7 Multisensory Texture Perception

SUSAN J. LEDERMAN AND ROBERTA L. KLATZKY

Introduction
Most of the scientific research on the perception of material qualities of objects has focused on the perception of texture. By texture, we mean the microstructure of surfaces as opposed to the large-scale macrostructure of objects (e.g., form, shape). The surface texture is independent of the macrostructure on which it is superimposed. The material from which objects are constructed also possesses other prominent properties, such as softness or hardness, slipperiness and friction, and thermal qualities such as thermal flow and thermal conductivity. The general issues we raise in this chapter with respect to texture perception pertain as well to these other properties of object materials. However, because there has been relatively little scientific research on these topics to date, we will consider them at the end of the chapter.

The perception of surface texture is a multidimensional task; however, the most salient dimensions remain somewhat uncertain (see Hollins, Fechtowski, Rao, & Young, 1993; Hollins & Ritter, 2000). When we describe the texture of a surface, we may focus on dimensions such as roughness/smoothness, bumpiness, or jaggedness, or perhaps on the degree of element cluster that precludes the surface microstructure. The perception of surface texture is also multisensory. That is, regardless of the properties to which we attend, we may use any or all of the following modalities during perception: haptic, vision, and audition. (The haptic system uses sensory inputs from mechanoreceptors in the skin, muscles, tendons, and joints, and typically involves voluntary manual exploration of surfaces, objects, and their spatial layout.)

The availability of more than one sensory source may serve a number of different functions. First, when the information obtained via different sensory modalities about a given property is exactly the same, the modalities will provide redundant cues about the targeted property. Second, even if the sensory sources provide concurrent information, one modality may provide more accurate or more precise information than another. Alternately or in addition, it may obtain that information faster than another modality. Third, the modalities may actually provide discrepant information about a targeted property. Fourth, the senses may simultaneously provide qualitatively different yet complementary information about a texture. For example, vision provides information about surface properties such as color, lightness, and reflectance, whereas touch provides information about roughness, hardness, slipperiness, and thermal properties.

The following discussion considers relevant empirical findings, theoretical perspectives, and the implications of such fundamental scientific work for the design of multisensory interfaces for teleoperation and virtual environments.

Empirical findings
Empirical studies have addressed two critical questions that pertain to multisensory texture perception. Our presentation of the experimental research findings is organized around these two questions.

1. How does the perceiver integrate information about surface texture from different sensory modalities?

The perceiver may choose to ignore information made available through one modality (e.g., Rock & Victor, 1964). Alternatively, the perceiver may integrate the different sensory sources in some form of compromise that is not identical to any one of them. Table 7.1 summarizes the multisensory dominance findings of this and other relevant experiments discussed in this chapter.

Sensory conflict and sensory dominance paradigms offer one experiential approach to the study of multisensory integration. Several studies have employed either artificial or natural discrepancies in the texture information presented simultaneously to the eye and the hand.

