler et al. showed that under minimal assumptions, the regularities of environment can predict human performance in simple perceptual tasks. The findings of Geisler et al. imply that optical stimulation does contain information for perception, in contrast to the view held by the Gestaltists. The information is available for perceptual systems to develop the corresponding sensitivities and to match the perceptual capacities of the organism to the structure of the environment.

As Geisler et al. (2001) pointed out, the rapidly growing research in neural networks shows that self-organizing networks (such as in Kohonen, 1997) are sensitive to the covariational structure of their inputs. This suggests that self-organizing neural networks could provide a useful tool in guiding our search for the match between the perceptual capacities of the biological organisms and the statistical structure of their environments.

REFERENCES


HEARING AS SOUND SOURCE DETERMINATION 121

SOURCES OF SOUND: THE PHYSICS OF THE COMPLEX SOUND WAVE 122
  Simple Vibrations 122
  Complex Vibrations 122
  Sound Propagation 123
  The Decibel 124
  Reflections, Standing Waves, Reverberation, and Sound Shadows 124

AUDITORY ANATOMY AND PHYSIOLOGY 125
  The Peripheral Auditory System: Transduction and Coding 125
  Auditory Nerve 128
  Central Auditory Nervous System 130

DETECTION 131
  Thresholds of Hearing 131
  Masking 132
  Temporal Modulation Transfer Functions 134

DISCRIMINATION 134

SOUND LOCALIZATION 134
  Sound Localization in Three-Dimensional Space 134
  Models of Sound Localization 135
  Lateralization Versus Localization 135
  The Effects of Precedence 136

SOUND SOURCE SEGREGATION 136
  Spectral Separation 136
  Temporal Separation 136
  Spatial Separation 137
  Pitch and Timbre: Harmonicity and Temporal Regularity 138
  Spectral Profile 139
  Common Onsets and Offsets 140
  Common Modulation 140
  Models or Theories of Sound Source Segregation 141

AN OVERVIEW OF THE FUTURE STUDY OF AUDITION 142

REFERENCES 144

HEARING AS SOUND SOURCE DETERMINATION

Hearing allows an organism to use sound to detect, discriminate, and segregate objects in its surrounding world (de Cheveigne, 2001). A simple nervous system could allow a primitive animal to detect the presence of the sound produced by prey on one side of the animal and to use a motor system, like a fin, on the opposite side of the animal to propel it toward the prey. Such a simple auditory detector would not be adaptive if the sound were from a predator. In this case, the system needs to be able to discriminate prey from predator and to activate a different response system (i.e., a fin on the same side of the body) to escape the predator. If the world consisted of either prey or predator, but not both, this primitive animal might survive. In the real world, however, prey and predator commingle. In the real world, the auditory system requires greater complexity in order to segregate prey from predator and then to make an appropriate neural decision to activate the proper response.

Sounds in the world do not travel from their sources to an animal along independent paths; rather, they are mixed into one complex sound wave before reaching the ears of an animal. As we will learn, the peripheral auditory system codes the spectral-temporal attributes of this complex sound wave. The rest of the auditory nervous system must interpret this code in order to reveal information about the sources of the complex sound wave in order that detection, discrimination, and especially segregation can occur (Yost, 1992a). As Bregman (1990) describes, the complex sound wave produces an auditory scene in which the images of this scene are the sound producing sources. Auditory scene analysis is based on perceptual mechanisms that process the spectral-temporal neural code laid down by the inner ear and auditory nerve.

Hearing therefore involves sound, neural structures that code for sound, and perceptual mechanisms that process this neural code. Then this information is integrated with that from other sensory systems and experiences to form a complete auditory system. This chapter begins with a discussion
of sound; follows with a description of the anatomy and physiology of the auditory system, especially the auditory periphery; and concludes with a discussion of auditory detection, discrimination, and segregation.

**SOURCES OF SOUND: THE PHYSICS OF THE COMPLEX SOUND WAVE**

**Simple Vibrations**

An object that vibrates can produce sound if it and the medium through which sound travels have mass and the property of inertia. A simple mass-and-spring model can be used to describe such a vibrating system, with the spring representing the property of inertia. When the mass that is attached to the spring is moved from its starting position and let go, the mass will oscillate back and forth. A simple sinusoidal function describes the vibratory oscillation of the mass after it is set into motion: \( D(t) = \sin(\sqrt{s/m} \cdot t + \theta) \), where \( D(t) \) is the displacement of the mass as a function of time \( t \), \( m \) is a measure of mass, and \( s \) a measure of the spring forces. In general, such a sinusoidal vibration is described by \( D(t) = A \sin(2\pi ft + \theta) \), where \( f \) is frequency \( (f = \sqrt{s/m}) \) and \( A \) is peak amplitude. Thus, a sinusoidal vibration has three mutually independent parameters: frequency \( (f) \), amplitude \( (A) \), and starting phase \( (\theta) \). Figure 5.1 shows two cycles of a sinusoidal relationship between displacement and time. Frequency and amplitude (also level and intensity) are the physical parameters of a vibration and sound. Pitch and loudness are the subjective and perceptual correlates of frequency and amplitude, and it is often important to keep the physical descriptions separated from the subjective. Pitch and loudness are discussed later in this chapter.

In addition to describing the vibration of the simple mass-and-spring model of a vibrating object, sinusoidal vibrations are the basic building blocks of any vibratory pattern that can produce sound. That is, any vibration may be defined as the simple sum of sinusoidal vibrations. This fact is often referred to as the Fourier sum or integral after Joseph Fourier, the nineteenth-century French chemist who formulated this relationship. Thus, it is not surprising that sinusoidal vibrations are the basis of most of what is known about sound and hearing (Hartmann, 1998).

Frequency is the number of cycles competed in one second and is measured in hertz (Hz), in which \( n \) cycles per second is \( n \) Hz. Amplitude is a measure of displacement, with \( A \) referring to peak displacement. Starting phase describes the relative starting value of the sine wave and is measured in degrees. When a sinusoid completes one cycle, it has gone through \( 360^\circ \) \( (2\pi \text{ radians}) \) of angular velocity, and a sinusoid that starts at time zero with an amplitude of zero has a zero-degree starting phase \( (\theta = 0^\circ) \). The period \( (Pr) \) of a sine wave is the time it takes to complete one cycle, such that period and frequency are reciprocally related \( [F = 1/Pr, Pr \text{ in seconds (sec)}, \text{or } F = 1000/Pr, Pr \text{ in milliseconds (msec)}] \). Thus, in Figure 5.1, frequency \( (f) \) is 500 Hz \( (Pr = 2 \text{ msec}) \), peak amplitude \( (A) \) is 10, and starting phase \( (\theta) \) is \( 0^\circ \).

**Complex Vibrations**

Almost all objects vibrate in a complex, nonsinusoidal manner. According to Fourier analysis, however, such complex vibrations can be described as the sum of sinusoidal vibrations for periodic complex vibrations:

\[
D(t) = \sum_{n=1}^{\infty} a_n \sin(2\pi nf, t) + b_n \cos(2\pi nf, t),
\]

where \( a_n \) and \( b_n \) are constants and \( \sin \) and \( \cos \) are sinusoidal functions.

Or as the complex integral for any complex vibration:

\[
f(t) = (1/2\pi) f(w)e^{iwt} \, dt,
\]

where \( w = 2\pi f, f(t) \) is a function of time, and \( f(w) \) is a function of frequency.

Any complex vibration can be described in either the time or the frequency domain. The time domain description provides the functional relationship between the amplitude of vibration and time. The frequency domain description contains the amplitude and phase spectra of the vibration. The amplitude spectrum relates the amplitude of each frequency component of
the complex vibration to its frequency. The phase spectrum provides the starting phases of each frequency component. That is, a complex vibration is the sum of sinusoidal vibrations. The amplitude spectrum describes the amplitudes of each sinusoid and the phase spectrum the starting phase of each sinusoidal component. When the instantaneous amplitudes of each sinusoidal component of the complex vibration are added point for point in time, the time domain description is determined. The time domain and the frequency domain descriptions of complex vibrations are transforms of each other, with each completely describing the vibration. Simple vibrations are sinusoidal vibrations and complex vibrations are the sum of simple or sinusoidal vibrations.

Several different complex signals are described in this chapter. Transient (click) signals are brief (usually less then 1 msec) signals that come on suddenly, stay on at a fixed level, and then go off suddenly. Transients have very broad amplitude spectra, with most of the spectral energy lying in the spectral region less than 1/T, where T is the duration of the transient expressed in seconds (thus, 1/T has the units of frequency). Noise stimuli have randomly varying instantaneous amplitudes and contain all frequencies (within a certain range). If the instantaneous amplitudes vary according to the normal (Gaussian) distribution, the noise is Gaussian noise. If the average level of each frequency component in the noise is the same, the noise is white noise. Noises can be generated (filtered) to be narrow band, such that a narrow-band noise contains frequency components in a limited frequency range (the bandwidth of the noise). The amplitudes or frequencies of a signal can vary as a function of time. For instance, a sinusoidal signal can have its amplitude modulated: A(t) sin(2πft); or it can have its frequency modulated: A sin(2πF(t) t), where A(t) is the amplitude-modulation pattern and F(t) is the frequency-modulation pattern. In general, any signal [x(t)] can be amplitude modulated: A(t)x(t). In this case, A(t) is often referred to as the signal envelope and x(t) as the signal fine structure. Such amplitude- and frequency-modulated sounds are common in nature.

**Sound Propagation**

Objects vibrate and the effects of this vibration travel through the medium (e.g., air) as a sound wave that eventually reaches the ears of a listener. Air consists of molecules in constant random motion. When an object vibrates in air, it causes the air molecules to move in the direction of the vibrating object’s outward and inward movements. An outward motion causes the air molecules to propagate from the source and to condense into areas of condensation where the density of molecules is greater than the average density of air molecules in the object’s surrounding environment. Thus, at a condensation, the air pressure is greater than the average static air pressure, because pressure is proportional to the density of molecules. When the object moves inward, rarefaction areas of lower density are produced, generating lower pressure. These areas of condensation and rarefaction propagate away from the source in a spherical manner as the object continues to vibrate. Figure 5.2 is a schematic depiction of these areas of condensation and rarefaction at one instant in time. Eventually, the pressure wave of alternating areas of condensations and rarefactions cause the eardrum ( tympanic membrane) to vibrate, and the process of hearing begins.

The distance between successive condensations (or successive rarefactions) is the wavelength (λ) of sound. Wavelength is proportional to the speed of sound in the medium (c) and inversely proportional to frequency (f): λ = c/f. The pressure of the sound wave decreases as a function of the square of the distance from the source, and this relationship is called the inverse square law of sound propagation.

![Diagram of sound propagation](image)

**Figure 5.2** Diagram of what one might see if air molecules were photographed as a sound source vibrated. The rarefaction and condensation are shown, as well as the direction (grey arrows above the source) in which the molecules were moving at the instant the picture was taken. The wave moves out in circular manner (actually as a sphere in the three-dimensional real world). As the wave moves out from the source it occupies a greater area, and thus the density of molecules at rarefactions and condensations lessens. The area around the border of the figure represents the static air motion before the propagated wave reaches this area. **Source:** Adapted from Yost (2000).
Sound intensity \( (I) \) is proportional to pressure \( (p) \) squared:
\[
I = \frac{p^2}{\rho c},
\]
where \( \rho \) is the density of the medium in which sound travels (e.g., air). Sound intensity is a power \( (P) \) measure of the rate at which work can be done and energy \( (E) \) is the measure of the amount of work: \( I = P = E/T \), where \( T \) is time.

**The Decibel**

In many situations involving sound, including hearing, the range of measurable sound intensity is very large. The range of sound intensity from the softest sound that one can detect to the loudest sound one can tolerate (the dynamic range of hearing) is on the order of \( 10^{13} \). This large range led to the decibel measure of sound intensity in which the decibel \( (\text{dB}) \) is 10 times the logarithm of the ratio of two sound intensities:
\[
\text{dB} = 10 \log_{10} \left( \frac{I}{I_0} \right),
\]
where \( I_0 \) is a referent sound intensity. Because sound intensity is proportional to pressure squared, \( \text{dB} = 20 \log_{10} \left( \frac{p}{p_o} \right) \), where \( p_o \) is a referent pressure. Thus, the dynamic range of hearing is approximately 130 dB.

The decibel is a relative measure of sound intensity or pressure. Several conventions have been adopted for the referent sound intensity \( (I_o) \) or pressure \( (p_o) \). The most common is the decibel measured in sound pressure level \( (\text{SPL}) \). In this case, \( p_o \) is 20 micropascals \( (20 \mu \text{Pa}) \). This is approximately the sound pressure required for the average young adult to just detect the presence of a tone (a sound produced by a sinusoidal vibration) whose frequency is in the region of 1000 to 4000 Hz. Thus, a measure such as 80 dB SPL means that the sound pressure being measured is 80 dB greater (or 10,000 times greater, \( 20 \log_{10} 10,000 = 80 \text{ dB} \)) than the threshold of hearing (i.e., 80 dB greater than 20 \( \mu \text{Pa} \)). Most often, decibels are expressed as dB SPL, but many other conventions are also used.

**Reflections, Standing Waves, Reverberation, and Sound Shadows**

As a sound wave travels from its source toward the ears of a listener, it will most likely encounter obstacles, including the head and body of the listener. Sound can be absorbed in, reflected from, diffracted around, or transmitted to the medium of the obstacle that the sound wave encountered. Each obstacle offers an impedance to the transmission of the sound wave to the medium of the obstacle. Impedance has three main components. The medium can offer a resistance \( (R) \) to the transmission of sound. The mass of the medium can offer a mass reactance \( (X_m) \) that impedes the sound, and the springlike inertia properties of the medium also produce spring reactance \( (X_s) \). The impedance \( (Z) \) of the medium equals \( \sqrt{R^2 + (X_m - X_s)^2} \). Thus, each obstacle has a characteristic impedance, and the greater the difference in characteristic impedance between two objects, the more sound is reflected from and not transmitted to the new medium. The characteristic impedance of an object is proportional to \( \rho c \), which is the denominator of the definition of sound intensity \( (I = \frac{p^2}{\rho c}) \). Thus, sound intensity is equal to pressure squared divided by characteristic impedance.

When sound is reflected from an object, the reflected sound wave can interact with the original sound wave, causing regions in which the two sound waves reinforce each other or at other locations cancel each other. Under the proper conditions, the reflected reinforcements and cancellations can establish a standing wave. A standing wave represents spatial locations in which the pressure is high (antinodes) due to reinforcements and spatial locations where the pressure is low nodes due to cancellations. The wavelength of a standing wave (distance between adjacent nodes or antinodes) is determined by the size of the environment in which the standing wave exists. Large areas produce long standing-wave wavelengths and hence low frequencies, and the converse is true for small areas. Thus, a standing wave in a short tube will produce a high-frequency standing wave, and a long tube will produce a low-frequency standing wave. This is the principal upon which organ pipes and horns operate to produce musical notes. Structures in the auditory system, such as the outer ear canal, can also produce standing waves.

The reflections from many surfaces can reinforce each other and sustain sound in an environment long after the sound has terminated. The time it takes this reverberation to decline by 60 dB relative to the source level is the reverberation time of the environment. Rooms can support high speech intelligibility and pleasant listening if there is some reverberation, but not if the reverberation time is too long.

If the size of an object is large relative to a sound’s wavelength, most of the sound will either be reflected from the object or be transmitted to the object. Sound will be diffracted around (bypass) an object whose size is much smaller than the sound’s wavelength. When the wavelength of sound is approximately the same as the size of an object, some of the sound is reflected from the object and some is diffracted around the object. The result is that there is an area on the side of the object opposite from where the sound originated where the sound pressure is lower. Thus, such an object produces a sound shadow in an area very near the object, where there is a lower sound pressure than there is in areas farther away from the object. The head, for instance, produces a sound shadow at the far ear when the frequency of sound arriving at the lead ear is generated by a sound with a wavelength that is approximately equal to or smaller than the size of the head.
AUDITORY ANATOMY AND PHYSIOLOGY

The auditory system (see Figure 5.3) has four main parts: The outer ear collects and funnels sound to the middle ear, which increases the force produced by air moving the tympanic membrane (eardrum) so that the fluid and tissues of the inner ear are efficiently vibrated; this enables the inner ear to transduce vibration into a neural code for sound, which the central auditory nervous system can process and integrate with other sensory and experiential information in order to provide motor, behavioral, and other outputs.

The Peripheral Auditory System: Transduction and Coding

Outer Ear

As sound travels from the source across the body and head, especially the pinna (see Figure 5.3), various body parts attenuate and delay the sound in a frequency-specific way caused by properties of reflection and diffraction. Thus, sound arriving at the outer ear canal is spectrally different from that leaving the source. These spectral alterations are described by head-related transfer functions (HRTFs), which specify the spectral (amplitude and phase) changes produced by the body and head for sources located at different points in space. The HRTFs may provide cues that are useful for sound localization (Wightman & Kistler, 1989a). Within the outer ear canal, resonances can be established that boost sound pressure in spectral regions near the 3000- to 5000-Hz resonant frequency of the outer ear canal (Shaw, 1974).

Middle Ear

The major function of the middle ear is to provide an increase in vibratory force so that the fluids and tissues of the inner ear can be effectively moved (Geisler, 1998; Pickles, 1988). The impedance of the inner ear structures is about 40 times greater...
than that of air (a 32-dB change). The middle ear compensates for this impedance difference via the lever action of the ossicular chain (a chain of three bones—malleus, incus, stapes—connecting the tympanic membrane to the inner ear) in combination with the pressure increase between the large area of the malleus’s connection to the tympanic membrane and the small area of the footplate of the stapes’ connection to the oval window of the inner ear. Over a significant portion of the audible frequency range, the middle ear in combination with the resonances of the outer ear canal delivers the sound to the inner ear with no pressure loss due to the high impedance of the inner ear structures. The eustachian tube connects the middle ear to the nasal cavities so that pressure on each side of the tympanic membrane remains the same, a necessary condition for efficient middle- and inner-ear functioning.

**Inner Ear**

The inner ear contains the hearing organs and those of the vestibular (balance-equilibrium) system (Fay & Popper, 1992; Webster, Fay, & Popper, 1992). The anatomy of the inner ear differs significantly across the animal kingdom. In mammals, the cochlea, a snail-like tube that spirals on itself three to four times, is the hearing organ of the inner ear (see Figure 5.3). The cochlea contains an inner tube, the cochlear partition, which contains supporting structures and the hair cells, the biological transducers for hearing. The cochlea is thus divided into three canals or scala: scala vestibuli (above the cochlear partition), scala media (the cochlear partition), and scala tympani (below the cochlear partition). Scala vestibuli and scala tympani contain a viscous fluid, perilymph, whereas scala media contains a different fluid, endolymph. In a cross-section (see Figure 5.4), the cochlear partition is bounded above by Resiner’s membrane and below by the basilar membrane. The metabolic engine for the cochlea resides within stria vascularis on the outer wall of the cochlea. Fibers from the auditory part of the VIIIth cranial nerve innervate the hair cells along the basilar membrane and course through the middle (modiolus) of the cochlea before picking up myelination on the way to the auditory brain stem. There are two types of hair cells (see Figure 5.5): outer hair cells, which in mammals are arranged in three rows toward the outside of the cochlear

Figure 5.4  Main components of the inner ear in relation to the other structures of the ear. (From Yost, 2000, adapted from drawings by Dorland, 1965, with permission.) Schematic diagram of middle ear and partially uncoiled cochlea, showing the relationship of the various scalae.  Source: From Yost (2000), adapted from similar drawings from Zemlin (1981), with permission.
partition; and inner hair cells, which are aligned in a single row. Several different supporting structures buttress the hair cells on the basilar membrane.

The vibration of the stapes causes the oval window to vibrate the fluids of the cochlea (Dallos, Popper, & Fay, 1996). This vibration sets up a pressure differential across the cochlear partition, causing the cochlear partition to vibrate. This vibration causes a shearing action between the basilar membrane upon which the hair cells set, and the tectorial membrane, which makes contact with the stereocilia (the hairs, so to speak, that protrude from the top of the hair cells; see Figure 5.5) such that the stereocilia are bent. The shearing of the stereocilia opens transduction channels, presumably toward the tips of the stereocilia, which initiates a generator potential in the hair cell and a resulting action potential in the auditory nerve fiber that innervates the hair cells (Pickles, 1988). Thus, the mechanical vibration of the stereocilia is transduced into a neural signal.

The properties of the cochlear partition involving its width and tension, as well as the fact that the cochlear partition does not terminate at the end of the cochlea, all result in a particular motion being imparted to the cochlear partition when it is vibrated by the action of the stapes (Dallos et al., 1996). The cochlear partition motion is described as a traveling wave, such that the vibration of the cochlear partition is distributed across the partition in a frequency-specific manner. High-frequency sounds generate maximal displacement toward the base of the partition where the stapes is, and the vibration does not travel very far along the partition. Low-frequency sounds travel along the partition towards its apex (end opposite of the stapes), such that maximal displacement is toward the apical end of the cochlear partition. Figure 5.6 provides a schematic depiction of the traveling wave for three different frequencies. The biomechanical traveling wave, therefore, sorts frequency according to the location of maximal displacement along the cochlear partition: High frequencies cause maximal vibration at the base, low frequencies at the apex, and middle frequencies at intermediate partition locations. Thus, the place of maximal displacement codes for the frequency content of the stimulating sound wave. If a sound wave is the sum of two frequency components, then there will be two locations of maximal displacement; three frequency components would generate a maximum of three, and so forth. The hair cells are distributed along the cochlear partition as if they were sensors of the cochlear displacement. Thus, different hair cells code for the frequency content of the incoming sound.

Figure 5.5  Light micrograph of a cross section of a chinchilla organ of Corti. Clearly shown are: IHC: inner hair cells; OHC: the three rows of outer hair cells. The stereocilla (Sc) of the outer and inner hair cells protrude through the recticular lamina that helps support the hair cells. Other supporting structures are shown. Source: From Yost (2000), photographs courtesy of Dr. Ivan Hunter-Duvar, Hospital for Sick Children, Toronto.

Figure 5.6  Instantaneous patterns and envelopes of traveling waves of three different frequencies shown on a schematic diagram of the cochlea. Note that the point of maximum displacement, as shown by the high point of the envelope, is near the apex for low frequencies and near the base for higher frequencies. Also note that low frequencies stimulate the apical end as well as the basal end, but that displacement from higher frequencies is confined to the base. Source: From Yost (2000), adapted from similar drawings Zemlin (1981), with permission.
Why should the system have two types of hair cells (inner and outer)? More than 90% of the inner hair cells are innervated by afferent auditory nerve fibers, indicating that the inner hair cells are the biological transducers for sound. The outer hair cells appear to perform a very different task (Dallos et al., 1996). The outer hair cells change their size (primarily in length) in reaction to stimulation, and the change in length happens on a cycle-by-cycle basis, even for high frequencies of 20,000 Hz and above. The shearing of the stereocilia of outer hair cells causes a neural action leading to a deformation of the walls of the outer hair cells, resulting in a change in length (Brownell, Bader, Bertrand, & de Ribaupierre, 1985; Geisler, 1998; Pickles, 1988). The length change most likely alters the connections between the basilar and tectorial membranes in a dynamic fashion, which in turn affects the shearing of the inner hair cell stereocilia (Zajic & Schacht, 1991). This type of positive feedback system appears to feed energy back into the cochlea, making the haircell function as an active process. The high sensitivity, fine frequency resolution, and nonlinear properties of the biomechanical action of the cochlear partition depend on viable outer hair cells. Thus, the outer hair cells act like a motor, varying the biomechanical connections within the cochlea that allow for the inner hair cells to transduce vibration into neural signals with high sensitivity and great frequency selectivity (Dallos et al., 1996).

A consequence of the motile outer hair cells may be the otoacoustic emissions that are measurable in the sealed outer ear canal of many animals and humans (Kemp, 1978). If a brief transient is presented to the ear and a recording is made in the closed outer ear canal, an echo to the transient can be recorded. This echo or emission is cochlear in origin. Emissions occur in response to transients (transient-evoked otoacoustic emissions; TOAE), steady-state sounds (usually measured as distortion product otoacoustic emissions; DPOAE), and they can occur spontaneously (spontaneous otoacoustic emissions; SOAE) in the absence of any externally presented sound. Otoacoustic emissions are also dependent on neural efferent influences on the outer hair cells. Presumably, the emissions result either from the spontaneous motion of outer hair cells or from other actions of the active processes associated with outer hair cell motility. These emissions can be used to access the viability of the cochlea and are used as a noninvasive measure of hearing function, especially in infant hearing screening programs (Lonsbury-Martin & Martin, 1990).

**Auditory Nerve**

Each inner hair cell is connected to about ten auditory nerve fibers, which travel in the XIIIth cranial nerve in a topographical organization to the first synapse in the cochlear nucleus of the auditory brain stem (Webster et al., 1992). Thus, the auditory nerve fibers carry information about the activity of the inner hair cells, which are monitoring the biomechanical displacements of the cochlear partition (Fay & Popper, 1992; Geisler, 1998; Pickles, 1988). Figure 5.7 shows tuning curves for individual auditory nerve fibers. A
tuning curve represents the tonal level required for a threshold number of neural discharges as a function of the frequency of the tone. The sharp V-shape tuning curves indicate that auditory nerve fibers are highly tuned to the frequency of stimulation; thus, they reflect the biomechanical traveling wave action of the cochlear partition.

The discharges of auditory nerve fibers are also synchronized to the vibratory pattern of acoustic stimulation. Figure 5.8 shows histograms for auditory nerve fibers indicating that the timing pattern of acoustic stimulation is preserved in these fibers up to about 5000 Hz (Geisler, 1998; Pickles, 1988). The auditory nerve discharges during only one phase of the sound, and the probability of discharge is proportional to the instantaneous amplitude of the sound. Thus, the temporal pattern of neural discharges in auditory nerve fibers depicts the temporal structure of sound’s pressure wave form for those frequency components that are lower in frequency than approximately 5000 Hz.

The number of discharges and number of discharging fibers increases in proportion to stimulus level. However, the discharge rate of individual auditory nerve fibers varies over only about a 40–50 dB range. Thus, although increased discharge rate does provide information about a sound’s overall amplitude, a simple relationship between discharge rate and sound amplitude cannot account for the range of sound intensity that most animals are capable of processing.

Thus, auditory nerve fibers are highly frequency-selective, discharge in synchrony with the acoustic stimulus (at least up to 5000 Hz), and change their discharge rates in proportion to sound level. The discharge rate of any particular auditory

Figure 5.8  Time-locked post-stimulus-time (PST) histograms to the sum of two pure tones. In the top row, tone 2 is 20 dB more intense than tone 1. In the middle row there is a 15 dB difference between the two tones, and in the bottom row the difference is 10 dB. For all cases the time domain waveform of the summed sinusoids is superimposed on top of the PST histogram. The nerve discharges only during one phase of the waveform (the positive-going sections of the waveform). The PST histogram displays the ability of the nerve to discharge in synchrony with the period of the input stimulus envelope, at least for low-frequency stimulation. Source: From Yost (2000), based on a figure from Hind, Anderson, Brugge, and Rose (1967), with permission.
nerve represents the relative amplitude of a particular frequency in the sound. The temporal discharge pattern of the fiber indicates the time domain properties of the sound in this frequency region. The overall level of neural discharge rate indicates the sound’s overall amplitude. Individual auditory nerve fibers are topographically organized within the auditory nerve bundle; fibers carrying low-frequency information are toward the middle of the XIIth bundle, and fibers carrying high-frequency information are toward the outside of the bundle. Thus, a spatial (spectral)-temporal representation of the stimulating sound is transmitted via the auditory nerve to the auditory brain stem. This spatial-temporal pattern represents the neural code for the sound waveform that is the composite of the sounds generated from all of the sources in the acoustic environment.

**Central Auditory Nervous System**

Figure 5.9 depicts a schematic diagram of the gross anatomy of the major components of central auditory nervous system. In addition to the afferent pathways indicated in Figure 5.9, there is a network of efferent centrifugal connections as well (Altschuler, Hoffman, Bobbin, & Clopton, 1989).

The cochlear nucleus has many different fiber types and connections in its three main subdivisions. There is evidence for lateral inhibitory networks in the cochlear nucleus that may aid it in performing different types of spectral pattern processing (Young, 1984). Processing of binaural information occurs in the olivary complex, where the first significant bilateral interactions occur. The medial superior olive is most sensitive to interaural (between the ears) time differences,
and the lateral superior olive is most sensitive to interaural level differences. Interaural time and level differences are important cues for sound localization (Yost & Gourevitch, 1987). The lateral limniscus is primarily a monaural pathway in many animals. The inferior colliculus appears to be a major processing nucleus for spatial hearing, modulation processing, and spectral pattern recognition (Fay & Popper, 1992). In mammalian systems, not a lot is known about the function of the medial geniculate body (Fay & Popper, 1992). The auditory cortex in primates is located deep within the Sylvian fissure, making it difficult to reach for physiological study. The auditory cortex, as are all parts of the central auditory nervous system, is tonotopically organized: Different cortical neurons are selective for different frequencies. There is evidence for modulation processing in the auditory cortex, and the auditory cortex may provide spatial maps for sound localization (Altschuler et al., 1989).

The study of animals with special adaptations for hearing, echo-locating bats (Suga, 1988), and the barn owl (Konishi, Takahashi, Wagner, Sullivan, & Carr, 1988), have provided valuable information about the functional role of the central auditory system. These studies have helped guide the study of the brain stems and cortices of other animals, including humans. The auditory nervous system is an anatomically complex system, perhaps reflecting the amount of neural computation that hearing appears to require.

DETECTION

Thresholds of Hearing

A basic measure of auditory detection is the threshold of hearing for pure tones, or the audiogram. The audiogram can be obtained in two conditions, each requiring its own calibration procedure. In the minimal audible field (MAF) process, listeners detect the presence of pure tones presented from loudspeakers, whereas in the minimal audible pressure (MAP) process, the sounds are presented over headphones. Figure 5.10 shows the thresholds of hearing (the audiogram) for the two procedures. The figure also shows estimates of the upper limit for hearing, indicating those sound levels that either are very uncomfortable or yield the sensation of pain. The thresholds of hearing have been standardized for both the MAP and MAF procedures, and the two estimates differ by on average about 6 dB. However, the differences are accounted for by calculating the diffraction of sound around the head and the resonance properties of the outer ear canal, which are substantially different in the MAP and MAF procedures (Yost & Killion, 1993). Figure 5.10 suggests that young humans can detect sounds from 20 to 20,000 Hz, and the dynamic range of hearing is about 130 dB in the middle of the range of the audible frequencies. At 0 dB SPL the pressure in the outer ear canal is 20 μPa, which indicates that the tympanic membrane at auditory threshold is moving a distance equal to approximately the diameter of a hydrogen atom. Females have slightly lower thresholds of hearing than do males. The thresholds of hearing increase as a function of age in a frequency-dependent manner (presbycusis), such that the thresholds for high-frequency sounds increase at an earlier age than do those for low-frequency sounds. This frequency dependence is consistent with the operation of the traveling wave, in which all sounds excite the base of the cochlear partition, where high-frequencies are coded, but only low-frequencies sounds excite the apex. Thus, the base of the cochlear partition, where high frequencies are coded, is more likely to be fatigued over time than is the apex.

Figure 5.10 shows the threshold levels for tonal detection. The subjectively perceived loudness of sound is also a joint function of the sound’s physical frequency and level. Figure 5.11 shows subjective equal-loudness contours; each contour describes the tonal levels and frequencies that are judged, in a loudness matching procedure, equally loud to a 1,000-Hz tone presented at the constant level indicated by the phon rating of the contour. Thus, all tones that have a loudness of x phons are judged equally loud to a 1,000-Hz tone presented at x dB SPL.

The thresholds of hearing are dependent on the duration of the sound—the shorter the sound, the higher the thresholds. Thresholds for tonal stimuli decrease as duration is increased until the duration is approximately 300 ms; then threshold remains constant as duration is increased further. To a first
approximation and over a considerable range of durations, if the energy of the sound remains constant, detection thresholds also remain constant (equal-energy rule). The duration at which thresholds no longer change with increases in duration (i.e., at 300 ms) is referred to as the integration time for detection (Viemeister & Plack, 1993).

**Masking**

When the threshold for detecting sound A is increased in the presence of another sound, sound B, sound B is said to be a masker for the signal, sound A. The amount of masking is the amount of the increase in signal detection threshold due to the presence of the masker. Figure 5.12 shows the thresholds for detecting a signal tone of one frequency as a function of tonal maskers of different frequencies. For these data, listeners were asked to detect the presence of a short-duration tonal signal presented just a few decibels above its threshold of hearing. The level of the tonal masker that yielded threshold performance for detecting the signal was determined for each masker frequency. Both the similarity of the shape of the data curves in Figure 5.12 to those in Figure 5.7 and the methodological similarities result in these psychophysical data being referred to as psychophysical tuning curves. It is assumed that the frequency selectivity suggested by psychophysical tuning curves results from the frequency selectivity measured in the auditory periphery (Moore, 1997).

The observation from Figure 5.12, that masking is greatest when the frequency of the masker is near that of the signal, was studied extensively by Harvey Fletcher in the 1940s. He formed the concept of the critical band (Fletcher, 1953), stating that only a band of frequencies near that of the signal was critical for masking. He further theorized that the amount of masking of a tonal signal was proportional to the power of the critical masking band. These observations have been confirmed by many experiments since the 1940s.

The tonal psychophysical curves are one method used to measure this critical band. However, several interactions can occur between a tonal signal and a tonal masker that can complicate interpretation of some tone-on-tone masking results (Wegel & Lane, 1924). If the signal and masker frequencies differ by 20 or fewer Hz, then the tones interact to produce slow fluctuations in overall intensity that result in the perception of beats (alteration in loudness), which can be used as a cue for detecting the presence of the tonal signal. In addition, the nonlinear properties of auditory transduction can produce aural distortion products that can also provide a detection cue. The tonal masker can produce aural harmonics, which are frequencies at the harmonics of the masker frequency caused by the nonlinear process. The nonlinear properties of transduction can produce difference tones, which are frequencies equal to differences between the frequencies of the masker and signal. The psychophysical tuning curve method reduces, but does not always eliminate, the effect of many of these stimulus interactions as possible detection cues.

The preferred method for measuring the critical band is the band-reject noise paradigm as shown in Figure 5.13. A band-reject noise has a frequency region filtered out of the noise, which for masking is a frequency region surrounding the signal frequency. This band-reject, noise-masking procedure (Moore, 1986) reduces or eliminates all of the interactive
effects obtained with tonal maskers and signals. As the width of the band-reject noise increases, signal threshold is lowered because there is less power in the critical band of frequencies near that of the signal. The width of the critical band is proportional to signal frequency as is consistent with frequency tuning measured in the auditory periphery (Glasberg & Moore, 1990). That is, as the frequency content of a signal increases, the range of frequencies that are critical for masking the signal also increases proportionally.

The concept of the critical band as a measure of auditory-processing channels that are frequency tuned is closely tied to the biomechanical and neural measures of processing in the auditory periphery. This combination of physiological and psychophysical evidence for frequency-tuned channels forms a significant part of all current theories and models of auditory processing (Moore & Patterson, 1986).

Many data from masking experiments, especially those involving Gaussian noises and tonal signals, can be explained using the energy detection model (Green & Swets, 1973) from the general theory of signal detection (TSD). For instance, in an experiment in which a Gaussian noise masks a tonal signal, the energy detection model assumes that the energy of the noise is compared to that of the signal plus noise. The noise masker energy is a random variable that can be described with a distribution with a known mean and standard deviation. The addition of the signal to the noise often increases the mean of the distribution, but not the standard deviation. As signal level increases, the normalized (normalized by the common standard deviation) difference in the means of the distributions increases. On any presentation, listeners use a sample of energy to decide whether the signal plus noise or just noise was presented. If the likelihood of the sampled energy is greater than a criterion value (set by the listener’s response proclivity or bias), the listener responds that the signal plus noise was presented, because high signal levels are more likely to produce high energy levels. A measure of performance ($d'$) can be obtained from the theoretical distributions of signal-plus-noise and noise-alone conditions, and then can be compared to a similar $d'$ measure obtained from the listener’s data. Various forms of the energy model and other models based on TSD have been successful in accounting for a variety of masking results (Green & Swets, 1973).

**Temporal Masking**

The masking data described so far are based on conditions in which the signal and masker occur at the same time. Masking also takes place when the signal and maskers do not temporally overlap. Forward masking occurs when the signal comes on after the masker is turned off and backward masking occurs when the signal precedes the masker. For the same temporal separation between signal and masker, there is usually more forward than backward masking. In the fringe conditions, a short-duration signal is presented near the onset (forward fringe) or offset (backward fringe) of a longer-duration masker. Most often, the greatest amount of masking occurs in these fringe conditions (masking overshoot).

As has already been described, the nonlinear properties of auditory transduction can have several psychophysical consequences. The existence of aural harmonics and difference tones is one such consequence. It is also probably the case that there are suppressive effects that are a function of some form of nonlinearity. That is, the masker may suppress or inhibit the excitatory effect of the signal under different conditions. The separation of the signal and masker in temporal masking conditions allows one to potentially isolate these suppressive effects. The fact that psychophysical tuning measured in forward masking generates measures of narrower tuning (smaller critical bands) than that obtained in simultaneous masking may be consistent with such suppressive effects existing in the simultaneous conditions (Moore, 1986).

Nonlinear peripheral processing is a compressive nonlinearity in which neural output is compressively related to sound.
input. Thus, the same decibel change in sound level produces a smaller change in neural output at high-sound levels than at low-sound levels. This compressive nonlinearity may also be the cause of a difference between simultaneous and forward tonal masking. In simultaneous tonal masking, the signal must change by about 1 dB for each 1 dB change in masker level in order for constant signal detection to occur. In forward masking, a change of less than 1 dB for each decibel change in masker level is required for constant detection. This change in masking slopes between simultaneous and forward masking may result because in simultaneous masking, both the signal and masker undergo the same form of compression. In forward masking, the temporal separation between the masker and signal results in the lower-level signal undergoing a different form of compression than that for the higher-level masker (Moore, 1995).

Temporal Modulation Transfer Functions

Most sounds change in their overall level over time (these sounds are amplitude modulated). The temporal modulation transfer function is one measure of the auditory system’s ability to detect such level changes. A noise waveform is amplitude modulated such that its overall amplitude varies from a low to a high level in a sinusoidal manner. Listeners are asked to detect whether such dynamic amplitude modulation occurs. The depth of modulation (the difference between the peak and valley levels) required for modulation detection (i.e., the ability to detect a difference between a noise with no modulation and a noise sinusoidally amplitude modulated) is determined as a function of the rate at which the amplitudes are modulated. As the modulation rate increases, the depth of modulation must increase to maintain a threshold ability to detect modulation. That is, at low rates of modulation, only a small depth of modulation is required to detect amplitude modulation. As the rate of modulation increases, the depth of modulation required for modulation detection also increases in a monotonic manner. The function relating threshold depth of modulation to the rate of modulation resembles that of a lowpass filter. The lowpass form of this function describes the temporal modulation transfer function for processing temporal amplitude changes (Dau, Kollmeier, & Kohlrausch, 1997; Viemeister & Plack, 1993).

DISCRIMINATION

Measures of the ability of listeners to discern differences in frequency, level, and the timing properties of sounds is often tied to the nineteenth-century observations of Weber and Fechner. The Weber fraction states that the just-noticeable difference between two stimuli is a fixed proportion of the value of the stimuli being judged. The Weber fraction for frequency, level, and duration have been measured for a variety of acoustic signals.

For sound level, listeners can detect between a 0.5- and 1.5-dB level difference (Jesteadt, Weir, & Green, 1977). For tonal stimuli, the Weber fraction is somewhat dependent on overall level, leading to a near miss to the Weber fraction. The Weber fraction for noise stimuli is constant at about 0.5 dB as a function of overall level, such that there is not a near-miss to Weber’s fraction for noise signals. The just-noticeable difference for tonal frequency is about 0.2–0.4% of the base frequency; for example, trained listeners can just discriminate a 1002-Hz tone from a 1000-Hz tone (Weir, Jesteadt, & Green, 1977). There is not a constant Weber fraction for most measures of temporal discrimination. Changes in duration can affect the detectability and loudness of sound, making it difficult to obtain unconfounded measures of duration discrimination (Abel, 1971; Viemeister & Plack, 1993).

SOUND LOCALIZATION

Sound Localization in Three-Dimensional Space

Sound has the properties of level, frequency, and time, but not space. Yet, the sound produced by an object can be used by most animals to locate that object in three-dimensional space (Blauert, 1997; Gilkey & Andersen, 1997). A different set of acoustic cues is used for sound localization in each plane. The location of a sound source is determined by neural computations based on these cues.