We begin with studies that focus on haptics and vision. Lederman and Abbott (1981, Experiment 1) created an artificial discrepancy between the two senses by simultaneously exposing two physically different (i.e., discrepant) black abrasive surfaces to vision and touch. The subjects in this modality discrepancy group were led to believe that they were examining two spatially separated areas on a single surface. Two other modality control groups (vision only, touch only) examined
<table>
<thead>
<tr>
<th>Study</th>
<th>Property</th>
<th>V</th>
<th>H</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lederman &amp; Abbott (1983)*</td>
<td>Texture (Exp. 1)</td>
<td>V = 50%</td>
<td>H = 50%</td>
<td>N/A</td>
</tr>
<tr>
<td>Lederman, Thorne, &amp; Jones (1986)*</td>
<td>Roughness (Exp. 2)</td>
<td>V[H] = 33.2%</td>
<td>H[V] = 73.2%</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>Roughness (Exp. 3)</td>
<td>V = 51.0%</td>
<td>H[V] = 69.0%</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>Roughness (Exp. 6)</td>
<td>V &lt; H</td>
<td>H[V] = 6.6%</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>Spatial density (Exp. 1)</td>
<td>V[H] = 76.6%</td>
<td>H[V] = 51.0%</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>Density (Exp. 4)</td>
<td>V[H] = 49.0%</td>
<td>H[V] = 51.0%</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>Spatial density (Exp. 5)</td>
<td>V &gt; H</td>
<td>H[V] = 6.6%</td>
<td>N/A</td>
</tr>
<tr>
<td>Klazny, Lederman, &amp; Reed</td>
<td>Object similarity</td>
<td>Geometry favored when biased to sort by vision</td>
<td>Material favored when biased to sort by haptics</td>
<td>N/A</td>
</tr>
<tr>
<td>Klazny, Lederman, &amp; Maunda (1993)</td>
<td>Material properties</td>
<td>Vision always available</td>
<td>Haptics rarely used, when tasks easy; H used on 80% of trials when tasks hard (H only used to judge material)</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>Vision always available</td>
<td>Haptics never used, regardless of task difficulty (only used to reorient objects to judge size)</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Geometric properties</td>
<td>Vision always available</td>
<td>Haptics never used, regardless of task difficulty (only used to reorient objects to judge size)</td>
<td>N/A</td>
</tr>
<tr>
<td>Lederman (1979)*</td>
<td>Roughness (Exp. 3)</td>
<td>N/A</td>
<td>H = 100%</td>
<td>A = 0%</td>
</tr>
<tr>
<td>Lederman et al. (2002)*</td>
<td>Roughness</td>
<td>N/A</td>
<td>H = 62%</td>
<td>A = 38%</td>
</tr>
<tr>
<td>Guest &amp; Spence (in press)</td>
<td>Roughness</td>
<td>V influenced by incongruent haptic distractors</td>
<td>H uninfluenced by incongruent V distractors</td>
<td>N/A</td>
</tr>
<tr>
<td>Joumaki &amp; Hari (1998)</td>
<td>Skin roughness and moisture</td>
<td>N/A</td>
<td>Bias of A by H not assessed</td>
<td>A biased H</td>
</tr>
<tr>
<td>Goset et al. (in press)</td>
<td>Roughness of abrasive surfaces (Exp. 1)</td>
<td>N/A</td>
<td>Bias of A by H not assessed</td>
<td>A biased H</td>
</tr>
<tr>
<td></td>
<td>Roughness and wetness of skin (Exp. 2)</td>
<td>N/A</td>
<td>Bias of A by H not assessed</td>
<td>A biased H</td>
</tr>
<tr>
<td></td>
<td>Roughness and wetness of skin (Exp. 5)</td>
<td>N/A</td>
<td>Bias of A by H not assessed</td>
<td>A biased H</td>
</tr>
</tbody>
</table>

Note: Percent bias of one modality by a second modality was calculated as discussed under Empirical Findings. Abbreviations: V, vision; H, haptics; A, audition; V[H], visual bias if haptics; H[V], haptic bias of vision.

*These modality bias effects were not independent and must sum to 100.

Either the visual or the tactile standard texture from the discrepant standard stimulus pair.

Subjects were required to select the surface that best matched the "texture" of the standard stimulus from among pairs of identical comparison surfaces. Within each of the three standard modality groups, three subgroups responded using vision only, touch only, or vision and touch together. There were thus nine groups in all. The results indicated that subjects selected a surface pair that lay midway (i.e., 50%) of the three standard modality groups.
between the two discrepant stimuli, regardless of the response-matching modality (Fig. 7.1). Because the effect of response modality was not significant, the matching responses are averaged over the three response-modality conditions in the figure. The Lederman and Abbott (1981) study showed that subjects altered their matching response in the bimodal disparity condition so that it lay halfway between those of the unimodal controls.

In a series of six experiments, Lederman, Thorne, and Jones (1986) demonstrated that there is in fact no fixed dominance hierarchy among modalities for the perception of textured surfaces, and with this observation underscored the multidimensional nature of texture perception. Changing the dimension that subjects were required to judge produced sizable adjustments in the relative weighting of the haptic and visual inputs. On each trial, subjects were asked to numerically estimate the perceived magnitude of the “spatial density” of raised two-dimensional dot patterns (Experiment 1) or of the “density” of abrasive surfaces (Experiment 4). Subjects examined the surface by touch alone, by vision alone, or by touch and vision together. In the bimodal trials, a pair of surfaces was presented, but unknown to the subject, the two were discrepant. Only after exploration was the subject informed whether to judge how the surface felt (touch in the presence of vision) or looked (vision in the presence of touch). The independent mutual biasing effects of touch and vision were calculated as follows:

Visual bias of haptics (% V/H) = \( \frac{(H_v - V[H])}{(H_v - V)} \times 100 \)

Haptic bias of vision (% H/V) = \( \frac{(V_v - H[V])}{(V_v - H)} \times 100 \)

where \( V[H] \) and \( H[V] \) represent the results of the bimodal conditions that indicate the mutual biasing effects of vision on haptics and of haptics on vision, respectively, and \( V_v \) and \( H_v \) represent the results of the visual and haptic conditions, respectively.