In the horizontal or azimuth plane, left-right judgments of sound location are most likely based on interaural differences of time and level (Wightman & Kistler, 1993; Yost & Gourevitch, 1987). The sound from a source will reach one ear (near ear) before it reaches the other ear (far ear), and the interaural difference in arrival time (or a subsequent interaural phase difference) is a cue for sound localization. However, given the small maximal interaural time difference due to the size of the head, this cue is probably only useful for low-frequency sounds. The sound level at the near ear will be greater than that at the far ear, primarily because the head produces a sound shadow at the far ear. The sound shadow is proportional to frequency, so that interaural level differences most likely provide cues for sound localization at high frequencies. The fact that interaural time provides a cue for sound location at low frequencies and interaural level differences a cue at high frequencies is referred to as the duplex theory of sound localization (Yost & Gourvitvch, 1987).
Sound localization accuracy is best at frequencies below 1000 Hz (the region where interaural time differences are useful cues) and above 2000 Hz (the region where interaural level differences are useful cues), and the transition region around 1500 Hz is consistent with the duplex theory of sound localization. Sound localization acuity is best for azimuthal judgements. Differences as small as 1° of visual angle can be discriminated when the sound sources are directly in front (discriminations of differences in sound source locations are referred to as minimal audible angles; see Mills, 1972). Differences in interaural time differences as small as 10 microseconds and differences in interaural level differences as small as 0.5 dB can be discriminated (Blauret, 1997).

All sounds that lie on cones of confusion (Mills, 1972) generate the same interaural time and level differences. One such cone is the midsagittal plane: the plane that is from directly in front, to directly overhead, to directly behind, to directly below a listener. All locations on the midsagittal plane produce zero differences of interaural time and level, and as such these interaural differences would not allow sound location within this plane. Yet, listeners can accurately locate sound sources in this plane without moving their heads (head movements would change the cone of confusion). Thus, cues other than the interaural differences are most likely used to locate sounds in the vertical plane (in the up-down direction).

The head-related transfer functions (HRTFs) discussed in relationship to the outer ear describe the spectral changes that sound undergoes as it travels from its source across the body and head of the listener toward the middle ear. The spectral characteristics of the HRTF are dependent on the location of the sound source. In particular, there are spectral peaks and valleys in frequency regions above 4000 Hz that change spectral location in a systematic and orderly manner as a function of the vertical position of the sound source (Wightman & Kistler, 1989b). Thus, the frequency location of these HRTF spectral peaks and valleys are probable cues for sound localization in the vertical direction. For instance, vertical sound localization is degraded if sound is low-passed filtered so that there is little or no energy above 4000 Hz where the spectral peaks and valleys are located. Acuity in the vertical direction is generally poorer than in the horizontal direction (Middlebrooks, 1992). The greatest number of sound localization errors occur along cones of confusions. For instance, there can be significant front-back and back-front confusions in the midsagittal plane (Wightman & Kistler, 1989b).

Sound localization accuracy of the distance of a sound source is poorer than either horizontal or vertical sound localization accuracy. The primary cues for distance perception are either the relative sound level or the ratio of reverberant to direct sound impinging on the listener (Loomis, Klatzky, Philbeck, & Golledge, 1998). If the sound source is within the near field of a listener (within about one meter), then interaural level differences may aid in distance judgements. Relative sound level is only a useful cue if there is some a priori knowledge of the level, because sound level can vary at the source, as well as a function of distance. The ratio of reflected or reverberant sound to that coming directly from the source varies as a function of distance, making this ratio a probable cue for distance judgements when there are reflective surfaces (e.g., the ground).

Models of Sound Localization

Neural coincidence networks have been suggested as one means by which the auditory system might compute interaural differences, especially interaural time differences (Colburn & Durlach, 1978). The network contains cells receiving bilateral inputs. The cells fire upon the simultaneous arrival of neural information from the two inputs. The network can serve as a neural crosscorrelator of the timing information arriving at each ear. Thus, sound arriving at the same time at each ear activates neurons in the middle of the network, whereas sound arriving at one ear ahead of that arriving at the other ear activates neurons to one side of the network. The activation of these neurons in the coincidence network could form a type of spatial neural map. Several computational models based on coincidence and crosscorrelation have been successful in accounting for a great deal of data based on manipulations of interaural time differences.

There is neural evidence in some species, especially birds, for just this sort of coincidence network (Konishi et al., 1988).

Lateralization Versus Localization

When sounds are presented over headphones and interaural differences of time or level are varied, listeners report that the sounds move left and right as they do in the real world. However, the sounds are lateralized inside the head rather than in space as any real-world sound source would be located—and therefore localized. Thus, lateralization is often used to refer to headphone-delivered sounds in the study of sound localization and localization when sound sources are in the external world (Blauert, 1997; Yost & Gouervitch, 1987).

One reason that sound delivered over headphones may be lateralized rather than localized is that the headphone-delivered sounds have not undergone the spectral transformations associated with the HRTFs that naturally occurring sounds undergo. If the information about the HRTF is put back into the sound delivered over headphones, then it is possible to produce a sound over headphones (using HRTF filters) that is spectrally identical to that which would have arrived at the middle ear from a real sound source at some
location in space. When such HRTF filters are used, listeners are much more likely to localize the sounds in space at a location appropriate for the specific HRTF used than they are to lateralize the sound inside the head. Thus, HRTF-filtered sound presentations over headphones can create a virtual auditory environment simulating sound localization in the real world. Under the proper conditions, sounds delivered over headphones are perceived as nearly indistinguishable from the sound delivered from actual sources (Gilkey & Andersen, 1997; Wightman & Kisltler, 1989b).

The Effects of Precedence

Although reflections from surfaces may aid distance judgements, they could also offer a confusing auditory scene for sound localization, because each reflection could be misinterpreted as a possible sound source location. In most real-world spaces, reflections do not have a significant effect on either the location or on the fidelity of the sound from the originating source. The sound from the source will reach a listener before that of any reflection due to the longer path any reflection must travel. Hence, it is as if the sound from the source takes perceptual precedence over that from reflections (Litovsky, Colburn, Yost, & Guzman, 1999).

The effects of precedence (Litovsky et al., 1999) include that fact that the reflections are rarely perceived as separate echoes (fusion), the perceived location of a sound source in a reflective environment is dominated by the location of the source and not by the location of reflections (location dominance), and information about reflections is suppressed relative to that about the source (discrimination suppression). Evidence also suggests that the effects of precedence may be influenced by a listener’s prior listening experience in an acoustic environment. A common paradigm (Litovsky et al., 1999) for studying the effects of precedence involves the presentation of a transient from one loudspeaker (the lead or source sound), followed a few milliseconds later by an identical transient presented from a different loudspeaker (the lag or reflected sound). In most cases in this lead-lag paradigm, a single transient is perceived (fusion), at the location of the lead loudspeaker (localization dominance), and the spatial acuity of the lag is reduced relative to conditions when the lag was presented in isolation of the lead (discrimination suppression).

SOUND SOURCE SEGREGATION

Any animal’s auditory experience probably involves processing several simultaneously or nearly simultaneously occurring sound sources. Several stimulus cues have been suggested as possibilities for segregating sound sources in the complex acoustic world: spectral separation, temporal separation, spatial separation, pitch and timbre (harmonicity and temporal regularity), spectral profiles, common onsets and offsets, and common modulation (Yost & Sheft, 1993; Yost, Popper, & Faye, 1993).

Recall from the description of the auditory periphery that the auditory nerve codes for the spectral-temporal properties of sound. Sounds from every sound source in an acoustic environment are combined into a single complex sound field that stimulates the inner ear. The auditory periphery codes for the spectral-temporal structure of this complex sound field. The spectral-temporal code must be analyzed to determine the potential sound sources. That is, the spectral-temporal neural properties must be deconvolved into subsets of spectral-temporal patterns representing the sound originating from each individual sound source. This form of analysis is presumably performed by the central auditory nervous system. Note that in order for this type of analysis to take place, computations must be made across frequency and over time (Bregman, 1990).

Spectral Separation

If two sound sources had very different and nonoverlapping spectral structures, the frequency-resolving ability of the auditory system might segregate the two sound sources very nicely into two patterns. Thus, in some cases the frequency-resolving abilities of the auditory system can aid in sound source segregation, but not in all cases. The difficulty arises when the spectra of sounds from different sources overlap in frequency and time.

Temporal Separation

Clearly, if the sound from two sources occurs at different times, and there is little, if any, temporal masking, then sound source segregation is possible. In many real-world situations, the sound from one source is intermittent and may overlap in time with sounds from other sources that are also intermittent. In addition to the question of segregation of one sound source from other sound sources, this stimulus situation also addresses the question of how an intermittent sound from a source continues to be identified as originating from that source, especially if other sounds occur at or about the same time. A series of studies referred to as auditory stream analysis investigates this type of stimulus condition (Bregman, 1990).

An early context for the study of auditory stream processing involved the presentation of two tones of different
frequencies, and each tone is pulsed on and off such that when one tone is on, the other is off. Two different percepts may occur in this situation. In one case, the perception is of one sound source that is alternating in pitch. In the other case, the perception is of two sound sources, each associated with the individual frequencies and with each containing a pulsing sound. In the latter case, it as if there were two sound sources, each producing a pulsing tone occurring at the same time, like two streams running side by side (stream segregation). By determining the stimulus conditions that yield stream segregation, investigators have attempted to study those stimulus conditions that promote sound source segregation. In addition to an appropriate frequency separation between the two sounds, differences in stimulus complexity, interaural differences (i.e., spatial differences), temporal amplitude modulation differences, and level differences may promote stream segregation. The temporal structure of the stimulus context plays a crucial role in stream segregation. In general, spectral differences promote stream segregation more than other stimulus attributes do.

In another methodology involving temporal sequences of sounds, an auditory pattern of tonal sounds is generated as a model of complex sounds, such as speech (Watson, 1976). In many conditions, a pattern of 10 tones presented in succession, each with a different frequency, is used as a tonal pattern. The frequency range over which the tones vary, the duration of each tone, and the overall duration of the 10-tone pattern are often similar to that occurring for many speech sounds, like words (see the Fowler chapter of this volume). Listeners are asked to discriminate a change in the frequency of 1 of the 10 tones in the pattern. In many conditions, the patterns change from trial to trial in a random manner. In this case, frequency discrimination of one tone in a pattern of changing tones is very poor, especially for tones at the beginning and at the end of the 10-tone pattern. However, as the random variation in the patterns is reduced, frequency discrimination improves, and the differences in discrimination as function of the temporal order of the tones are also reduced. When the same 10-tone pattern is presented on each trial (i.e., there is no randomization of the pattern frequencies) and only one tone is subjected to a frequency change, frequency discrimination thresholds for any one tone in the 10-tone pattern is nearly equal to that achieved when that tone is presented in isolation. These 10-tone pattern experiments show that the uncertainty about the stimulus context can have a large effect on performance in identifying complex sounds.

Information masking is used to describe the decrease in performance attributable to the stimulus context rather than to the actual values of the stimulus parameters. Thus, the changes in performance due to certain versus uncertain contexts in the 10-tone pattern experiments for the same stimulus values is due to informational masking. Another example of informational masking involves a tonal signal and a tonal-complex masker. If the tonal complex is a 100-tone masker and all 100 tones are mixed together at one time to form the masker, a certain signal level is required for signal detection (assume the signal frequency is in the center of the range of the frequencies used for the tonal-complex masker). If only 1 of the 100 tones in the tonal-complex masker is chosen at random and presented alone on each trial and masking of the signal is measured over the random presentation of the 100 tones, then signal threshold may be elevated by 20 or more decibels relative to the case when all 100 tones were mixed together at the same time to form the single tonal-complex masker. The increase in threshold is referred to as informational masking due to the uncertainty in the masking stimulus from trial to trial, despite the fact that the frequency range over which the masker varies is the same in both conditions, and on many trials the signal should be easy to detect because its frequency would be very different from that of the masker on that trial (Neff & Green, 1987).

Spatial Separation

The section on sound localization described the ability of listeners to locate a sound source based on the sound that is produced. When sound sources are located at different locations, does this spatial separation aid sound source segregation? Cherry (1953) stated that spatial separation would aid sound source segregation when he coined the term cocktail party effect. That is, spatial separation was a way to segregate one sound from the conglomeration of other sounds at a noisy cocktail party. Spatially separating sound sources does aid in the identification of the individual sound sources, especially when there are more than two sound sources (Yost, Dye, & Sheft, 1996).

The masked threshold for detecting a signal presented with one set of interaural differences can vary greatly as a function of the interaural differences of the masker. If the signal and masker are presented with a different set of interaural differences, then signal threshold is lower than in conditions in which the signal and masker are presented with the same interaural differences. The decibel difference in masked threshold between a condition in which the signal and masker have different interaural differences compared to that in which they have the same interaural differences is the masking-level difference, MLD (Green & Yost, 1975; Yost & Dye, 1991). For instance, if the masker (M) and the signal (S) each have no interaural differences (subscript 0), the condition is
M0S. If the masker is M0, but the signal is presented with a
180° (π radians) interaural phase difference, the condition is
M0S2. The threshold for detection of the signal in the M0S2
condition is 15–18 dB lower than it is in the M0S0 condition
(the MLD is 15–18 dB). The MLD has been studied for a
wide variety of stimulus conditions and interaural configura-
tions, and the MLD is always positive when the signal and
masker have a different set of interaural differences, as com-
pared to conditions in which the signal and masker have the
same set of interaural differences.

Because an interaural difference is associated with a ho-
izontal location in space, signals and maskers that have dif-
ferent interaural differences are similar to stimuli that are at
different positions in space. Thus, the results from the MLD
literature suggest that when the signal and masker are in dif-
f erent spatial locations (have different interaural differences),
signal threshold is lower than when the signal and masker oc-
cupy the same spatial locations (have the same interaural dif-
ferences). Such threshold differences do exist when signals
and maskers are presented from loudspeakers in real-world
spaces (Gilkey & Andersen, 1997). These results appear con-
sistent with Cherry’s observation about the role spatial sepa-
ration plays in solving the cocktail party problem. Models
that are variations of the coincidence models used to account
for processing interaural time differences have also been suc-
sessful in accounting for a great deal of the data from the
MLD literature (Colburn & Durlach, 1978).

Pitch and Timbre: Harmonicity and
Temporal Regularity

Pitch is a subjective attribute of sound that varies along a
low-high dimension and is highly correlated with the spectral
content of sound. The pitch of a target sound is often given in
terms of hertz, such that the pitch of a target sound is \( x \) Hz, if
a tone of \( x \) Hz is judged perceptually equal in pitch to the tar-
get sound. Musical scales, such as the 12-note scale, can also
be used to denote the pitch of a sound.

Timbre is defined as that subjective attribute of a sound
that differentiates two sounds that are otherwise equal in
pitch, loudness, and duration. Thus, the difference between
the sound from a cello playing the note G for the same dura-
tion and loudness as the sound from a violin playing the same
note G, is said to be a difference in timbre. The sound of the
cello differs in timbre from that of a violin. There are no units
for measuring timbre, and timbre is often correlated with the
spectral or temporal complexity of the sound.

Although the pitch of a sound is often highly correlated
with frequencies that are the most intense in a sound’s spec-
trum, many complex sounds produce a strong pitch in the
absence of such a concentration of spectral energy. Consider a
complex sound with frequency components of 300, 400, 500,
and 600 Hz. This sound will often have a 100-Hz pitch, even
though there is no spectral component at 100 Hz. Note that
100 Hz is the fundamental of this sound (all of the existing
spectral components are harmonics of a 100-Hz fundamental),
but the fundamental is missing. Thus, this type of complex
pitch is referred to as the “pitch of the missing fundamental.”
Many sound sources (e.g., most musical instruments) contain
a spectrum of harmonics. The pitches associated with these
sounds are derivatives of the pitch of the missing fundamental.

The stimulus described above that leads to the pitch of the
missing fundamental will often have a periodic time envelope,
which in this case will have a 100-Hz repetition (a 10-ms
period). Thus, the pitch may be associated with the temporal
regularity in the envelope. However, stimuli with very little
envelope periodicity can still produce a complex pitch like
that of the pitch of the missing fundamental. Such stimuli may
not have a smooth spectrum like that of the tonal complex
described above. Thus, neither envelope periodicity nor a
smooth spectrum appear to be necessary and sufficient con-
ditions for producing a complex pitch. However, such stimuli
without periodic temporal envelopes may contain a tempo-
rally regular, but nonperiodic, fine structure that may be the
basis for complex pitch (an analysis of this stimulus, such as
autocorrelation, will reveal this otherwise difficult-to-
determine temporal regularity; see Yost, 1996).

In addition to influencing the pitch of complex sounds, har-
monic structure also influences timbre. Thus, a complex
harmonic sound with high-amplitude, high-frequency har-
monics may have a brighter timbre than a complex sound
with high-amplitude, low-frequency harmonics, which would
dhave a dull timbre. Certain forms of temporal regularity (e.g.,
noise vs. periodic sounds) can also influence a sound’s timbre.

Therefore, harmonic structure and temporal regularity are
important stimulus properties that help determine the pitch
and timbre of complex sounds. Complex sounds differ in
pitch and timbre, and, as such, these two subjective attributes
may allow for sound source segregation. Indeed, complex
pitch and timbre have both been used to segregate sound
sources in auditory stream experiments (Bregman, 1990).
The two-vowel paradigm (Summerfield & Assmann, 1991; 
Yost & Sheft, 1993) is another procedure used to study the in-
fluence of harmonicity on sound source segregation. In the
two-vowel procedure, two artificially generated (via com-
puter) vowels are mixed. Often it is difficult or impossible
to identify the two vowels generated in this manner. Any stim-
ulus manipulation that allows for vowel recognition in the
two-vowel stimulus is arguably a crucial stimulus condition
for sound source segregation. If the fundamental voicing
frequency of the two vowels is made to differ, then often vowel recognition is improved. Because a change in the fundamental voicing frequency also alters the harmonic structure of each vowel, this result suggests that harmonicity does support sound source segregation.

**Spectral Profile**

Most of the time, the identifiable properties of a sound source are level independent. For instance, an uttered sentence has about the same intelligibility at many different overall loudness levels. Thus, the overall spectral-temporal structure of the sound from a source, which remains constant as overall level is varied, is the important determiner of sound-source identification. An area of study referred to as profile analysis (Green, 1989) has been used to study this property of auditory perception.

In a typical profile analysis experiment, several tones that are all of the same level but of different frequencies are mixed together. The frequency spacing of the tones is usually logarithmic to avoid generating sounds with harmonic structure that may have complex pitches. The level of a tone in the spectral middle of the complex (the signal tone) is increased, and the level of this signal tone required for the complex with the signal to be just discriminable from the complex with all tones equal in level is measured in several different conditions. The key aspect of these profile studies is that the overall level of both complexes is randomly varied across stimulus presentations over a large range, such as 40 dB. The random variation would affect two possible cues for detection. If detection were based on just attending to the signal tone, the overall random-level variation would require a very large signal-level increment for discrimination. An increase in the level of the signal will increase the overall level of the complex as compared to the complex in which all tones are presented at the same level. Thus, overall level (or loudness) could be a cue for detection. However, the random overall level variation would again require a very large signal-level increment if this loudness cue were to be the basis for discrimination. If, on the other hand, listeners could use the relative change in level between the level of the signal as compared to the level of the other tones in the complex, then the random overall level variation would not affect this cue. The complex with the signal increment would have a pointed spectral profile, whereas the complex without the signal increment would have a flat profile. Thus, if this spectral profile cue were used, then discrimination between the signal and nonsignal complexes might occur for small changes in signal level.

The data in Figure 5.14 suggest that such a spectral profile cue is being used. The level of the tonal signal required for detection of the signal increment is shown as a function of the number of total tones in the complex. With 11 tones in the complex, the threshold is about the same as it was when the signal was presented in isolation of any flanking tones. When there are fewer or more flanking tones in the complex than 11, thresholds are higher. When a large number of tones fit into the same bandwidth, the tones are so close together that they directly interact, so that tones near that of the signal mask the signal. The increase in threshold with increases in tonal density is consistent with other masking data. When there are only a few tones in the complex, it is argued that the profile is difficult to determine; for example, there is a large spectral difference between the signal tone and its nearest neighbor, making it difficult to discern the spectral profile. A model of listener performance, based on how listeners weigh the spectral information in these tonal profile complexes, suggests that listeners do monitor the spectral profile of these stimuli as the basis for their discrimination judgments (Green, 1989).

Experiments like these profile experiments suggest that the auditory system is very sensitive to subtle changes in the spectral shape of complex signals. Thus, sounds from different sources can be segregated based on changes in spectral shape. Note that the use of spectral shape requires the auditory system to process information across a wide spectral range.
Common Onsets and Offsets

It is often the case that although sounds from different sources may occur at about the same time, one sound may come on or go off at a slightly different time than another sound. When this happens, all of the temporal-spectral characteristics of one sound come on and go off at a different time than that occurring for the other sound. Thus, the common onset or offset of these spectral-temporal cues could be used for sound source segregation.

Asynchronous onsets, and in some cases offsets, have been shown to provide powerful cues for sound source segregation (Yost & Sheft, 1993). In some cases, onset cues can be used to amplify other cues that might be used for sound source segregation. As described above in the section on pitch, a harmonic sequence can produce a complex pitch equal to the fundamental frequency of the complex. If two complexes with different fundamentals are mixed, in most conditions listeners do not perceive the two pitches corresponding to the original two fundamental frequencies. The spectral characteristics of the new complex consisting of the mixture of the two harmonic sequences appear to be analyzed as a whole (synthetically). However, if one of the harmonic complexes is turned on slightly before (50 ms) the other harmonic complex, listeners often perceive the two pitches, even though for most of the time (perhaps for a second) the two harmonic complexes occur together (Darwin, 1981).

Common Modulation

Most everyday sound sources impart a slow amplitude and frequency modulation (change) to the overall spectral-temporal properties of the sound from the source. Each sound source will produce a different pattern of modulation, and these modulation patterns may allow for sound source segregation (Yost & Sheft, 1993). When a person speaks, the vocal cords open and close in a nearly periodic manner that determines the pitch of a voice (Fowler chapter in this volume). However, the frequency of these glottal openings varies (frequency modulation, voicing vibrato) slightly, and the amplitude of air released by each opening also randomly varies (amplitude modulation, voicing jitter) over a small range. Each person has a different pattern of vibrato and jitter. Speech sounds can be artificially generated (via computer) such that the speech (see Fowler chapter in this volume) is produced with constant glottal frequency and amplitude. If two such constant speech sounds are generated and mixed, it is often difficult to segregate the two sounds into the two different speech signals. However, if random variation is introduced into the computer-generated glottal openings and closing (random vibrato and jitter), segregation can occur (McAdams, 1984).

Thus, common amplitude and frequency modulation may be possible cues for sound source segregation. However, frequency modulation per se is probably not a cue used for sound source segregation (Carylon, 1991), but amplitude modulation is most likely a useful cue. Two experimental procedures have been extensively studied to investigate the role of amplitude modulation in auditory processing: comodulation masking release (CMR) and modulation detection interference (MDI).

In a typical CMR experiment (Hall, Haggard, & Fernandes, 1984; Yost & Sheft, 1993) listeners are asked to identify a tonal signal spectrally centered in the middle of a narrow band of noise (target band). In one condition, the detection of the signal is compared to a case in which another narrow band of noise (the flanking band) is simultaneously added in another region of the spectrum. The addition of this flanking band has little effect on signal threshold in the target band, if the target and flanking bands are completely independent. This is consistent with the critical-band view of auditory processing, in that the flanking band falls outside the spectral region of the critical band of the target band and therefore should have little influence on signal detection within the target band. However, if the target and flanking band are dependent in that they have the same pattern of amplitude modulation (they are comodulated), then signal threshold for the target band is lowered by 10–15 dB. This improvement in signal threshold due to comodulation is referred to as CMR, and the results from a typical experiment are shown in Figure 5.15.

The CMR results suggest that the common modulation increases the listener’s ability to detect the signal. One explanation of these results is based on the assumption that comodulation groups the flanking and target bands into one perceived sound source that contains more information than that in a single band. Independent (non-comodulated) bands of noise would not come from a single sound source and therefore would not be grouped together. The additional information in the combined (grouped) sound might aid signal detection. For instance, it might make the valleys of low amplitude in the modulated noises more obvious, increasing the ability of the auditory system to detect the tone occurring in these valleys. The addition of the signal changes the correlation between the signal-plus-masking stimuli and the masking-alone stimuli. The combined stimulus may increase this correlation, increasing signal detection.

In an MDI condition (Yost, 1992b; Yost & Sheft, 1993), listeners are asked to discriminate between two amplitude-modulated tonal carrier signals (the probe stimuli) on the basis of the depth of the amplitude modulation. Threshold performance is typically a 3% change in the depth of amplitude modulation. If a tone of a different frequency and not
Sound Source Segregation

If the masker is now amplitude modulated with the same amplitude modulation pattern (same rate of modulation) of the probe, then threshold is increased to around 20% (a 15–16 dB increase). The increase in modulation-depth threshold due to the common pattern of amplitude modulation between the probe and masker is referred to as modulation detection interference, MDI, and typical results are shown in Figure 5.16.

One argument for why MDI occurs is that the probe and masker are grouped together as a single sound source based on the common pattern of amplitude modulation. Because the common pattern of amplitude modulation is the basis of the grouping, the auditory system has difficulty detecting changes in modulation unless it affects the pattern of modulation. Because changes in amplitude modulation depth of one tone would have a small effect on the modulation pattern of the mixture of the two tones, it is difficult for the auditory system to detect changes in the depth of amplitude modulation for the probe. One test of this argument is to make the pattern (rate) of masker modulation different from that of the probe. In this case, the masker and probe would not be grouped as a single sound source, and MDI would be less or disappear. The data shown in Figure 5.16 are consistent with this argument.

Models or Theories of Sound Source Segregation

One key aspect of accounting for sound source segregation is the recognition that such processing requires the auditory system to process sound across a wide spectral range and...
over time. This is in contrast to the critical-band approach to explaining auditory processing, in which only a narrow region of the spectrum (in the critical band) is processed in a very short period of time. As pointed out above, explanations of profile analysis, CMR, and MDI all assume wide-band spectral processing and procedures like auditory stream segregation emphasize the importance of information integration over a long period of time.

Bregman (1990) has approached explanations of sound source segregation from a perceptual point of view, borrowing many concepts from the Gestalt school of perception. Several computational models have been developed to account for aspects of sound source segregation, especially those from auditory stream segregation experiments. These models are usually based on pattern recognition computations that interrogate spectral-temporal patterns generated by modeling the processes of the auditory periphery (Patterson, Allerhand, & Giguere, 1995). Computational models of the auditory periphery simulate the frequency-resolving properties of the cochlear partition (often using a bank of band-pass filters) and simulations of hair cell transduction of stereocilia displacement to neural discharges in the auditory nerve (Meddis & Hewitt, 1992). The pattern recognizers are neural nets or similar methods of computation that attempt to segregate the spectral-temporal neural patterns into subparts, whereby each subpart may reveal the spectral-temporal structure of a particular sound source. The cues discussed in this chapter, as well as a priori information about the stimulus context or prior learning about the stimulus context, are used to segregate the overall spectral-temporal pattern into these subparts. These models clearly imply that sound source segregation is based on processing the spectral-temporal code provided by the auditory periphery, and hence sound source segregation is a central process (Meddis & Hewitt, 1992). As of yet, little direct physiological data are available that can be used to help guide these modeling efforts.

AN OVERVIEW OF THE FUTURE STUDY OF AUDITION

A great deal of what is known about hearing comes from understanding the causes of hearing loss and its treatment. The major links in the hearing process that are most vulnerable to damage are the intricate structures of the inner ear, especially the hair cells. The study of the function of inner and outer hair cells and the exact consequences each plays in hearing will continue to be a major research focus in audition. The recent suggestions that the compressive nonlinear properties of cochlear transduction are derived from outer hair cell function have led to a better understanding of auditory perception in both people with normal hearing and those with impaired hearing. Perhaps, however, the most exciting discovery concerning hair cells is the fact that hair cells in birds, fish, and probably amphibians regenerate after damage due to overexposure to either sound or ototoxic drugs (Tsue, Osterle, & Rubel, 1994). These regenerated hair cells in birds appear to function normally in support of normal hearing, but additional work is needed to fully understand the perceptual abilities of these animals with regenerated hair cells. Hair cells in mammals do not regenerate. The quest is on to determine why hair cell regeneration occurs in some nonmammals but not in mammals. The ability to regrow hair cells could, for many different types of hearing loss, be the ultimate hearing aid.

The study of hair cell regeneration is one of many areas in which genetic techniques are supplying new and important facts about auditory function. In addition to revealing...
important clues for understanding normal audition, the genetic revolution has made significant strides in identifying the genetic basis for several different forms of inheritable deafness. A gene that may control the motile response of the outer hair cells has been identified (Zheng, Shen, He, Long, & Pallos, 2000), opening a whole array of possibilities (e.g., genic manipulation) for better understanding outer hair-cell function. In many areas, perceptual research should provide improved ways to determine different phenotypes in order to better define structure-function auditory relationships.

The development of better hearing aids, both amplification hearing aids and the cochlear prosthesis, has stimulated new knowledge about audition; these technologies have benefited from the past research on hearing as well. The cochlear prosthesis in particular offers unique opportunities to study the hearing process (Miller & Spelman, 1989). The cochlear prosthesis is a wire with multiple electrodes that is surgically inserted into the cochlear partition of a patient with a hearing loss. The electrodes stimulate selected portions of the cochlear partition, based on the transduction of sound into electrical current via a sound processor worn by the patient. The success achieved by thousands of cochlear prosthetic users worldwide suggests that these devices provide a useful means of aural communication for many people with hearing impairments. Because the use of cochlear prostheses bypasses the biomechanical properties of the inner ear, understanding the auditory abilities of successful implant users provides valuable information about the early neural stages of the auditory process. Many successful users of the cochlear prostheses had been deprived of useful hearing for many years before their implantation. The significant improvement in auditory abilities achieved by these cochlear prostheses users, after implementation and training, suggests a degree of auditory plasticity that is receiving a great deal of attention. The importance of this issue has increased now that young children are being implanted.

In addition to providing potential utility for hearing aids, research on spatial hearing and the HRTF have provided improvements for devices used in many sound localization situations (Gilkey & Andersen, 1997). For instance, many traditional hearing aids (especially if only one hearing aid is used) do not allow users to accurately localize sound sources. Proper use of HRTF technology may enable hearing aid users to more accurately localize sound sources, and such accuracy may also improve their ability to detect sounds in noisy environments (the aforementioned cocktail party effect). HRTF technology has also been adopted in the audio entertainment and other industries.

The use of the HRTF offers complete control of the sound cues that are important for sound localization. Such control offers several advantages for studying hearing (Gilkey & Andersen, 1997). One interesting use of HRTF-transformed sound is in the study of auditory adaptation and neural plasticity to alterations of the normal cues for sound localization (Hofman, Van Riswick, & Van Opstal, 1998). If the HRTF is altered such that the location of a sound source is now perceived at a new location, listeners can adapt to the change and after a few days demonstrate near-normal sound localization abilities. When the nonnormal alterations are removed, listeners quickly return to being able to accurately localize as they had before the alteration. Such sound localization adaptation research with human and animal (e.g., the barn owl) listeners is revealing and will continue to reveal important insights about the plasticity of neural sound localization processes (Knudsen, Esterly, & Olsen, 1994).

In the late 1980s and early 1990s, several authors (Hartmann, 1988; Moore, 1997; Yost, 1992a), most notably Bregman (1990), suggested that our ability to determine the sources of sounds, especially in multisource acoustic environments, was a major aspect of hearing about which very little was known. Although the early history of the study of hearing suggests that so-called object identification was an important aspect of hearing, for most of the last century and a half the study of audition focused on the detection and discrimination of the attributes of sound—frequency, level, and timing—and how those attributes were coded in the auditory periphery (Yost, 1992a). While there is still not a lot known about how the auditory scene is achieved, current research in hearing is no longer focused on processing in narrow frequency bands and over very short temporal durations. Psychophysical and physiological investigators have examined and will continue to investigate auditory mechanisms that integrate acoustic information across the spectrum and over time, because such processing is crucial for sound source determination.

The progress in understanding auditory scene processing may be hindered by a lack of appropriate techniques to study these problems. New correlation techniques in psychophysics, multiple electrode technology, new physiological techniques, new ways of extracting information from neural data, and neural imaging are some of the new methods that may open up opportunities for understanding sound source determination and audition. Auditory science also knows very little about the functional purposes of the auditory nuclei in the ascending auditory pathway and within the auditory cortex. With a few notable exceptions of several animal models (e.g., bats and echo processing; barn owls and sound localization), very little is known about the roles various neural centers play in hearing. A great deal is known about the anatomy of many neural circuits and the physiological properties of many types of fibers in most neural centers, but far less is known about what
function those circuits and fibers play in hearing. This is perhaps understandable because most of functional hearing is based on significant neural computation of the incoming auditory signal. The neural centers that perform the computations necessary for sound localization, especially computations for interaural time and level differences, are beginning to be sorted out, probably because a great deal is known about the type of computations that are required for accurate sound localization. Similar work for understanding the functionality of other auditory neural centers will continue to be an intense area of interest for auditory science. Progress will require a better understanding of auditory neural circuits in the central auditory system, additional knowledge about the auditory cues used for sound source determination, and testable models and theories of sound source determination.

Models of auditory processing, especially computational models, have provided significant new insights into possible mechanisms of central auditory processing (Hawkins, McMullen, Popper, & Fay, 1996). Several computational models of the auditory periphery have been shown to produce accurate representations of the spectral-temporal code provided by the inner ear and auditory nerve. These models can be used instead of the laborious collection of physiological data to explore possible models of central mechanisms for processing sound. Additional work on these models will continue to be an active area of research in audition.

There is a growing interest in neural imaging (e.g., PET; positron-emission tomography), especially fMRI (functional magnetic resonance imaging), as a potentially potent tool for probing neural mechanisms of auditory processing. Some of the most recent work is focusing on basic auditory processing (Griffith, Buchel, Frankowiak, & Patterson, 1998), as opposed to speech and language processing based on spoken language (see Fowler chapter in this volume). Such imaging research will be most useful when the spatial and temporal scales of measurement allow one to study the individual neural circuits involved with hearing. New imaging techniques, such as cardiac triggering and sparse imaging, are just now demonstrating the promise this technology might provide for better understanding auditory processing, especially in human listeners.

Although remarkable progress has been made in understanding audition, the field really is in its infancy when one considers all that is not known. While more is to be learned about the auditory periphery and the detectability and discriminability of sounds, a major challenge facing audition is unraveling the function of the central auditory nervous system and how it supports our abilities to process the sound sources that constantly bombard us with crucial information about the world in which we live.

REFERENCES


CHAPTER 6

Touch

ROBERTA L. KLATZKY AND SUSAN J. LEDERMAN

TOUCH DEFINED AS AN ACTIVE, MULTISENSORY SYSTEM 148
THE NEUROPHYSIOLOGY OF TOUCH 148
The Skin and Its Receptors 148
Pathways to Cortex and Major Cortical Areas 150
SENSORY ASPECTS OF TOUCH 150
Cutaneous Sensitivity and Resolution 150
Effects of Body Site and Age on Cutaneous Thresholds 150
Sensory-Guided Grasping and Manipulation 151
HAPTIC PERCEPTION OF PROPERTIES OF OBJECTS AND SURFACES 152
Haptically Perceptible Properties 152
Roughness 153
Weight 154
Curvature 155
Role of Manual Exploration in Perceiving Object Properties 156
Relative Availability of Object Properties 158

HAPTIC SPACE PERCEPTION 159
HAPTIC PERCEPTION OF TWO- AND THREE-DIMENSIONAL PATTERNS 160
Vibrotactile Patterns 161
Two-Dimensional Patterns and Freestanding Forms 161
Two-Dimensional Outline Drawings of Common Objects 161
Three-Dimensional Objects 162
VISUAL-HAPTIC INTERACTIONS 164
Attention 164
Cross-Modal Integration 165
HAPTIC MEMORY 165
APPLICATIONS OF RESEARCH ON TOUCH 167
Aids for the Visually Impaired 167
Haptic Augmentation of Speech Perception 168
Teleoperation and Virtual Environments 169
SUMMARY AND FUTURE DIRECTIONS 169
REFERENCES 170

This chapter describes a sensory modality that underlies the most common everyday activities: maintaining one’s posture, scratching an itch, or picking up a spoon. As a topic of psychological research, touch has received far less attention than vision has. However, the substantial literature that is available covers topics from neurophysiology, through basic psychophysics, to cognitive issues such as memory and object recognition. All these topics are reviewed in the current chapter.

We begin by defining the modality of touch as comprising different submodalities, characterized by their neural inputs. A brief review of neurophysiological and basic psychophysical findings follows. The chapter then pursues a number of topics concerning higher-level perception and cognition. Touch is emphasized as an active modality in which the perceiver seeks information from the world by exploratory movements. We ask how properties of objects and surfaces—like roughness or size—are perceived through contact and movement. We discuss the accuracy of haptic space perception and why movement might introduce systematic errors or illusions. Next comes an evaluation of touch as a pattern-recognition system, where the patterns range from two-dimensional arrays like Braille to real, free-standing objects. In everyday perception, touch and vision operate together; this chapter offers a discussion of how these modalities interact. Higher-level cognition, including attention and memory, is considered next. The chapter concludes with a review of some applications of research on touch.

A number of common themes underlie these topics. One is the idea that perceptual modalities are similar with respect to general functions they attempt to serve, such as conveying information about objects and space. Another is that by virtue of having distinct neural structures and relying on movement for input, touch has unique characteristics. The chapter makes the point that touch and vision interact cooperatively in extracting information about the world, but that the two
The modality of touch encompasses several distinct sensory systems. Most researchers have distinguished among three systems—cutaneous, kinesthetic, and haptic—on the basis of the underlying neural inputs. In the terminology of Loomis and Lederman (1986), the cutaneous system receives sensory inputs from mechanoreceptors—specialized nerve endings that respond to mechanical stimulation (force)—that are embedded in the skin. The kinesthetic system receives sensory inputs from mechanoreceptors located within the body’s muscles, tendons, and joints. The haptic system uses combined inputs from both the cutaneous and kinesthetic systems. The term haptic is associated in particular with active touch. In an everyday context, touch is active; the sensory apparatus is intertwined with the body structures that produce movement. By virtue of moving the limbs and skin with respect to surfaces and objects, the basic sensory inputs to touch are enhanced, allowing this modality to reveal a rich array of properties of the world.

When investigating the properties of the peripheral sensory system, however, researchers have often used passive, not active, displays. Accordingly, a basic distinction has arisen between active and passive modes of touch. Unfortunately, over the years the meaning and use of these terms have proven to be somewhat variable. On occasion, J. J. Gibson (1962, 1966) treated passive touch as restricted to cutaneous (skin) inputs. However, at other times Gibson described passive touch as the absence of motor commands to the muscles (i.e., efferent commands) during the process of information pickup. For example, if an experimenter shaped a subject’s hands so as to enclose an object, it would be a case of active touch by the first criterion, but passive touch by the second one. We prefer to use Loomis and Lederman’s (1986) distinctions between types of active versus passive touch. They combined Gibson’s latter criterion, the presence or absence of motor control, with the three-way classification of sensory systems by the afferent inputs used (i.e., cutaneous, kinesthetic, and haptic). This conjunction yielded five different modes of touch: (a) tactile (cutaneous) perception, (b) passive kinesthetic perception (kinesthetic afferents respond without voluntary movement), (c) passive haptic perception (cutaneous and kinesthetic afferents respond without voluntary movement), (d) active kinesthetic perception, and (e) active haptic perception. The observer only has motor control over the touch process in modes d and e.

In addition to mechanical stimulation, the inputs to the touch modality include heat, cooling, and various stimuli that produce pain. Tactile scientists distinguish a person’s subjective sensations of touch per se (e.g., pressure, spatial acuity, position) from those pertaining to temperature and pain. Not only is the quality of sensation different, but so too are the neural pathways. This chapter primarily discusses touch and, to a lesser extent, thermal subsystems, inasmuch as thermal cues provide an important source of sensory information for purposes of haptic object recognition. Overviews of thermal sensitivity have been provided by Sherrick and Cholewiak (1986) and by J. C. Stevens (1991). The topic of pain is not extensively discussed here, but reviews of pain responsiveness by Sherrick and Cholewiak (1986) and, more recently, by Craig and Rollman (1999) are recommended.