Felt spatial density judgments were strongly dominated by the concomitant visual inputs (i.e., 76.5%, Experiment 1), while felt density judgments of abrasive surfaces were somewhat less so but still strongly affected by vision (i.e., 49.0%, Experiment 4). (Two reasons for the difference in the magnitude of the visual bias on touch were suggested. The simpler term density, which was used subsequently to avoid possible confusion among subjects concerning what they were to judge, may have also de-emphasized the importance of the spatial cues. Alternatively, judging the (spatial) density of abrasive surfaces may have been more difficult than judging the more clearly distinct photogravure raised-dot surfaces.) The influence of haptic inputs on the independently obtained but corresponding estimates of seen spatial density was very minor, only 7.0% (Experiment 1); for density judgments of abrasive surfaces, haptic dominance was equivalent to visual dominance, 51.0% (Experiment 4), for reasons we suggest later. In sum, subjects weighted touch and vision cues about equally when asked to judge surface texture (Lederman & Abbott, 1981). In contrast, when asked to judge (spatial) density, they tended to weight the visual inputs more strongly. And when asked to judge the roughness of these same surfaces, they weighted the tactile inputs more strongly (Lederman et al., 1986).

The method used to assess the relative contributions of vision and touch to bimodal judgments of discrepant textured surfaces required that the subject believe that the visually and tactually derived information depicted the same surface. Experiments 5 and 6 used a different method known as functional measurement (Anderson, 1974). In this method, subjects are knowingly presented with pairs of stimuli that may or may not be the same. Anderson demonstrated that across a variety of tasks (e.g., weight assessment), information from at least two sources is integrated according to simple additive models. Subjects are presented with pairs of values drawn from two independently manipulated stimulus dimensions. The task requires subjects to provide a combined judgment of pairs of stimuli produced.
by factorially combining all possible values from the two dimensions.

Lederman et al. (1986) told this paradigm as a converging method for comparing the relative weighting of vision versus haptic judgments of "spatial density" (Experiment 5) versus "roughness" (Experiment 6). The stimulus pairs were drawn from the same set of two-dimensional raised-dot patterns used in Experiments 1 and 2. When subjects assessed surface roughness bidimensionally by vision and touch, they treated the two modality inputs as independent. The results of these experiments confirmed the relative dominance of touch by vision for the spatial density task and of vision by touch for the roughness task. The results were best described by a linear averaging model in which subjects averaged the two unimodal control judgments to produce a weighted average judgment in the corresponding bimodal condition. The theoretical details of this model are discussed later in the section titled Modeling Multisensory Texture Perception.

Sensory dominance and intersensory integration have been explored with respect to texture perception using a number of other methodologies. For example, Klazky, Lederman, and Reed (1987) demonstrated the relatively greater importance of surface texture and hardness for touch than vision using a different experimental paradigm. They custom-designed and constructed a large set of multiattribute artificial objects that varied along four different object dimensions in psychophysically equivalent ways, whether objects were grouped according to their surface texture (three values), hardness (three values), shape (three values), or size (three values). The three different values for each property were combined across all dimensions, producing a set of 81 objects. Each object had one of three textures, one of three hardness values, one of three shape values, and one of three size values. Subjects were required to sort the objects into piles according to their perceived similarity. Objects that were most similar to each other were to be grouped together. Different instruction sets were used to deliberately bias subjects toward favoring their haptically derived or visually derived object representations in their sorting judgments. When biased to process and represent objects in terms of haptic similarity, subjects preferred to sort objects in terms of the available material properties as opposed to the objects' geometric properties. In contrast, when biased to sort on the basis of visual similarity, subjects sorted primarily by shape and secondarily by texture. People tend to be more efficient—more accurate or faster—when accessing precise material variation with their hands than with their eyes, as other research has tended to show (e.g., Heller, 1989; Lederman & Klazky, 1987; but see also Jones & O'Neil, 1985, discussed later in this chapter under question 2: "How does multisensory integration affect performance?"). Conversely, when biased to process and represent objects using vision, subjects strongly preferred to sort by differences in object shape. This finding too makes sense, because vision is considerably more efficient at processing geometric information (e.g., Klazky & Lederman, 2000; Lederman & Klazky, 1987; Walk & Pick, 1981). The results of this study were subsequently confirmed by Lederman, Summala, and Klazky (1986) with threedimensional objects.