The Skin and Its Receptors

The skin is the largest sense organ in the body. In the average adult, it covers close to 2 m and weighs about 3–5 kg (Quilliam, 1978). As shown in Figure 6.1, it consists of two major layers: the epidermis (outer) and the dermis (inner). The encapsulated endings of the mechanoreceptor units, which are believed to be responsible for transducing mechanical energy into neural responses, are found in both layers, as well as at the interface between the two. A third layer lies underneath the dermis and above the supporting structures made up of muscle and bone. Although not considered part of the formal medical definition of skin, this additional layer (the hypodermis) contains connective tissue and subcutaneous fat, as well as one population of mechanoreceptor end organs (Pacinian corpuscles).

We focus here on the volar portion of the human hand, because the remainder of this chapter considers interactions of the hand with the world. This skin, which is described as glabrous (hairless), contains four different populations of cutaneous mechanoreceptor afferent units. These populations are differentiated in terms of both relative receptive field size
and adaptation responses to sustained and transient stimulation (see Table 6.1).

The two fast-adapting populations (FA units) show rapid responses to the onset, and sometimes the offset, of skin deformation. In addition, FAI (fast adapting type I) units have very small, well-defined receptive fields, whereas FAII (fast adapting type II) units have large receptive fields with poorly defined boundaries. FAI units respond particularly well to rate of skin deformation, and they are presumed to end in Meisner’s corpuscles. FAII units respond reliably to both the onset and offset of skin deformation, particularly acceleration and higher-derivative components, and have been shown to terminate in Pacinian corpuscles. The two slow-adapting populations (SA units) show a continuous response to sustained skin deformation. SAI (slow adapting type I) units demonstrate a strong dynamic sensitivity, as well as a somewhat irregular response to sustained stimulation. They are presumed to end in Merkel cell neurite complexes (see Figure 6.1). SAII (slow adapting type II) units show less dynamic sensitivity but a more regular sustained discharge, as well as spontaneous discharge sometimes in the absence of skin deformation; they are presumed to end in Ruffini endings. Bolanowski, Gescheider, Verrillo, and Checkosky (1988) have developed a four-channel model of mechanoreception, which associates psychophysical functions with the tuning curves of mechanoreceptor populations. Each of the four mechanoreceptors is presumed to produce different psychophysical responses, constituting a sensory channel, so to speak.

Response to thermal stimulation is mediated by several peripheral cutaneous receptor populations that lie near the body surface. Researchers have documented the existence of separate “warm” and “cold” thermoreceptor populations in the skin; such receptors are thought to be primarily responsible for thermal sensations. Nociceptor units respond only to extremes (noxious) in temperature (or sometimes mechanical) stimulation, but these are believed to be involved in pain rather than temperature sensation.

Response to noxious stimulation has received an enormous amount of attention. Here, we simply note that two populations of peripheral afferent fibers (high-threshold nociceptors) in the skin have been shown to contribute to pain transmission: the larger, myelinated A-delta fibers and the narrow, unmyelinated C fibers.

Mechanoreceptors in the muscles, tendons, and joints (and in the case of the hand, in skin as well) contribute to the kinesthetic sense of position and movement of the limbs. With respect to muscle, the muscle spindles contain two types of sensory endings: Large-diameter primary endings code for rate of change in the length of the muscle fibers, dynamic stretch, and vibration; smaller-diameter secondary endings are primarily sensitive to the static phase of muscle activity. It is now known that joint angle is coded primarily by muscle length. Golgi tendon organs are spindle-shaped receptors that lie in series with skeletal muscle fibers. These receptors code muscle tension. Finally, afferent units of the joints are now known to code primarily for extreme, but not intermediate, joint positions. As they do not code for intermediate joint positions, it has been suggested that they serve mainly a protective function—detecting noxious stimulation. The way in which the kinesthetic mechanoreceptor units mediate perceptual outcomes is not well understood, especially

Table 6.1 Four Mechanoreceptor Populations in the Glabrous Skin of the Human Hand, with Their Defining Characteristics

<table>
<thead>
<tr>
<th>Receptive Field</th>
<th>Adaptation Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small, well defined</td>
<td>Fast; No response to sustained stimulation</td>
</tr>
<tr>
<td>Large, diffuse</td>
<td>FAI</td>
</tr>
<tr>
<td></td>
<td>FAII</td>
</tr>
</tbody>
</table>

*Note: FA = fast adapting; SA = slow adapting; and I and II index types within each classification.*
in comparison to cutaneous mechanoreceptors. For further details on kinesthesia, see reviews by Clark and Horch (1986) and by Jones (1999).

Pathways to Cortex and Major Cortical Areas

Peripheral units in the skin and muscles congregate into single nerve trunks at each vertebral level as they are about to enter the spinal cord. At each level, their cell bodies cluster together in the dorsal root ganglion. These ganglia form chains along either side of the spinal cord. The proximal ends of the peripheral units enter the dorsal horn of the spinal cord, where they form two major ascending pathways: the dorsal column-medial lemniscal system and the anterolateral system. The dorsal column-medial lemniscal system carries information about tactile sensation and limb kinesthesia. Of the two systems, it conducts more rapidly because it ascends directly to the cortex with few synapses. The anterolateral system carries information about temperature and pain—and to a considerably lesser extent, touch. This route is slower than the dorsal column-medial lemniscal system because it involves many synapses between the periphery and the cortex. The two pathways remain segregated until they converge at the thalamus, although even there the separation is preserved.

The primary cortical receiving area for the somatic senses, S-I, lies in the postcentral gyrus and in the depths of the central sulcus. It consists of four functional areas, which when ordered from the central sulcus back to the posterior parietal lobe, are known as Brodmann’s areas 3a, 3b, 1, and 2. Lateral and somewhat posterior to S-I is S-II, the secondary somatic sensory cortex, which lies in the upper bank of the lateral sulcus. S-II receives its main inputs from S-I. The posterior parietal lobe (Brodmann’s areas 5 and 7) also receives somatic inputs. It serves higher-level associative functions, such as relating sensory and motor processing, and integrating the various somatic inputs (for further details, see Kandel, Schwartz, & Jessell, 1991).

SENSORY ASPECTS OF TOUCH

Cutaneous Sensitivity and Resolution

Tests of absolute and relative sensitivity to applied force describe people’s threshold responses to intensive aspects of mechanical deformation (e.g., the depth of penetration of a probe into the skin). In addition, sensation magnitude has been scaled as a function of stimulus amplitude, in order to reveal the relation between perceptual response and stimulus variables at suprathreshold levels. Corresponding psychophysical experiments have been performed to determine sensitivity to warmth and cold, and to pain. A review chapter by Sherrick and Cholewiak (1986) has described basic findings in this area in detail (see also Rollman, 1991; Stevens, 1991).

The spatial resolving capacity of the skin has been measured in a variety of ways, including the classical two-point discrimination method, in which the threshold for perceiving two punctate stimuli as a single point is determined. However, Johnson and Phillips (1981; see also Craig & Johnson, 2000; Loomis, 1979) have argued persuasively that grating orientation discrimination provides a more stable and valid assessment of the human capacity for cutaneous spatial resolution. Using spatial gratings, the spatial acuity of the skin has been found to be about 1 mm.

The temporal resolving capacity of the skin has been evaluated with a number of different methods (see Sherrick & Cholewiak, 1986). For example, it has been assessed in terms of sensitivity to vibratory frequency. Experiments have shown that human adults are able to detect vibrations up to about 700 Hz, which suggests that they can resolve temporal intervals as small as about 1.4 ms (e.g., Verrillo, 1963). A more conservative estimate (5.5 ms) was obtained when determining the minimum separation time between two 1-ms pulse stimuli that is required for an observer to perceive them as successive.

Overall, the experimental data suggest that the hand is poorer than the eye and better than the ear in resolving fine spatial details. On the other hand, it has proven to be better than the eye and poorer than the ear in resolving fine temporal details.

Effects of Body Site and Age on Cutaneous Thresholds

It has long been known that the sensitivity, acuity, and magnitude of tactile and thermal sensations can vary quite substantially as a function of the body locus of stimulation (for details, see van Boven & Johnson, 1994; Stevens, 1991; Weinstein, 1968; Wilska, 1954). For example, the face (i.e., upper lip, cheek, and nose) is best able to detect a low-level force, whereas the fingers are most efficient at processing spatial information. The two-point threshold is shown for various body sites in Figure 6.2.

More recently, researchers have addressed the effect of chronological age on cutaneous thresholds (for details, see Verrillo, 1993). One approach to studying aging effects is to examine the vibratory threshold (the skin displacement at which a vibration becomes detectable) as a function of age. A number of studies converge to indicate that aging particularly affects thresholds for vibrations in the range detected by the Pacinian corpuscles (i.e., at frequencies above 40 Hz; see
Figure 6.2 The minimal separation between two points needed to perceive them as separate (2-point threshold), when the points are applied at different sites of the body. Source: From Weinstein (1968), in D. R. Kenshalo, The Skin Senses, 1968. Courtesy of Charles C. Thomas, Publisher, Ltd., Springfield, Illinois. Reprinted with permission.

Gescheider, Bolanowski, Verrillo, Hall, & Hoffman, 1994; Verrillo, 1993). The rise in the threshold with age has been attributed to the loss of receptors. By this account, the Pacinian threshold is affected more than are other channels because it is the only one whose response depends on summation of receptor outputs over space and time (Gescheider, Edwards, Lackner, Bolanowski, & Verrillo, 1996). Although the ability to detect a vibration in the Pacinian range is substantially affected by age, the difference limen—the change in amplitude needed to produce a discriminable departure from a baseline value—varies little after the baseline values are adjusted for the age-related differences in detection threshold (i.e., the baselines are equated for magnitude of sensation relative to threshold; Gescheider et al., 1996).

Cutaneous spatial acuity has also been demonstrated to decline with age. Stevens and Patterson (1995) reported an approximate 1% increase in threshold per year over the ages of 20 to 80 years for each of four acuity measures. The measures were thresholds, as follows: minimum separation of a 2-point stimulus that allows discrimination of its orientation on the finger (transverse vs. longitudinal), minimum separation between points that allows detection of gaps in lines or disks, minimum change in locus that allows discrimination between successive touches on the same or different skin site, and difference limen for length of a line stimulus applied to the skin.

The losses in cutaneous sensitivity that have been described can have profound consequences for everyday life in older persons because the mechanoreceptors function critically in basic processes of grasping and manipulation.

Sensory-Guided Grasping and Manipulation

Persons who have sustained peripheral nerve injury to their hands are often clumsy when grasping and manipulating objects. Such persons will frequently drop the objects; moreover, when handling dangerous tools (e.g., a knife), they can cut themselves quite badly. Older adults, whose cutaneous thresholds are elevated, tend to grip objects more tightly than is needed in order to manipulate them (Cole, 1991). Experiments have now confirmed what these observations suggest: Namely, cutaneous information plays a critical role in guiding motor interactions with objects following initial contact. Motor control is discussed extensively in the chapter written by Heuer in this volume.
Neurophysiological evidence by Johansson and his colleagues (see review by Johansson & Westling, 1990) has clearly shown that the mechanoreceptor populations present in glabrous skin of the hand, particularly the FAI receptors, contribute in vital ways to the skill with which people are able to grasp, lift, and manipulate objects using a precision grip (a thumb-forefinger pinch). The grasp-lift action requires that people coordinate the grip and load forces (i.e., forces perpendicular and tangential to the object grasped, respectively) over a sequence of stages. The information from cutaneous receptors enables people to grasp objects highly efficiently, applying force just sufficient to keep them from slipping. In addition to using cutaneous inputs, people use memory for previous experience with the weight and slipperiness of an object in order to anticipate the forces that must be applied. Johansson and Westling have suggested that this sensorimotor form of memory involves programmed muscle commands. If the anticipatory plan is inappropriate—for example, if the object slips from the grasp or it is lighter than expected and the person overgrips—the sensorimotor trace must be updated. Overt errors can often be prevented, however, because the cutaneous receptors, particularly the FAIs, signal when slip is about to occur, while the grip force can still be corrected.

**HAPTIC PERCEPTION OF PROPERTIES OF OBJECTS AND SURFACES**

Up to this point, this chapter has discussed the properties of touch that regulate very early processing. The chapter now turns to issues of higher-level processing, including representations of the perceived world, memory and cognition about that world, and interactions with other perceptual modalities. A considerable amount of work has been done in these areas since the review of Loomis and Lederman (1986). We begin with issues of representation. What is it about the haptically perceived world—its surfaces, objects, and their spatial relations—that we represent through touch?

Klatzky and Lederman (1999a) pointed out that the haptic system begins extracting attributes of surfaces and objects from the level of the most peripheral units. This contrasts with vision, in which the earliest output from receptors codes the distribution of points of light, and considerable higher-order processing ensues before fundamental attributes of objects become defined.

The earliest output from mechanoreceptors and thermal receptors codes attributes of objects directly through various mechanisms. There may be different populations of peripheral receptors, each tuned to a particular level of some dimension along which stimuli vary. An example of this mechanism can be found in the two populations of thermoreceptors, which code different (but overlapping) ranges of heat flow. Another example can be found in the frequency-based tuning functions of the mechanoreceptors (Johansson, Landstrom, & Lundstrom, 1982), which divide the continuum of vibratory stimuli. Stimulus distinctions can be made within single units as well: for example, by phase locking of the unit’s output to a vibratory input (i.e., the unit fires at some multiple of the input frequency). The firing rate of a single unit can indicate a property such as the sharpness of a punctate stimulus (Vierck, 1979). Above the level of the initial receptor populations are populations that combine inputs from the receptors to produce integrative codes. As is later described, the perception of surface roughness appears to result from the integration at cortical levels of inputs from populations of SAI receptors. Multiple inputs from receptors may also be converted to maps that define spatial features of surfaces pressed against the fingertip, such as curvature (LaMotte & Srinivasan, 1993; Vierck, 1979).

Ultimately, activity from receptors to the brain leads to a representation of a world of objects and surfaces, defined in spatial relation to one another, each bound to a set of enduring physical properties. We now turn to the principal properties that are part of that representation.

**Haptically Perceptible Properties**

Klatzky and Lederman (1993) suggested a hierarchical organization of object properties extracted by the haptic system. At the highest level, a distinction is made between geometric properties of objects and material properties. Geometric properties are specific to particular objects, whereas material properties are independent of any one sampled object.

At the next level of the hierarchy, the geometric properties are divided into size and shape. Two natural scales for these properties are within the haptic system, differentiated by the role of cutaneous versus kinesthetic receptors, which we call micro- and macrogeometric. At the microgeometric level, an object is small enough to fall within a single region of skin, such as the fingertip. This produces a spatial deformation pattern on the skin that is coded by the mechanoreceptors (particularly the SAI) and functions essentially as a map of the object’s spatial layout. This map might be called 2-1/2 D, after Marr (1982), in that the coding pertains only to the surfaces that are in contact with the finger. The representation extends into depth because the fingertip accommodates so as to have differential pressure from surface planes lying at different depth. At the macrogeometric level, objects do not fall within a single region of the skin, but rather are...
enveloped in hands or limbs, bringing in the contribution of kinesthetic receptors and skin sites that are not somatotopically continuous, such as multiple fingers. Integration of these inputs must be performed to determine the geometry of the objects.

The hierarchical organization of Klatzky and Lederman further differentiates material properties into texture, hardness (or compliance), and apparent temperature. Texture comprises many perceptually distinct properties, such as roughness, stickiness, and spatial density. Roughness has been the most extensively studied, and we treat it in some detail in a following section. Compliance perception has both cutaneous and kinesthetic components, the relative contributions of which depend on the rigidity of the object’s surface (Srinivasan & LaMotte, 1995). For example, a piano key is rigid on the surface but compliant, and kinesthesia is a necessary input to the perception that it is a hard or soft key to press. Although cutaneous cues are necessary, they are not sufficient, because the skin bottoms out, so to speak, whether the key is resistant or compliant. On the other hand, a cotton ball deforms as it is penetrated, causing a cutaneous gradient that may be sufficient by itself to discriminate compliance. Another property of objects is weight, which reflects geometry and material. Although an object’s weight is defined by its total mass, which reflects density and volume, we will see that perceived weight can be affected by the object’s material, shape, and identity.

A complete review of the literature on haptic perception of object properties would go far beyond the scope of this chapter. Here, we treat three of the most commonly studied properties in some detail: texture, weight, and curvature. Each of these properties can be defined at different scales, although the meaning of scale varies with the particular dimension of interest. The mechanisms of haptic perception may be profoundly affected by scale.

**Roughness**

A textured surface has protruberant elements arising from a relatively homogeneous substrate. The surface can be characterized as having macrotexture or microtexture, depending on the spacing between surface elements. Different mechanisms appear to mediate roughness perception at these two scales. In a microtexture, the elements are spaced at intervals on the order of microns (thousands of a millimeter); in a macrotexture, the spacing is one or two orders of magnitude greater, or more. When the elements get too sparse, on the order of 3–4 mm apart or so, people tend to be reluctant to characterize the surface as textured. Rather, it appears to be a smooth surface punctuated by irregularities.

Early research determined some of the primary physical determinants of perceived roughness with macrotextures (i.e., ≥ 1 mm spacing between elements). For example, Lederman (Lederman, 1974, 1983; Lederman & Taylor, 1972; see also Connor, Hsiao, Philips, & Johnson, 1990; Connor & Johnson, 1992; Sathian, Goodwin, John, & Darian-Smith, 1989; Sinclair & Burton, 1991; Stevens & Harris, 1962), using textures that took the form of grooves with rectangular profiles, found that perceived roughness strongly increased with the spacing between the ridges (groove width). Increases in ridge width—that is, the size of the peaks rather than the troughs in the surface—had a relatively modest effect, tending to decrease perceived roughness. Although roughness was principally affected by the geometry of the surface, the way in which the surface was explored also had some effect. Increasing applied fingertip force increased the magnitude of perceived roughness, and the speed of relative motion between hand and surface had a small but systematic effect on perceived roughness. Finally, conditions of active versus passive control over the speed-of-hand motion led to similar roughness judgments, suggesting that kinesthesia plays a minimal role, and that the manner in which the skin is deformed is critical.

Taylor and Lederman (1975) constructed a model of perceived roughness, based on a mechanical analysis of the skin deformation resulting from changes in groove width, fingertip force, and ridge width. Their model suggested that perceived roughness of gratings was based on the total amount of skin deformation produced by the stimulus. Taylor and Lederman described the representation of roughness in terms of this proximal stimulus as “intensive” because the deformation appeared to be integrated over the entire area of contact, resulting in an essentially unidimensional percept.

The neural basis for coding roughness has been modeled by Johnson, Connor, and associates (Connor et al., 1990; Connor & Johnson, 1992). The model assumes that initial coding of the textured surface is in terms of the relative activity rates of spatially distributed SAI mechanoreceptors. The spatial map is preserved in S-I, the primary somatosensory cortex (specifically, area 3b), which computes differences in activity of adjacent (1 mm apart) SAI units. These differences in spatially distributed activity are passed along to neurons in S-II, another somatosensory cortical area that integrates the information from the primary cortex (Hsiao, Johnson, & Twombly, 1993).

Although vibratory signals exist, psychophysical studies suggest that humans tend not to use vibration to judge macrotextures presented to the bare skin. Roughness judgments were unaffected by the spatial period of stimulus gratings (Lederman, 1974, 1983) and minimally affected by movement speed (Katz, 1925/1989; Lederman, 1974, 1983).
both of which should alter vibration; they were also unaffected by either low- or high-frequency vibrotactile adaptation (Lederman, Loomis, & Williams, 1982). Vibratory coding of roughness does, however, occur with very fine microtextures. LaMotte and Srinivasan (1991) found that observers could discriminate a featureless surface from a texture with height .06–.16 microns and interelement spacing ~100 microns. Subjects reported attending to the vibration from stroking the texture. Moreover, measures of mechanoreceptor activity in monkeys passively exposed to the same surfaces implicated the FAII (or PC) units, which respond to relatively high-frequency vibrations (peak response ~ 250 Hz; Johansson & Vallbo, 1983). Vibrotactile adaptation affected perceived roughness of fine but not coarse surfaces (Hollins, Bensmaia, & Risner, 1998).

Somewhat surprisingly, the textural scale where spatial coding of macrotexture changes to vibratory coding of microtexture appears to be below the limit of tactile spatial resolution (.5–1.0 mm). Dorsch, Yoshioka, Hsiao, and Johnson (2000) reported that SAI activity, which implicates spatial coding, was correlated with roughness perception over a range of gratings that began with a .1-mm groove width. Using particulate textures, Hollins and Risner (2000) found evidence for a transition between vibratory and spatial coding at a similar particle size.

Weight

The perception of weight has been of interest for a time approaching two centuries, since the work of Weber (1834/1978). Weber pointed out that the impression of an object’s heaviness was greater when it was wielded than when it rested passively on the skin, suggesting that the perception of weight was not entirely determined by its objective value. In the late 1800s (Charpentier, 1891; Dresslar, 1894), the discovery of the size-weight illusion—that given equal objective weight, a smaller object seems heavier—pointed to the fact that multiple physical factors determine heaviness perception. Recently, Amazeen and Turvey (1996) have integrated a body of work on the size-weight illusion and weight perception by accounting for perceived weight in terms of resistance to the rotational forces imposed by the limbs as an object is held and wielded. Their task requires the subject to wield an object at the end of a rod or handle, precluding volumetric shape cues. Figure 6.3 shows the experimental setup for a wielding task. Formally, resistance to wielding is defined by an entity called the inertia tensor, a three-by-three matrix whose elements represent the resistance to rotational acceleration about the axes of a three-dimensional coordinate system that is imposed on the object around the center of rotation. Although the inertia tensor will vary with the coordinate system that is imposed on the object, its eigenvalues are invariant. (The eigenvalues of a matrix are scalars that, together with a set of eigenvectors—essentially, coordinate axes—can be used to reconstruct it.) They correspond to the principal moments of inertia: that is, the resistances to rotation about a nonarbitrary coordinate system that uses the primary axes of the object (those around which the mass is balanced). In a series of experiments in which the eigenvalues were manipulated and the seminal data on the size-weight illusion were analyzed (Stevens & Rubin, 1970), Amazeen and Turvey found that heaviness was directly related to the product of power functions of the eigenvalues (specifically, the first and third). This finding explains why weight is not dictated simply by mass alone; the reliance of heaviness perception on resistance to rotation means that it will also be affected by geometric factors.

But the story is more complicated, it seems, as weight perception is also affected by the material from which an object is made and the way in which it is gripped. A material-weight relation was documented by Wolfe (1898), who covered objects of equal mass with different surface materials and found that objects having surface materials that were more dense were judged lighter than those with surfaces that were less dense (e.g., comparing brass to wood). Flanagan and associates (Flanagan, Wing, Allison, & Spencely, 1995; Flanagan & Wing, 1997; see also Rinkenauer, Mattes, & Ulrich, 1999).
suggested that material affected perceived weight because objects that were slipperier required a greater grip force in order to be lifted, and a more forceful grip led to a perception of greater weight (presumably because heavier objects must be gripped more tightly to lift them). Ellis and Lederman (1999) reported a material-weight illusion, however, that could not be entirely explained by grip force, because the slipperiest object was not felt to be the heaviest. Moreover, they demonstrated that the effects of material on perceived heaviness vanished when (a) objects of high mass were used, or (b) even low-mass objects were required to be gripped tightly. The first of these effects, an interaction between material and mass, is a version of scale effects in haptic perception to which we previously alluded.

However, cognitive factors cannot be entirely excluded either, as demonstrated by an experiment by Ellis and Lederman (1998) that describes the so-called golf-ball illusion, a newly documented misperception of weight. Experienced golfers and nongolfers were visually shown practice and real golf balls that looked alike, but that were adjusted to be of equal mass. The golfers judged the practice balls to be heavier than the real balls, in contrast to the nongolfers, who judged them to be the same apparent weight. These results highlight the contribution of a cognitive component to weight perception, inasmuch as only experienced golfers would know that practice balls are normally lighter than real golf balls.

Collectively, this body of studies points to a complex set of factors that affect the perception of weight via the haptic system. Resistance to rotation is important, particularly when an object is wielded (as opposed, e.g., to being passively held). Grip force and material may reflect cognitive expectancies (i.e., the expectation that more tightly gripped objects and denser objects should be heavier), but they may also affect more peripheral perceptual mechanisms. A pure cognitive-expectancy explanation for these factors would suggest equivalent effects when vision is used to judge weight, but such effects are not obtained (Ellis & Lederman, 1999). Nor would a pure expectancy explanation explain why the effects of material on weight perception vanish when an object is gripped tightly. Still, a cognitive expectancy explanation does explain the differences in the weight perceptions of the experienced golfers versus the nongolfers. As for lower-level processes that may alter the weight percept, Ellis and Lederman (1999) point out that a firm grip may saturate mechanoreceptors that usually provide information about slip. And Flanagan and Bandomir (2000) have found that weight perception is affected by the width of the grip, the number of fingers involved, and the contact area, but not the angle of the contacted surfaces; these findings suggest the presence of additional complex interactions between weight perception and the motor commands for grasping.

**Curvature**

Curvature is the rate of change in the angle of the tangent line to a curve as the tangent point moves along it. Holding shape constant, curvature decreases as scale increases; for example, a circle with a larger radius has a smaller curvature. Like other haptically perceived properties, the scale of a curve is important. A curved object may be small enough to fall within the area of a fingertip, or large enough to require a movement of the hand across its surface in order to touch it all. If the curvature of a surface is large (e.g., a pearl), then the entire surface may fall within the scale of a fingertip. A surface with a smaller curvature may still be presented to a single finger, but the changes in the tangent line over the width of the fingertip may not make it discriminable from a flat surface.

One clear point is that curvature perception is subject to error from various sources. One is manner of exploration. For example, when curved edges are actively explored, curvature away from the explorer may lead to the perception that the edge is straight (Davidson, 1972; Hunter, 1954). Vogels, Kappers, and Koenderink (1996) found that the curvature of a surface was affected by another surface that had been touched previously, constituting a curvature aftereffect. The apparent curvature of a surface also depends on whether it lies along or across the fingers (Pont, Kappers, & Koenderink, 1998), or whether it touches the palm or upper surface of the hand (Pont, Kappers, & Koenderink, 1997). When small curved surfaces, which have relatively high curvature, are brought to the fingertip, slowly adapting mechanoreceptors provide an isomorphic representation of the pressure gradient on the skin (LaMotte & Srinivasan, 1993; Srinivasan & LaMotte, 1991; Vierck, 1979). This map is sufficient to make discriminations between curved surfaces on the basis of a single finger’s touch. Goodwin, John, and Marceglia (1991) found that a curvature equivalent to a circle with a radius of .2 m could be discriminated from a flat surface when passively touched by a single finger.

When larger surfaces (smaller curvature) are presented, they may be explored by multiple fingers of a static hand or by tracing along the edge. Pont et al. (1997) tested three models to explain curvature perception when static, multifinger exposure was used.

To understand the models, consider a stimulus shaped like a semicircle, the flat edge of which lies on a tabletop with the curved edge pointing up. This situation is illustrated in Figure 6.4. Assume that the stimulus is felt by three fingers, with the middle finger at the highest point (i.e., the midpoint) of the curve. There are then three parameters to consider. The first is height difference: The middle finger is higher (i.e., at a greater distance from the tabletop) than the
Figure 6.4 Definition of three different measures of curvature detectable from touch (Pont et al., 1999; Figure 5, top)—the height difference, the attitude of the fingers, and the radius of curvature. The circles represent three fingers touching a curved surface. Reprinted with permission.

other fingers by some height. The second is the difference in the angles at which the two outer fingers lie: These fingers’ contact points have tangent lines tilted toward one another, with the difference in their slopes constituting an attitude difference, so to speak. In addition, the semicircle has some objective curvature. All three parameters will change as the semicircle’s radius changes size. For example, as the radius increases and the surface gets flatter, the curvature will decrease, the difference in height between the middle and outer fingers will decrease, and the attitudes of the outer fingers approach the horizontal from opposing directions, maximizing the attitude difference. The question is, which of these parameters—height difference, attitude difference, or curvature—determines the discriminability between edges of different curvature? Pont et al. concluded that subjects compared the difference in attitudes between surfaces and used that difference to discriminate them. That is, for each surface, subjects considered the difference in the slope at the outer points of contact. For example, this model predicts that as the outer fingers are placed further apart along a semicircular edge of some radius, the value of the radius at which there is a threshold level of curvature (i.e., where a curved surface can just be discriminated from a flat one) will increase. As the fingers move farther apart, only by increasing the radius of the semicircle can the attitude difference between them be maintained.

As we report in the following section, when a stimulus has an extended contour, moving the fingers along its edge is the only way to extract its shape; static contact does not suffice. For simple curves, at least, it appears that this is not the case, and static and dynamic curvature detection is similar. Pont (1997) reported that when subjects felt a curved edge by moving their index finger along it, from one end to the other of a window of exposure, the results were similar to those with static touch. She again concluded that it was the difference in local attitudes, the changing local gradients touched by the finger as it moved along the exposed edge, that were used for discrimination. A similar conclusion was reached by Pont, Kappers, and Koenderink (1999) in a more extended comparison of static and dynamic touch. It should be noted that the nature of dynamic exploration of the stimulus was highly constrained in these tasks, and that the manner in which a curved surface is touched may affect the resulting percept (Davidson, 1972; Davidson & Whitson, 1974). We now turn to the general topic of how manual exploration affects the extraction of the properties of objects through haptic perception.

Role of Manual Exploration in Perceiving Object Properties

The sensory receptors under the skin, and in muscles, tendons, and joints, become activated not only through contact with an object but through movement. Lederman and Klatzky (1987) noted the stereotypy with which objects are explored when people seek information about particular object properties. For example, when people seek to know which of two objects is rougher, they typically rub their fingers along the objects’ surfaces. Lederman and Klatzky called such an action an “exploratory procedure,” by which they meant a stereotyped pattern of action associated with an object property.

The principal set of exploratory procedures they described is as follows (see Figure 6.5):

Lateral motion—associated with texture encoding; characterized by production of shearing forces between skin and object.
Static contact—associated with temperature encoding; characterized by contact with maximum skin surface and without movement, also without effort to mold to the touched surface.

Enclosure—associated with encoding of volume and coarse shape; characterized by molding to touched surface but without high force.

Pressure—associated with encoding of compliance; characterized by application of forces to object (usually, normal to surface), while counterforces are exerted (by person or external support) to maintain its position.

Unsupported holding—associated with encoding of weight; characterized by holding object away from supporting surface, often with arm movement (hefting).

Contour following—associated with encoding of precise contour; characterized by movement of exploring effector (usually, one or more fingertips) along edge or surface contour.

The association between these exploratory procedures and the properties they are used to extract has been documented in a variety of tasks. One paradigm (Lederman & Klatzky, 1987) required blindfolded participants to pick the best match, among three comparison objects, to a standard object. The match was to be based on a particular property, like roughness, with others being ignored. The hand movements of the participants when exploring the standard object were recorded and classified as exploratory procedures. In another task, blindfolded participants were asked to sort objects into categories defined by haptically perceptible properties, as quickly as possible (Klatzky, Lederman, & Reed, 1989; Lederman, Klatzky, & Reed, 1993; Reed, Lederman, & Klatzky, 1990). The objects were custom fabricated and varied systematically (across several sets) in shape complexity, compliance, size, hardness, and surface roughness. In both of these tasks, subjects were observed to produce the exploratory procedure associated with the targeted object property.

Haptic exploratory procedures are also observed when vision is available, although they occur only for a subset of the properties, and then only when the judgment is relatively difficult (i.e., vision does not suffice). In particular (Klatzky, Lederman, & Matula, 1993), individuals who were asked which of two objects was greater along a designated property—size, weight, and so on—used vision alone to make judgments of size or shape, whether the judgments were easy or difficult. However, they used appropriate haptic exploratory procedures to make difficult judgments of material properties, such as weight and roughness.

One might ask what kind of exploration occurs when people try to identify common objects. Klatzky, Lederman, and Metzger (1985) observed a wide variety of hand movements when participants tried to generate the names of 100 common objects, as each object was placed in their hands in turn. Lederman and Klatzky (1990) probed for the hand movements used in object identification more directly, by placing an object in the hands of a blindfolded participant and asking for its identity with one of two kinds of cues. The cue referred either to the object’s basic-level name (e.g., Is this writing implement a pencil?) or to a name at a subordinate level (e.g., Is this pencil a used pencil?). An initial phase of the experiment determined what property or properties people thought were most critical to identifying the named object at each level; in this phase, a group of participants selected the most diagnostic attributes for each name from a list of properties that was provided. This initial phase revealed that shape was the most frequent diagnostic attribute for identifying objects at the basic level, although texture was often diagnostic as well. At the subordinate level, however, the set of object names was designed to elicit a wider variety of diagnostic attributes; for example, whereas shape is diagnostic to identify a food as a noodle, compliance is important when identifying a noodle as a cooked noodle. In the main phase of the experiment, when participants were given actual exemplars of the named object and probed at the basic or subordinate level, their hand movements were recorded and classified. Most identifications began with a grasp and lift of the object. This initial exploration was often followed by more specific exploratory procedures, and those procedures were the ones that were associated with the object’s most diagnostic attributes.

Why are dedicated exploratory procedures used to extract object properties? Klatzky and Lederman (1999a) argued that each exploratory procedure optimizes the input to an associated property-computation process. For example, the exploratory procedure associated with the property of apparent temperature (i.e., static holding) uses a large hand surface. Spatial summation across the thermal receptors means that a larger surface provides a stronger signal about rate of heat flow. As another example, lateral motion—the scanning procedure associated with the property of surface roughness—has been found to increase the firing rates of slowly adapting receptors (Johnson & Lamb, 1981), which appear to be the input to the computation of roughness for macrotextured surfaces (see Hsiao et al., 1993, for review). (For a more complete analysis of the function of exploratory procedures, see Klatzky & Lederman, 1999a.)

The idea that the exploratory procedure associated with an object property optimizes the extraction of that property is supported by an experiment of Lederman and Klatzky...
(1987, Experiment 2). In this study, participants were constrained to use a particular exploratory procedure while a target property was to be compared. Across conditions, each exploratory procedure was associated with each target property, not just the property with which the procedure spontaneously emerged. The accuracy and speed of the comparison were determined for each combination of procedure and property. When performance on each property was assessed, the optimal exploratory procedure in this forced-choice exploration task (based on accuracy, with speed used to disambiguate ties) was found to be the same one that emerged when subjects freely explored to compare the given property. That is, the spontaneously executed procedure was in fact the best one to use, indicating that the procedure maximizes the availability of relevant information. The use of contour following to determine precise shape was found not only optimal, but also necessary in order to achieve accurate performance.

Turvey and associates, in an extensive series of studies, have examined a form of exploration that they call “dynamic touch,” to contrast it with both cutaneous sensing and haptic exploration, in which the hand actively passes over the surface of an object (for review, see Turvey, 1996; Turvey & Carello, 1995). With dynamic touch, the object is held in the hand and wielded, stimulating receptors in the tendons and muscles; thus it can be considered to be based on kinesthesia. The inertia tensor, described previously in the context of weight perception, has been found to be a mediating construct in the perception of several object properties from wielding. We have seen that the eigenvalues of the inertia tensor—that is, the resistance to rotation around three principal axes (the eigenvectors)—appear to play a critical role in the perception of heaviness. The eigenvalues and eigenvectors also appear to convey information about the geometric properties of objects and the manner in which they are held during wielding, respectively. Among the perceptual judgments that have been found to be directly related to the inertia tensor are the length of a wielded object (Pagano & Turvey, 1993; Solomon & Turvey, 1988), its width (Turvey, Burton, Amazeen, Butwill, & Carello, 1998), and the orientation of the object relative to the hand (Pagano & Turvey, 1992). A wielded object can also be a tool for finding out about the external world; for example, the gap between two opposing surfaces can be probed by a handheld rod (e.g., Barac-Cikoja & Turvey, 1993).

Relative Availability of Object Properties

Lederman and Klatzky (1997) used a variant of a visual search task (Treisman & Gormican, 1988) to investigate which haptically perceived properties become available at different points in the processing stream. In their task, the participant searched for a target that was defined by some haptic property and presented to a single finger, while other fingers were presented with distractors that did not have the target property. For example, the target might be rough, and the distractors smooth. From one to six fingers were stimulated on any trial, by means of a motorized apparatus. The participant indicated target presence or absence by pressing a thumb switch, and the response time—from presentation of the stimuli to the response—was recorded. The principal interest was in the search function; that is, the function relating response time to the number of fingers that were stimulated. Two such functions could be calculated, one for target-present trials and the other for target-absent trials. The functions were generally strongly linear.

Twenty-five variants on this task were performed, representing different properties. The properties fell into four broad classes. One was material properties: for example, rough-smooth (a target could be rough and distractors smooth, or vice versa), hard-soft, and cool-warm (copper vs. pine). A second class required subjects to search for the presence or absence of abrupt surface discontinuities, such as detecting a surface with a raised bar among flat surfaces. A third class of discriminations was based on planar or three-dimensional spatial position. For example, subjects might be asked to search for a vertical edge (i.e., a raised bar aligned along the finger) among horizontal-edge distractors, or they might look for a raised dot to the right of an indentation among surfaces with a dot to the left of an indentation (Experiments 8–11). Finally, the fourth class of searches required subjects to discriminate between continuous three-dimensional contours, such as seeking a curved surface among flat surfaces.

From the resulting response-time functions, the slope and intercept parameters were extracted. The slope indicates the additional cost, in terms of processing time, of adding a single finger to the display. The intercept includes one-time processes that do not depend on the number of fingers, such as adjusting the orientation of the hand so as to better contact the display. Note that although the processes entering the intercept do not depend on the number of fingers, they may depend on the particular property that is being discriminated. The intercept will include the time to extract information about the object property being interrogated, to the extent the process of information extraction is done in parallel and it does not use distributed capacity across the fingers (in which case, the processing time would affect the slope).

The relative values of the slope and intercept indicate the availability ordering among properties. A property whose discrimination produces a higher slope extracts a higher
finger-by-finger cost and hence is slower to extract; a property producing a higher intercept takes longer for one-time processing and hence is slow to be extracted. Both the slopes and intercepts of this task told a common story about the relative availability among haptically accessible properties. There was a progression in availability from material properties, to surface discontinuities, to spatial relations. The slopes for material properties tended to be low (≤36 ms), and several were approximately equal to zero. Similarly, the intercepts of material-property search functions tended to be among the lowest, except for the task in which the target was cool (copper) and the distractors warm (pine). This exception presumably reflects the time necessary for heat to flow from the subject’s skin to the stimulus, activating the thermoreceptors. In contrast, the slopes and intercepts for spatially defined properties tended to be among the highest.

Why should material properties and abrupt spatial discontinuities be more available than properties that are spatially defined? Lederman and Klatzky (1997) characterized the material and discontinuity properties as unidimensional or intensive: That is, they can be represented by a scalar magnitude that indicates the intensity of the perceptual response. In contrast, spatial properties are, by definition, related to the two- or three-dimensional layout of points in a reference system. A spatial discrimination task requires that a distinction be made between stimuli that are equal in intensity but vary in spatial placement. For example, a bar can be aligned with or across the fingertip, but exerts the same amount of pressure in either case.

The relative unavailability of spatial properties demonstrated in this research is consistent with a more general body of work suggesting that spatial information is relatively difficult to extract by the haptic system, in comparison both to spatial coding by the visual system and to haptic coding of non-spatial properties (e.g., Cashdan, 1968; Johnson & Phillips, 1981; Lederman, Klatzky, Chataway, & Summers, 1990).

A variety of studies have established that the perception of manipulatory space is nonveridical. The distortions have been characterized in various ways. One approach is to attempt to determine a distance metric for lengths of movements made on a reached surface. Brambring (1976) had blind and sighted individuals reach along two sides of a right triangle and estimate the length of the hypotenuse. Fitting the hypotenuse to a general distance metric revealed that estimates departed from the Euclidean value by using an exponent less than 2. Brambring concluded that the operative metric was closer to a city block. Subsequent work suggests, however, that no one metric will apply to haptic spatial perception, because distortions arise from several sources, and perception is not uniform over the explored space; that is, haptic spatial perception is anisotropic.

One of the indications of anisotropy is the vertical-horizontal illusion. Well known in vision, although observed long ago in touch as well (e.g., Burtt, 1917), this illusion takes the form of vertical lines’ being overestimated relative to length-matched horizontals. Typically, the illusion is tested by presenting subjects with a T-shaped or L-shaped form and asking them to match the lengths of the components. The T-shaped stimulus introduces another source of judgment error, however, in that the vertical line is bisected (making it perceptually shorter) and the horizontal is not. The illusion in touch is not necessarily due to visual mediation (i.e., imagining how the stimulus would look), because it has been observed in congenitally blind people as well as sighted individuals (e.g., Casla, Blanco, & Travieso, 1999; Heller & Joyner, 1993). Heller, Calcaterra, Burson, & Green (1997) demonstrated that the patterns of arm movement used by subjects had a substantial effect on the illusion. Use of the whole arm in particular augmented the magnitude of the illusion. Millar and Al-Attar (2000) found that the illusion was affected by the position of the display relative to the body, which would affect movement and, potentially, the spatial reference system in which the display was represented.