Collectively, these studies suggest that when multiattribute objects vary in perceptually equivalent ways, subjects choose to develop representations of these objects more in terms of the objects' variation in material properties (e.g., texture, hardness) than in terms of their geometric properties (e.g., size, shape) when focused on the haptic inputs. Our data indicate that texture plays an important role in both types of representations; however, it appears to be more cognitively salient to haptically than to visually derived representations of the same objects.

Klazky, Lederman, and Mautla (1993) used a different experimental paradigm to address the issue of multisensory texture perception. They considered situations in which people chose to use touch, even though visual texture was available. In the study, subjects were required to make discriminations between commonobject pairs with respect to six targeted perceptual dimensions—four material dimensions (roughness, hardness, thermal properties, and weight) and two geometric dimensions (shape and size). When the discriminations were very easy, subjects chose not to use touch very often, relying solely on vision. The situations in which touch was used for easy judgments tended to involve decisions regarding the material characteristics of objects. When the discriminations were difficult, subjects chose to use touch in addition to vision in about 80% of the trials that required judgments about material properties, including roughness. In contrast, subjects did not use touch to make perceptual discriminations about shape or size. This study indicates that people elect to use textual information about surface roughness when relatively precise information about the material features of objects is required, even when visual information is available.

Finally, Guest and Spencer (2003) have examined visual-tactual integration of texture information in a speeded texture task. Participants were required to make speeded discrimination judgments of the roughness of abstract (painted) textile samples by one modality (either vision or touch) in the presence of congruent or
incongruent information about a textile distracter presented to the other modality (touch or vision, respectively). Visual discrimination of textile roughness was altered by incongruous tactile distracters. However, the reverse did not occur, even when the visual distracters were more discriminable than the tactile targets. Guest and Spence concluded that the asymmetric interference effect implied that ecological validity must play a role. In addition to modality appropriateness. In keeping with Lederman’s initial interpretation (1979), which will be discussed in the following section on haptics and audition studies, Guest and Spence argued that the assessment of textile surfaces is better suited to tactile than to visual assessment.

In summary, there appears to be no fixed sensory dominance with respect to the multisensory perception of texture by vision and haptics. Rather, the lability with which one modality dominates the other with respect to order and magnitude is influenced by the selected emphasis on some aspect of surface texture (e.g., texture, roughness, spatial density).

We turn now to studies that have involved haptic and auditory processing. Another potentially valuable source of information about surface texture is the accompanying set of sounds that are generated when one touches a surface. For example, Katz (1989) showed that people are remarkably skilled at identifying different materials using the touch-produced sounds alone.

Lederman (1979) used a natural discrepancy paradigm to investigate whether people can use such sounds alone to evaluate the roughness magnitude of metal gratings. The paradigm also allowed consideration of the relative contributions of haptic and tactically produced sounds to the perception of roughness when both sources of information were simultaneously available. In one experiment (Experiment 3), subjects were required to estimate the magnitude of perceived roughness of rigid, two-dimensional gratings that varied in terms of groove width (constant ridge width). For such stimulus surfaces groove width is known to be the primary parameter that affects the haptic perception of roughness by touch (in contrast to ridge width, spatial period, and groove-to-ridge width ratio, which have little or no effect; see, e.g., Lederman, 1974). Subjects participated in each of three modality conditions—haptics only, audition only (sounds generated by experimenter), and haptics plus audition (sounds generated by the subject). The psychophysical functions (perceived roughness magnitude as a function of the groove width of linear engraved gratings, on log-log scales) increased considerably less rapidly for audition alone (particularly for the plates with the narrower grooves) than for either haptic alone or haptics plus audition.

These results indicate that subjects could use the auditory cues generated by gratings, although not as well as they could the haptic cues. Indeed, when both auditory and haptic cues were available, the psychophysical function for roughness was virtually identical to that for haptics alone (Fig. 7.2; see also Table 7.1). That these two functions were identical suggests that in contrast to the results of von Schiller (1992), subjects in Lederman’s (1979) experiment ignored the available auditory information when estimating roughness magnitude.