Another anisotropy is revealed by the radial-tangential effect in touch. This refers to the fact that movements directed toward and away from the body (radial motions) are overestimated relative to side-to-side (tangential) motions of equal extent (e.g., Cheng, 1968; Marchetti & Lederman, 1983). Like the vertical-horizontal illusion, this appears to be heavily influenced by motor patterns. The perception of distance is greater when the hand is near the body, for example (Cheng, 1968; Marchetti & Lederman, 1983). Wong (1977) found that the slower the movement, the greater the judged extent; he suggested that the difference between radial and tangential distance judgments may reflect different execution times. Indeed, when Armstrong and Marks (1999) controlled
for movement duration, the difference between estimates of radial and tangential extents vanished.

A third manifestation of anisotropy in haptic space perception is the oblique effect, also found in visual perception (e.g., Appelle & Countryman, 1986; Gentaz & Hatwell, 1995, 1996, 1998; Lechelt, Eliuk, & Tanne, 1976). When people are asked to reproduce the orientation of a felt rod, they do worse with obliques (e.g., 45°) than with horizontal or vertical lines. As with the other anisotropies that have been described, the pattern in which the stimulus is explored appears to be critical to the effect. Gentaz and Hatwell (1996) had subjects reproduce the orientation of a rod when the gravitational force was either natural or nulled by a counterweight. The oblique effect was greater when the natural gravitational forces were present. In a subsequent experiment with blind subjects (Gentaz & Hatwell, 1998), it appeared that the variability of the gravitational forces, rather than their magnitude, was critical: The oblique effect was not found in the horizontal plane, even with an unsupported arm; in this plane the gravitational forces do not vary with the direction of movement. In contrast, the oblique effect was found in the frontal plane, where gravitational force impedes upward and facilitates downward movements, regardless of arm support.

A study by Essock, Krebs, and Prather (1997) points to the fact that anisotropies may have multiple processing loci. Although effects of movement and gravity point to the involvement of muscle-tendon-joint systems, the oblique effect was also found for gratings oriented on the finger pad. This is presumably due to the filtering of the cutaneous system. The authors suggest a basic distinction between low-level anisotropies that arise at a sensory level, and ones that arise from higher-level processing of spatial relations.

The influence of high-level processes can be seen in a phenomenon described by Lederman, Klatzky, and Barber (1985), which they called “length distortion.” In their studies, participants were asked to trace a curved line between two endpoints, and then to estimate the direct (Euclidean) distance between them. The estimates increased directly with the length of the curved line, in some cases amounting to a 2:1 estimate relative to the correct value. High errors were maintained, even when subjects kept one finger on the starting point of their exploration and maintained it until they came to the endpoint. Under these circumstances, they had simultaneous sensory information about the positions of the fingers before making the judgment; still, they were pulled off by the length of the exploratory path. Because the indirect path between endpoints adds to both the extent and duration of the travel between them by the fingers, Lederman et al. (1987) attempted to disambiguate these factors by having subjects vary movement speed. They found that although the duration of the movement affected responses, the principal factor was the pathway extent. In short, it appears that the spatial pattern of irrelevant movement is taken into account when the shortest path is estimated.

Bingham, Zaal, Robin, and Shull (2000) suggested that haptic distortion might actually be functional: namely, as a means of compensating for visual distortion in reaching. They pointed out that although visual distances are distorted by appearing greater in depth than in width, the same appears to be true of haptically perceived space (Kay, Hogan, & Fasse, 1996). Given an error in vision, then, the analogous error in touch leads the person to the same point in space. Suppose that someone reaching to a target under visual guidance perceives it to be 25% further away than it is—for example, at 1.25 m rather than its true location of 1 m. If the haptic system also feels it to be 25% further away than it is, then haptic feedback from reaching will guide a person to land successfully on the target at 1 m while thinking it is at 1.25 m. However, the hypothesis that haptic distortions usefully cancel the effects of visual distortions was not well supported. Haptic feedback in the form of touching the target after the reach compensated to some extent, but not fully, for the visual distortion.

_Virtually all of the anisotropies that have been described are affected by the motor patterns used to explore haptic space. The use of either the hand or arm, the position of the arm when the hand explores, the gravitational forces present, and the speed of movement, for example, are all factors that have been identified as influencing the perception of a tangible layout in space. What is clearly needed is research that clarifies the processes by which a representation of external space is derived from sensory signals provided by muscle-tendon-joint receptors, which in turn arise from the kinematics (positional change of limbs and effectors) and dynamics (applied forces) of exploration. This is clearly a multidimensional problem. Although it may turn out to have a reduced-dimensional solution, the solution seems likely to be relatively complex, given the evidence that high-level cognitive processes mediate the linkages between motor exploration, cutaneous and kinesthetic sensory responses, and spatial representation._

**HAPTIC PERCEPTION OF TWO- AND THREE-DIMENSIONAL PATTERNS**

Pattern perception in the domain of vision is presented in the chapter by Stephen in this volume. Perception of pattern by the haptic system has been tested within a number of stimulus domains. The most common stimuli are vibrotactile patterns, presented by vibrating pins. Other two-dimensional patterns that have been studied are Braille, letters, unfamiliar outlines, and outline drawings of common objects. There is also work on fully three-dimensional objects.
Vibrotactile Patterns

A vibrotactile pattern is formed by repeatedly stimulating some part of the body (usually the finger) at a set of contact points. Typically, the points are a subset of the elements in a matrix. The most commonly used stimulator, the Optacon (for optical-to-tactile converter), is a array with 24 rows and 6 columns; it measures 12.7 * 29.2 mm (Cholewiak & Collins, 1990). The row vibrators are separated by approximately 1.25 mm and the column pins by approximately 2.5 mm. The pins vibrate approximately 230 times per second. Larger arrays were described by Cholewiak and Sherrick (1981) for use on the thigh and the palm.

A substantial body of research has examined the effects of temporal and spatial variation on pattern perception with vibrating pin arrays (see Craig & Rollman, 1999; Loomis & Lederman, 1986). When two temporally separated patterns are presented, they may sum to form a composite, or they may produce two competing responses; these mechanisms of temporal interaction appear to be distinct (Craig, 1996; Craig & Qian 1997). These temporal effects can occur even when the patterns are presented to spatial locations on two different fingers (Craig & Qian, 1997).

Spatial interactions between vibratory patterns may occur because the patterns stimulate common areas of skin, or because they involve a common stimulus identity but are not necessarily at the same skin locus. The term communality (Geldard & Sherrick, 1965) has been used to measure the extent to which two patterns have active stimulators in the same spatial location, whether the pattern identities are the same or different. The ability to discriminate patterns has been found to be inversely related to their communality at the finger, palm, and thigh (Cholewiak & Collins, 1995; see that paper also for a review). The extent to which two patterns occupy common skin sites has also been found to affect discrimination performance. Horner (1995) found that when subjects were asked to make same-different judgments of vibrotactile patterns, irrespective of the area of skin that was stimulated, they performed best when the patterns were presented to the same site, in which case the absolute location of the stimulation could be used for discrimination. As the locations were more widely separated, performance deteriorated, suggesting a cost for aligning the patterns within a common representation when they were physically separated in space.

Two-Dimensional Patterns and Freestanding Forms

Another type of pattern that has been used in a variety of studies is composed of raised lines or points. Braille constitutes the latter type of pattern. Loomis (1990) modeled the perception of characters presented to the fingertip—not only Braille patterns, but also modified Braille with adjacent connected dots, raised letters of English and Japanese, and geometric forms. Confusion errors in identifying members of these pattern sets, tactually and visually when seen behind a blurring filter (to simulate filtering properties of the skin), were compiled. The data supported a model in which the finger acts like a low-pass filter, essentially blurring the input; the intensity is also compressed. Loomis has pointed out that given the filtering imposed by the skin, the Braille patterns that have been devised for use by the blind represent a useful compromise between the spatial extent of the finger and its acuity: A larger pattern would have points whose relative locations were easier to determine, but it would then extend beyond the fingertip.

The neurophysiological mechanisms underlying perception of raised, two-dimensional patterns at the fingertip have been investigated by Hsaio, Johnson, and associates (see Hsiao, Johnson, Twombly, & DiCarlo, 1996). The SAI mechanoreceptors appear to be principally involved in form perception. These receptors have small receptive fields (about 2 mm diameter), respond better to edges than to continuous surfaces (Phillips & Johnson, 1981), and given their sustained response, collectively produce an output that preserves the shape of embossed patterns presented to the skin. Hsaio et al. (1996) have traced the processing beyond the S1 mechanoreceptors to cortical areas S1 and SII in succession. Isomorphism is preserved in area S1, whereas SII neurons have larger receptive fields and show more complex responses that are not consistently related to the attributes of the stimulus.

Larger two-dimensional shapes, felt with the fingers of one or more hands, have also been used to test the pattern-recognition capabilities of the haptic system. These larger stimuli introduce demands of memory and integration (see following paragraphs), and often, performance is poor. Klatzky, Lederman, and Balakrishnan (1991) found chance performance in a successive matching task with irregularly shaped planar forms (like wafers) on the order of 15 cm in diameter. Strategic exploration may be used to reduce the -memory demands and detect higher-order properties of such stimuli. Klatzky et al. found that subjects explored as symmetrically as possible, often halting exploration with one hand so that the other, slowed by a more complex contour, could catch up, so to speak, to the same height in space. Ballesteros, Manga, and Reales (1997) and Ballesteros, Millar, and Reales (1998) found that such bimanual exploration facilitated the ability to detect the property of symmetry in raised-line shapes scaled well beyond the fingertip.

Two-Dimensional Outline Drawings of Common Objects

If unfamiliar forms that require exploration beyond the fingertip are difficult to identify and compare, one might
expect better performance with familiar objects. Studies that examine object-identification performance with raised, two-dimensional depictions of objects have led to the conclusion that performance is considerably below that with real objects (see following discussion), but well above chance. Lederman et al. (1990) found that sighted individuals recognized only 34% of raised-line drawings of objects, even when they were allowed up to 2 minutes of exploration. The blind participants did substantially worse (10% success). Loomis, Klatzky, and Lederman (1991) implicated memory and integration processes as limiting factors in two-dimensional haptic picture recognition. This study compared visual and tactual recognition with identical line drawings of objects. In one condition with visual presentation, the contours of the object were revealed through an aperture scaled to have the same proportion, relative to the size of the object, as the fingertip. As the participant moved his or her hand or a single finger, which reduced the objects’ information content primarily to three-dimensional contour (although some surface information, such as coefficient of friction, was available). Performance was approximately 75% accurate, well above the level achieved when exploring raised-line depictions of the same objects.

Lakatos and Marks (1999) investigated whether, when individuals explore three-dimensional objects, they emphasize the local features or the global form. The task was to make similarity judgments of unfamiliar geometric forms (e.g., cube; column) that contained distinctive local features such as grooves and spikes (see Figure 6.6). The data suggested a greater salience for local features in early processing, with global features becoming more equal in salience as processing time increased. Objects with different local features but similar in overall shape were judged less similar when explored variation in the plane as a three-dimensional spatial entity. This ability is consistent with demonstrations that blind people often create drawings that illustrate pictorial conventions such as perspective and metaphorical indications of movement (Heller, Calcaterra, Tyler, & Burson, 1996; Kennedy, 1997).

Three-Dimensional Objects

Real, common objects are recognized very well by touch. Klatzky et al. (1985) found essentially perfect performance in naming common objects placed in the hands, with a modal response time of 2 s. This level of performance contrasts with the corresponding data for raised-line portrayals of common objects (i.e., low accuracy even with 2 minutes of exploration), raising the question as to what is responsible for the difference. No doubt there are several factors. Experience is likely to be one; note that experience is implicated in previously described studies with raised-line objects.

Another relevant factor is three-dimensionality. A two-dimensional object follows a convention of projecting variations in depth to a picture plane, from which the third dimension must be constructed. This is performed automatically by visual processes, but not, apparently, in the domain of touch. Lederman et al. (1990) found that portrayals of objects that have variations in depth led to lower performance than was found with flat objects that primarily varied in two dimensions (e.g., a bowl vs. a fork). Shimizu, Saida, and Shimura (1993) used a pin-element display to portray objects as two-dimensional outlines or three-dimensional relief forms. Ratings of haptic legibility were higher for the three-dimensional objects, and their identification by early blind individuals was also higher. Klatzky, Loomis, Lederman, Wake, and Fujita (1993) asked participants to identify real objects while wearing heavy gloves and exploring with only a single finger, which reduced the objects’ information content primarily to three-dimensional contour (although some surface information, such as coefficient of friction, was available). Performance was approximately 75% accurate, well above the level achieved when exploring raised-line depictions of the same objects.

Lakatos and Marks (1999) investigated whether, when individuals explore three-dimensional objects, they emphasize the local features or the global form. The task was to make similarity judgments of unfamiliar geometric forms (e.g., cube; column) that contained distinctive local features such as grooves and spikes (see Figure 6.6). The data suggested a greater salience for local features in early processing, with global features becoming more equal in salience as processing time increased. Objects with different local features but similar in overall shape were judged less similar when explored...
haptically than when vision was available. Longer exposure time (increasing from 1 s to 16 s) produced greater similarity ratings for objects that were locally different but globally similar, indicating the increasing salience for global shape over time.

When people do extract local features of three-dimensional objects, they appear to have a bias toward encoding the back of the object—the reverse of vision. Newell, Ernst, Tian, and Bulthoff (2001) documented this phenomenon using objects made of Lego blocks. The participants viewed or haptically explored the objects, and then tried to recognize the ones to which they had been exposed. On some trials, the objects were rotated 180° (back-to-front) between exposure and the recognition test. When exposure and test were in the same modality (vision or touch), performance suffered if the objects were rotated. When the modality changed between exposure and test, however, performance was better when the objects were rotated as well: In this case, the surface that was felt at the back of the object was viewed at the front. Moreover, when exploration and testing were exclusively by touch, performance was better for objects explored from the back than for those explored from the front.

Although the previously described studies emphasized the role of shape, no doubt a critical factor in recognizing real, common objects by touch is material. Material is locally available, whereas extraction of the shape of an object requires following its contours or enclosing it in the hand (Lederman & Klatzky, 1987). A number of studies by Klatzky, Lederman, and associates point to the importance of
material properties in identification and similarity judgments (Klatzky & Lederman, 2000). Klatzky, et al. (1985) observed that individuals who were freely identifying common objects often reported attending to the objects’ material properties. Klatzky, Loomis, Lederman, Wake, and Fujita (1993) found that the performance of individuals who explored common objects with a single finger while wearing a heavy glove improved significantly when the tip of the glove was cut off to expose the object’s material. Lederman and Klatzky (1990) found that when an object’s identity was particularly revealed by its material (e.g., as compliance is diagnostic of a cooked noodle), people attempting to identify the object executed the exploratory procedure that was associated with the relevant material property (e.g., to identify the cooked noodle, pressing on it). And Klatzky and Lederman (1995) found that a 200-ms touch with the fingertip was sufficient to identify 25% of a set of objects selected to have large surfaces and to be particularly identifiable by texture (e.g., sandpaper).

VISUAL-HAPTIC INTERACTIONS

Attention

The chapter in this volume by Egeth and Lamy provides a general overview of attention. Vision and touch have been shown to be linked in attentional processing. Spence, Pavani, and Driver (2000) used a paradigm based on early work of Posner (1978) to demonstrate cross-modal interactions in endogenous (self-directed rather than stimulus-driven) spatial attention. Subjects discriminated between sustained or pulsed targets that were presented either visually (by a light) or tactually (by a force to the fingertip) on the right or left side of the body. They indicated the target’s form (sustained vs. pulsed), not its spatial location, with a foot pedal. A centralized visual pre-cue, a right or left arrow, correctly predicted the target location on 80% of trials, which should trigger a voluntary orienting of attention to the precued location. Both visual and tactual pre-cueing effects were obtained, in the form of facilitation when the cue was valid (correctly predicted the target location). In a subsequent experiment, shown in Figure 6.7, in each hand the participant held a foam cube, which could produce either vibrotactile stimulation or a light on the upper or lower edge. The response was to indicate elevation of the target (upper vs. lower edge), regardless of the cube on which it appeared or of its modality. Again, a central arrow cue, predicting the likely hand to be stimulated, was facilitative when it was valid. These experiments indicated that spatial attention could be endogenously directed in the visual or tactual modality.

Just as a visual cue can direct attention to a tactile stimulus, incongruent visual stimulation can interfere with tactile detection. Pavani, Spence, and Driver (2000) asked individuals to indicate the location of a tactile stimulus on the hand while it held a cube underneath a table (i.e., the hand could not be seen). Simultaneously with the tactile stimulus, a light could flash on a visible cube located on the table top. When the light flashed at one location on the cube while the tactile stimulus occurred at another location, tactile detection was slowed. This interference from an incongruent visual stimulus increased when the participants saw rubber hands holding the visible cubes on top of the table, aligned with their own hands. Moreover, some participants reported feeling that the rubber hands were their own!

Another study of Spence et al. (2000) specifically tested cross-modal cueing of attention: The target appeared in one of the two modalities (the so-called primary modality) on 73% of trials, and participants were instructed to direct their attention primarily in that modality and not in the other (the so-called secondary modality). A critical manipulation was that the cue indicating the likely spatial location of the target within the primary modality was incorrect (actually reversed by a ratio of 2:1) for the secondary one. For example, if touch was primary—that is, a tactile stimulus occurred most
often—a cue predicting a right-side tactile stimulus actually implied that if a visual stimulus occurred instead, it would be on the left side. Thus, participants had countermotivation to direct their attention in the primary and secondary modalities to the same side. The results indicated that individuals responded faster when the target appeared on the side cued within the primary modality. This occurred not only for stimuli in the primary modality, in which the cue’s prediction was valid, but also for the secondary modality, in which the cue was invalid most of the time—although the primary cueing effect was stronger. Thus, for example, when touch was primary and the cue indicated a right-side stimulus, a visual stimulus on the right was responded to faster than on the left—even though a right-side cue for touch implied a left-side stimulus for vision. On the whole, the results indicate that subjects did not have two dissociated attentional mechanisms that could be directed to opposite areas of space. Rather, the attentional focus directed by the primary modality applied to both modalities.

Cortical sites that may underlie these early attentional interactions between vision and touch were identified by Macaluso, Frith, and Driver (2000). They began with the observation that a touch on one hand can improve visual discrimination in nearby locations (e.g., Butter, Buchtel, & Santucci, 1989). Functional MRI was used while subjects were presented with visual stimulation alone, or visual-plus-tactile stimulation on the same or different sides. When tactile stimulation occurred on the same side as visual, there was elevated activity in the visual cortex. Visual-plus-tactile stimulation on opposite sides did not produce such an elevated response. The authors suggested that this influence of touch on early visual processing arises from pathways that arise in the parietal lobe and project backward.

Cross-Modal Integration

Visual-haptic interactions have been investigated at higher levels of stimulus processing, in which sensory inputs produce a unitary perceptual response. A common paradigm in this research uses a discrepancy between visual and haptic stimuli—sizes or textures, for example—to determine the relative weighting of the modalities under different conditions. In early work, Rock (Rock & Harris, 1967; Rock & Victor, 1964) reported total dominance of haptic perceptions by visual inputs, when participants judged the size of a square that was simultaneously felt and viewed through a reducing lens. However, subsequent research has challenged the early claim of strong visual dominance. Friedes (1974) and Welch and Warren (1980) have argued that a better predictor of relative weighting of modality pairs (e.g., vision-touch, touch-audition, vision-audition) is the relative appropriateness (i.e., defined in terms of accuracy, precision, and cue availability) of the task for each modality. More recently, Heller, Calcatera, Green, and Brown (1999) showed that the modality and precision of the response strongly influenced the weighting of the input stimuli. When subjects responded by viewing a ruler, vision dominated, whereas when they indicated size with a pinch posture, touch dominated. This suggests that the relative contributions of the modalities can be modulated by attention.

A response by age interaction was found in a size-discrepancy study by Misceo, Hershberger, and Mancini (1999). Children from 6 to 12 years of age matched a viewed and touched square to a set of comparison squares that were either felt or viewed. While visual dominance was found across age groups with the visual response, the haptic response led to an age progression from visual to haptic dominance. Thus it appears that experience, maturation, or both alter the extent to which the haptic input can be weighted.

Cognitive factors were also identified in a texture-discrepancy study by Lederman, Thorne, and Jones (1986). One group of subjects was asked to judge the so-called spatial density of a set of textured surfaces by vision, by touch, and by vision and touch together. A second group was asked to judge the same stimuli in terms of roughness, once again by vision, touch, and vision and touch together. The spatial-density instructions produced strong dominance of vision over touch, presumably because fine spatial resolution is required by the task, something that vision does considerably better than touch. In contrast, the roughness instructions produced equally strong tactual dominance over vision; this time it was argued because the sense of touch can differentiate fine differences in surface roughness better than vision can (Heller, 1989b).

Further work on visual-haptic interactions is related to representations in memory. A particularly important issue is whether the two channels converge on a common representation. Memory is reviewed in the next section.

HAPTIC MEMORY

Chapters in this volume that provide broad coverage of human memory are those by Nairne; McNamara and Holbrook; Roediger and Marsh; and Johnson. The literature in this area has tended to neglect the haptic modality, being dominated by verbal stimuli in the auditory and visual modalities. In particular, there has been little effort to build an information-processing systems approach that would identify, for example, sensory stores and different forms of long-term memory.
Research on memory within the modality of touch is also plagued by the possibility, even the likelihood, of mediating representations in the verbal or spatial domains. In an informative review, Millar (1999) has summarized a substantial body of research within the memory-systems approach. She points to evidence for the existence of short-term memory in the tactual modality with a limiting span of two to three items (Millar, 1975a; Watkins & Watkins, 1974). A counterpart to the very short-term iconic and echoic memories, found in vision and audition, has not been clearly demonstrated.

One of the general issues in memory research is the nature of the internal representation. When information is encountered through the sense of touch, one version of this question is whether the representation is intrinsic to the modality or whether it is more general (e.g., spatial). There is evidence for specifically tactual coding during early learning of small patterns like Braille forms; that is, coding that is in terms of tactual features such as texture or dot density, rather than being spatially mediated (see Millar, 1997). Millar (1999) suggested that when patterns can be organized within spatial reference frames, memory for touched patterns is further aided by spatial coding.

Another issue is whether the representation resulting from touch is cross-modal, in the sense of being accessible by other modalities—especially vision. The answer has been demonstrated to be cross-modal. Specifically, haptically presented patterns can be subsequently recognized through the visual modality, although the effect is regulated by a number of factors such as discriminability (see Millar, 1997). Millar (1999) suggested that when patterns can be organized within spatial reference frames, memory for touched patterns is further aided by spatial coding.

In a study with 5-year-olds, Bushnell and Baxt (1999) demonstrated that the children were virtually error-free at discriminating between previously presented and newly presented common objects, whether the modality changed between vision and touch or was held constant between presentation and test. Cross-modal recognition became less accurate (although still above chance levels) when the objects were unfamiliar, or when the old and new objects were different tokens of the same category name. The authors suggest that these decrements due to unfamiliarity and categorical similarity arise from different sources. The categorical effect is likely to be due to mediation at a conceptual level or explicit naming, which children were observed to do. Use of the same name for old and new objects would lead to misrecognition.

On the other hand, the decrement due to using unfamiliar objects is thought to depend on the use of a perceptual code, which emphasizes different aspects of the objects under vision and touch. Such a representation is suggested by experiments on haptic object categorization, which indicate that people use different attributes to group objects, depending on whether vision is available and on whether the participants are instructed to think about what the objects feel like versus what they look like (Klatzky, Lederman, & Reed, 1987; Lederman, Summers, & Klatzky, 1996). Other research suggests that age as well as modality affects the relative emphasis of haptically accessible attributes in object categorization (Schwarzer, Kuefer, & Wilkening, 1999).

A major distinction in memory systems that has emerged in the past two decades or so is made between implicit and explicit memory. Explicit memory is indicated by conscious recollection or recognition: that is, by knowledge that memory is being tapped. Implicit memory is indicated by priming—a change in the performance of some task, due to prior exposure to the task materials. For example, having studied a list of words, participants may be asked to generate completions for three-letter word stems; they tend to generate more completions that match the studied words than would be expected by chance, regardless of whether they explicitly remember those words.

This paradigm has been extended to the haptic modality in several studies. Srinivas, Greene, and Easton (1997a) investigated the effects of elaborative (meaningful) processing on an implicit and explicit memory test with two-dimensional forms. In verbal learning studies, elaborative processing generally leads to better performance on explicit tests of memory, but not on implicit tests. In the Srinivas et al. experiment, participants studied the forms by feeling them and verbally describing their features. They then went on to do elaborative encoding: generating a function for the form (e.g., coat hanger)—or shallow encoding: reporting the number of horizontal and vertical lines in the form. When tested, participants either recognized whether a form was studied or new (i.e., an explicit test), or they drew the form as accurately as possible after 4 seconds of study (i.e., an implicit test). The nature of encoding, whether elaborative or shallow, substantially affected the explicit test but not the implicit test. This indicates that implicit memory extends to the haptic modality (see also Easton, Srinivas, & Greene, 1997).

A subsequent experiment (Srinivas, Greene, and Easton, 1997b) showed that both the explicit and implicit tactual memory tests were affected by changes in the orientation and size of the forms between study and test. Indeed, when the forms were left-right reversed or rescaled, the priming produced by implicit memory vanished. In contrast, a visual version of the test was affected by orientation changes but not size changes; this suggests that the basis for implicit memory in touch is not identical to that in vision, and that the functional representation in touch preserves the physical structure and scale of the touched object.

Cross-modal priming between the visual and haptic modalities has also been of interest. Such priming would
Aids for the Visually Impaired

Printed media are an unquestioned aspect of life for sighted individuals; reading text and viewing images like maps and pictures are taken for granted. Efforts to provide tactual substitutes for text can be traced to the eighteenth century (see Millar, in press). Interest in maps for blind individuals has lagged considerably; it is noteworthy that the first international conference on maps for the blind was held only in the 1980s (Wiedel, 1983).

Millar (in press) pointed to the need for understanding basic processes in haptic perception in order to understand how advanced Braille readers succeed and to apply this understanding to Braille education. She emphasized the inaccuracy of the naive assumption that Braille patterns constitute gestalt, or wholistic, shapes that are read character by character. On the contrary, detailed observation indicates that skilled Braille reading involves interactive scanning by the two hands, which share higher-order goals. One goal is to maintain spatial orientation on the lines of text, and the other is to extract verbal content. Typically, an advanced reader will alternate these functions over the two hands, with one hand starting to find the next line of text while the other finishes the extraction of meaning from the preceding one. This scanning process is moderated by the task goal: for example, to read for meaning or to find a target letter. In order to learn Braille, then, students must master not only the decoding of individual letters, but also the monitoring and controlling of their orientation relative to the text, as well as maintaining a smooth scan.

In designing letters or graphics for the blind, the nature of the pattern is critical. With respect to letters, legibility is the principal issue. The Braille cell uses dot separations that are well within the discrimination of a typical fingerpad, although the dots may be too dense for people with lowered cutaneous acuity, like elderly or diabetic persons. The inventor of Braille designed dotted patterns in preference to embossed continuous letters, with which he had been taught. The punctate nature of Braille dots has been found preferable for matching characters (see Millar, in press). However, Loomis (1990) reported that sighted, blindfolded individuals identified continuous versions of Braille patterns as well as the original dots.

When it comes to graphical aids for the blind other than printed characters, such as icons used on maps, many factors in addition to legibility are important. In early work in this area, Heath (1958) tested a variety of symbolic patterns for discriminability, and various groups have made recommendations for the symbol system of tangible graphics on this basis (Jansson, 1972; Nolan & Morris, 1971; Edman, 1992). Another consideration is function. Golle (1991) suggested that the blind traveler would find it particularly useful to have strip maps providing navigable routes between landmarks that are reoriented relative to the user’s current perspective, rather than survey maps that convey the relative positions of the landmarks in a fixed spatial reference system. Lederman and Campbell (1982) found that the relative effectiveness of a format for raised graphs depended on the use to which the graph was put. When the ordinate value for a given abscissa value had to be determined, a raised grid

APPLICATIONS OF RESEARCH ON TOUCH

Applications of experimental psychology are the topic of the chapter in this volume by Nickerson and Pew. Work on the sense of touch can be applied in many areas. A long-standing application has been to human factors, for the design of handles or knobs. Work on vibrotactile stimulation has led to development of reading aids for blind persons, like the Optacon, and speech-augmentation devices for deaf persons. Increasingly, computer-driven force stimulators have led to applications in the form of virtual reality and teleoperation. Understanding of the basic capacities and information-processing mechanisms of the haptic perceptual system is highly useful, if not necessary, for developing successful applications in these areas. Conversely, the need for application has motivated basic research.

Aids for the Visually Impaired

Printed media are an unquestioned aspect of life for sighted individuals; reading text and viewing images like maps and pictures are taken for granted. Efforts to provide tactual substitutes for text can be traced to the eighteenth century (see Millar, in press). Interest in maps for blind individuals has lagged considerably; it is noteworthy that the first international conference on maps for the blind was held only in the 1980s (Wiedel, 1983).

Millar (in press) pointed to the need for understanding basic processes in haptic perception in order to understand how advanced Braille readers succeed and to apply this understanding to Braille education. She emphasized the inaccuracy of the naive assumption that Braille patterns constitute gestalt, or wholistic, shapes that are read character by character. On the contrary, detailed observation indicates that skilled Braille reading involves interactive scanning by the two hands, which share higher-order goals. One goal is to maintain spatial orientation on the lines of text, and the other is to extract verbal content. Typically, an advanced reader will alternate these functions over the two hands, with one hand starting to find the next line of text while the other finishes the extraction of meaning from the preceding one. This scanning process is moderated by the task goal: for example, to read for meaning or to find a target letter. In order to learn Braille, then, students must master not only the decoding of individual letters, but also the monitoring and controlling of their orientation relative to the text, as well as maintaining a smooth scan.

In designing letters or graphics for the blind, the nature of the pattern is critical. With respect to letters, legibility is the principal issue. The Braille cell uses dot separations that are well within the discrimination of a typical fingerpad, although the dots may be too dense for people with lowered cutaneous acuity, like elderly or diabetic persons. The inventor of Braille designed dotted patterns in preference to embossed continuous letters, with which he had been taught. The punctate nature of Braille dots has been found preferable for matching characters (see Millar, in press). However, Loomis (1990) reported that sighted, blindfolded individuals identified continuous versions of Braille patterns as well as the original dots.

When it comes to graphical aids for the blind other than printed characters, such as icons used on maps, many factors in addition to legibility are important. In early work in this area, Heath (1958) tested a variety of symbolic patterns for discriminability, and various groups have made recommendations for the symbol system of tangible graphics on this basis (Jansson, 1972; Nolan & Morris, 1971; Edman, 1992). Another consideration is function. Golle (1991) suggested that the blind traveler would find it particularly useful to have strip maps providing navigable routes between landmarks that are reoriented relative to the user’s current perspective, rather than survey maps that convey the relative positions of the landmarks in a fixed spatial reference system. Lederman and Campbell (1982) found that the relative effectiveness of a format for raised graphs depended on the use to which the graph was put. When the ordinate value for a given abscissa value had to be determined, a raised grid
aided performance, but the same grid was confusing when subjects had to find the coordinates of an intersection. Patterns of scanning also affect the utility of a display (Berlá, 1982).

Basic research on haptic perception indicates that the iconic value of a symbol may be different in touch than in vision. For example, the limited spatial resolution of the fingertip may make it difficult to determine the direction of a raised arrow, whereas a pattern that is rough in one direction and smooth in the other can be read quickly as a directional signal (Schiff, Kaufer, & Mosák, 1966). Research by Lambert and Lederman (1989) extended this notion by designing raised point symbols for designating interior landmarks of buildings, in which the three-dimensional structure inherently suggested the intended meaning of the symbol.

Technological advances have permitted increasingly sophisticated aids for blind persons. Touch tablets are an electronic means of displaying maps to visually impaired individuals. The display is divided into a matrix, and when a cell is touched, a synthesized speech message is invoked. This is the basis for the NOMAD system (Parkes, 1988). Another system, MoBIC (The MoBIC Consortium, 1997), combines a computer-based map, global positioning sensing, and speech synthesis. It can be used to preview and plan a journey (virtual travel), or it can be consulted en route by means of a keypad in order to get information about current position and a travel plan. The initial system did not include a tactual map. In an experimental test, however, Bringhammar, Jansson, and Douglas (1997) found that after planning and walking with the system, visually impaired participants had high ratings for usability and satisfaction, but augmenting the system with a tactual map increased satisfaction ratings and measures of route understanding.

The development of aids for the blind will undoubtedly benefit from a two-pronged approach, in which applied research is coupled with work on the basic capabilities of the haptic system. Sensory limitations such as spatial thresholds are important, but so are many higher-level factors. In order to develop effective displays for people without vision, one must deal with issues such as what properties of stimuli are available, how these properties emerge in perceptual processing, how exploration alters what is encoded, and how haptic information is remembered, particularly in the context of a real, multimodal environment.

**Haptic Augmentation of Speech Perception**

Speech perception and production are the focus of the chapter by Fowler in this volume. The use of haptic stimulation to augment speech perception is motivated in part by the success of the Tadoma method for speech communication to deaf and blind individuals (Reed et al., 1985). In this method, shown in Figure 6.8, the receiver of communication places his or her hand on the face and neck of a speaker and monitors the movements of the speech musculature. In addition to changing position of the jaw and lips, users have access to changes in air flow, temperature, and vibration. Experienced users can achieve high levels of speech understanding based on multiple sources of sensory information.

Efforts to create haptic stimulators to produce similar effects have varied both with respect to the type of device and the aspect of the speech signal that they attempt to convey. The first formant of the speech signal (F0, or fundamental frequency) has been conveyed by vibration and more recently by vibration-spatial coupling (i.e., both the location and the frequency of the vibration are manipulated; see Auer, Bernstein, & Coulter, 1998, for review). An advantage of conveying F0 is that it is related to several aspects of speech, including voicing, stress, intonation (question vs. statement), and syntactic boundaries. Auer et al. found that when vision was combined with a spatio-temporal display of F0, intonation identification was augmented relative to vision alone.

![Figure 6.8](image-url)
Bernstein (1992) summarized data from a number of extant devices, along with Tadoma data, in terms of information transmitted (Miller & Nicely, 1955). The Tadoma method was superior to any of the aids tested. She reported it "perplexing" (p. 171) that those studies comparing tactile-visual stimulation to that of visual and tactile alone showed only modest gains when the tactile device was added to visual stimulation. Bernstein suggested this might reflect either cross-modal interactions, which would suppress the contribution of one modality in the presence of the other, or redundancy in the visual and tactual speech signals.

The limitations on augmentation of speech by a haptic device reflect, of course, the device itself. Tan, Durlach, Reed, and Rabinowitz (1999) devised a haptic speech device, the Tactuator, that through vibrations and movements of the finger, combines cutaneous and kinesthetic features, hence enriching the stimulus dimensionality. Independent actuators move the fingerpads of the thumb, index finger, and middle finger; the thumb moves perpendicularly to the other fingers so that the hand posture is natural. The system has a temporal response range of up to 400 Hz and can displace the finger by 26 mm. From absolute identification tasks, the authors estimated the information transmission rate at 12 bits/s, comparable to that of Tadoma. The capabilities of the system for augmenting natural speech remain to be demonstrated.

**Teleoperation and Virtual Environments**

A haptic interface is a device that enables manual interaction with virtual or remote environments (Durlach & Mavor, 1994). The device feeds back information to the operator about the consequences of interaction in the remote world. Although the feedback modality is unspecified in principle, it can take the form of haptic feedback, which indicates the forces and vibrations that are imposed on the effector in the remote or simulated world. This type of feedback has been used in two contexts. One is known as teleoperation—that is, when a human operator controls a remote device. The other is virtual haptic environments, in which contact with computer-generated objects and surfaces is simulated. In either case, haptic feedback enhances a sense of telepresence, the feeling that the operator is in a physical environment.

Three types of information are potentially provided by a haptic display. One is directional force feedback, indicating forces that the remote or simulated effector encounters in the environment. Commercial force stimulators are available, such as the PHANToM™, and new laboratory models have been developed (e.g., Berkelman & Hollis, 2000). Another type of information is the sustained, distributed spatial pattern of local forces that generates skin deformation across the fingertip. To generate this information requires a stimulator in the form of a matrix of pins; such devices have been difficult for engineers to implement, although there are some examples (Kontarinis & Howe, 1993). Perhaps the most promising display for immediate application is one that produces vibrotactile stimulation (Cholewiak & Wollowitz, 1992). Vibrotactile stimulation can be produced relatively cheaply, and the frequency and amplitude can be set to optimally activate human mechanoreceptors. An example of this type of display is the Optacon. A more recent development is the vibrating mouse, although that does not present a spatial array of forces.

Haptic displays promise to be useful in many applications in which conveying a sense of physical interaction is important. Haptic feedback has already been found to be essential for performing some tasks, and it is highly useful for others (e.g., Kontarinis & Howe, 1995; Sheridan, 1992). Vibrations in particular, have been shown to improve performance in industrial teleoperation (Dennerlein, Millman, & Howe, 1997), in which a human operator controls a remote robot. Vibratory signals are effective cues to the moment of puncture in medical applications (Kontarinis & Howe, 1995), and they can aid remote manipulation by conveying the forces encountered by a robot effector (Murray, 1999). Other potential applications of haptic displays are to electronic commerce, in which the quality or aesthetic value of produces could be displayed, and haptic augmentation of visual displays of complex data sets (Infed et al., 1999).

Basic research on haptic perception is necessary to guide the development and use of haptic interfaces. For example, Klatzky, Lederman, and associates (Klatzky & Lederman, 1999b; Lederman, Klatzky, Hamilton, & Ramsay, 1999) investigated how people perceived the roughness of a surface composed of raised elements by rubbing it with a rigid probe. These circumstances were meant to model a haptic virtual display in which vibration is the cue to texture. The psychophysical function relating perceived roughness to the spacing of raised elements was quadratic in form, which contrasts with the function typically obtained for roughness perception via the bare skin. The obtained function has direct implications for efforts to simulate texture by altering vibrations to the hand, because it means that any vibratory roughness system must deal with nonmonotonic responses to changes in frequency, amplitude, or both.

**SUMMARY AND FUTURE DIRECTIONS**

This chapter has attempted to provide a view of the modality of touch as a sensory and cognitive system, one that shares many features of perceptual systems but is also, by virtue of
underlying neurophysiology and linkage to the motor system, unique. The brief review of the neurophysiology of touch proves sufficient to show that this modality is based on a variety of receptors, responding to mechanical, thermal, and noxious stimulation. Classical psychophysics has described thresholds for the basic receptors and higher-level properties. Much of this chapter has focused on the role of touch in perceiving properties of objects and surfaces. It has emphasized that touch is particularly adapted for receiving and processing information about the material of which the world is made, more than its form. Nonetheless, form and space perception are performed through touch, and a wide variety of patterns can be discriminated and recognized. The latter part of the chapter portrayed touch as a fully cognitive system, playing a role in the direction of attention and providing a substrate for conscious and implicit memory. The chapter’s conclusion, which identified a number of applications for touch, should make clear the many contexts in which research on human haptic capability is relevant to daily life.

Future research will no doubt characterize the neurophysiology of touch, particularly at cortical levels, much more fully. Comparative neurophysiological work, which relates human and nonhuman systems with respect to this modality, is also ongoing. Research on touch as a cognitive system appears to be just breaking stride; only 20 years ago the basic object-recognition abilities possible through touch had not been widely recognized. Forthcoming research is likely to emphasize even more, as did David Katz (1925/1989) in the early twentieth century, that the sense of touch is an active, richly informative, and highly useful perceptual modality. The burgeoning field of applied haptics will no doubt prove this further by bringing forth new applications to fields such as entertainment, electronic commerce, and telesurgery.

REFERENCES


Kontarinis, D. A., & Howe, R. D. (1993). Tactile display of con-
Klatzky, R. L., & Lederman, S. J. (1985). Identifying ob-
Klatzky, R. L., & Reed, C. L. (1987). There’s more
to touch than meets the eye: The salience of object attributes
Kontarinis, D. A., & Howe, R. D. (1993). Tactile display of con-
LaMotte, R. H., & Srinivasan, M. A. (1991). Surface microgeome-
LaMotte, R. H., & Srinivasan, M. A. (1993). Responses of cuta-
neous mechanoreceptors to the shape of objects applied to the primate fingertip. Acta Psychologica, 84, 41–51.
Lechelt, E. C., Eliuk, J., & Tanne, G. (1976). Perceptual orienta-