In Lederman’s (1979) study, subjects used their bare fingers to explore the textured surfaces. However, people also frequently use intermediate tools to perceive and interact with objects. When judging surface properties haptically with a probe, people must use vibrations as the source of their information, one that is accessible to both hand and ear. Jen recently Lederman, Klatsky, Hamilton, and Morgan (2002) used Lederman’s (1979) psychophysical paradigm (magnitude estimation) in their investigation of the remote perception of surface texture via a rigid stylus. Different groups of subjects explored the roughness magnitude of spatially jittered, raised two-dimensional dot patterns under one of three different modality conditions: unimodal haptic, unimodal audition (sounds produced by the experimenter), and bimodal haptics plus audition (sounds produced by the subjects). The results indicated that, as with direct contact between hand and surface exploration with a stylus created a natural perceptual discrepancy between unimodal haptic and auditory estimates of roughness; that is, the two psychophysical control functions (haptics only and audition only) differed. However, unlike the results of the earlier bare finger study, when both sensory sources were simultaneously
available, subjects chose to use both remote sources of information in their bimodal judgments, weighting the haptic inputs somewhat more (62%) than the auditory inputs (38%) (Fig. 7.3).

Converging evidence that people use auditory information to judge surface roughness remotely via a probe in bimodal touch/audition conditions was reported in a second experiment in this study (Morgan, 2001).

Subjects bimodally explored the same raised two-dimensional dot patterns with a rigid probe in each of three sound-amplification level conditions. Their roughness estimates increased systematically as a function of the increasing amplification of the touch-produced sounds produced as the subject explored the surfaces with a rigid probe. The results converge with those reported by Lederman et al. (2002), in that subjects did indeed use the touch-produced sounds to make their remote judgments bimodally, using cues generated with a rigid probe and concomitant touch-produced sounds.

Additional support for the intersensory bias of haptic estimates of surface roughness by concomitant sounds comes from an intriguing phenomenon discovered by Joussamaki and Hari (1998), known as the parchment-skin illusion. As subjects rubbed their hands back and forth together, the sounds were recorded and played back to the subjects through headphones as normal or altered sounds. One parameter was the amplitude of the sound frequencies above 2000 Hz, with which the natural level (i.e., identical to the original sounds), amplification to 15 dB above natural level, and attenuation to 15 dB below the natural level. A second parameter was the average sound level, which was adjusted relative to a comfortable listening level ("normal"). Values included the average sound level adjusted to normal, 20 dB below normal, and 40 dB below normal. Subjects were required to judge skin roughness on a scale from 0 (tough or moist) to 10 (smooth or dry). Subjects judged the skin surface to increase in smoothness/dryness as the values of both parameters were increased. This increase was compared to the feel of parchment paper. The results of this study clearly indicate the bias of touch by audition and a role for sound amplitude in the perception of the roughness of this naturally occurring surface.

Guest, Cannutt, Lloyd, and Spence (2002) extended the work of Joussamaki and Hari by investigating the role of haptically produced sounds on the perceived roughness of abrasive surfaces. In contrast to Joussamaki and Hari, they found that in a speeded discrimination task, reducing the high-frequency auditory feedback biased subjects toward judging surfaces by touch as smoother (Experiment 1). Experiment 2 replicated Joussamaki and Hari's hand-rubbing experiment but asked subjects to scale both the wetness and the roughness of their hands. The results for wetness judgments replicated Joussamaki and Hari's findings, however, those for roughness confounded level, which was greater in Experiment 1. Finally, in Experiment 3, they showed that delaying the auditory feedback for hand rubbing reduced the size of the parchment-skin illusion in Experiment 2. Collectively, the results strongly indicate the influence of concurrent sounds (i.e., auditory frequency) on the tactile assessment of roughness and hand moisture. Gäver (e.g., 1995) has provided additional examples of "textural" properties obtained via sound.

To summarize, as noted with the vision and haptic intersensory bias studies, results of the audition and haptics studies likewise indicate that the relative order and magnitude of sensory dominance is labile. As we later discuss later under Modeling Multisensory Texture Perception, transient or short-term factors may also play a role by influencing the relative perceptual accessibility of the stimulus information available in the given task.

2. How does multisensory integration affect performance?

If the perceiver integrates information from more than one modality, do the multiple sources improve, impair,
<table>
<thead>
<tr>
<th>Study</th>
<th>Task</th>
<th>Performance Variables</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2. Rate of growth of perceived &quot;roughness&quot; (magnitude estimation)</td>
<td>1. Variability</td>
<td>$H = V = (H+V)$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2. Exponent of psychophysical power function</td>
<td>$2. H = V = (H+V)$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2. Variability</td>
<td></td>
</tr>
<tr>
<td>Jones &amp; O'Neil (1985)</td>
<td>1. 2-AFC for relative &quot;roughness&quot; (Exp. 1)</td>
<td>1. Accuracy</td>
<td>$H = V = (H+V)$</td>
</tr>
<tr>
<td></td>
<td>2. Same/different for relative roughness (Exp. 2)</td>
<td>1. Response time</td>
<td>$1. V &lt; (H+V) &lt; H$</td>
</tr>
<tr>
<td></td>
<td>3. 3-AFC for relative roughness (Exp. 5)</td>
<td>2. Accuracy</td>
<td>$2. H = V &lt; (H+V)$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3. Response time</td>
<td>$V &lt; (H+V) &lt; H$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3. Accuracy</td>
<td>$H &lt; (H+V) &lt; V$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3. Response time</td>
<td></td>
</tr>
<tr>
<td>Heller (1982)</td>
<td>3-AFC for relative &quot;smoothness&quot; (course textures)</td>
<td>Accuracy</td>
<td>$H = V$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$V &lt; (H+V)$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$H$ and $V &lt; (H+V)$</td>
<td></td>
</tr>
<tr>
<td>Heller (1989)</td>
<td>2-AFC relative &quot;smoothness&quot;</td>
<td>Accuracy (course stimuli)</td>
<td>$H = V$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>accuracy (very smooth stimuli)</td>
<td>$H &gt; V$</td>
</tr>
<tr>
<td>Lederman (1979)</td>
<td>Magnitude estimation of roughness (Exp. 1)</td>
<td>Exponent of psychophysical power function</td>
<td>$A &lt; H$</td>
</tr>
<tr>
<td>Lederman, Klatzky, Hamilton,</td>
<td>Magnitude estimation of roughness (Exp. 2)</td>
<td>Exponent of psychophysical power function</td>
<td>$H = (H+A)$</td>
</tr>
<tr>
<td>Morgan (2002)</td>
<td>Confidence estimation of magnitude estimates of roughness</td>
<td>Confidence ratings</td>
<td>Mean of $H$ and $A$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rate of learning</td>
<td>$H = (A+H) &lt; A$ (but see text for proper interpretation)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Confidence rating</td>
<td>$A &lt; H = (A+H)$</td>
</tr>
<tr>
<td>Wu, Baugnon, &amp; Srisurapan (1999)</td>
<td>2-AFC relative stiffness</td>
<td>Resolution</td>
<td>$H$ and $V &lt; (H+V)$</td>
</tr>
</tbody>
</table>

Abbreviations: A, audition only; H, haptic only; V, vision only; 2(3)-AFC, two (three)-alternative forced-choice task. Bimodal conditions are represented in parentheses.

or have no influence on performance (e.g., accuracy, variability, confidence, response time)? For the purpose of comparing unimodal and bimodal performance on texture-related perceptual tasks, Table 7.2 summarizes the results of all experiments in this section. Many of these studies were discussed in the preceding section.

Once again, we begin with studies on vision and haptics. An additional experiment in the Lederman and Abbott study (1981, Experiment 2) assessed the associated relative performance by the two modalities on a texture-matching task. Subjects participated in one of three modality conditions: touch only, vision only, or haptics plus vision. In this experiment, however, the bimodal condition presented identical pairs of relatively coarse, black abrasive surfaces to both vision and touch.

The results indicated virtually identical performance among the two unimodal and one bimodal groups, both in terms of texture-matching accuracy and in terms of response precision (i.e., consistency). Finally, Lederman and Abbott (1981, Experiment 3) compared haptics, vision, and touch plus vision in terms of the rate of growth of the corresponding psychophysical functions for perceived "roughness" magnitude (i.e., the increase in reported roughness magnitude with increased growth).
size of the particles that formed the textured surface. Subjects were required to judge the roughness magnitudes of black abrasive surfaces using a magnitude estimation procedure. The results indicated that there were no differences among the modality conditions in either rate of growth of roughness magnitude or relative precision. In addition, there was no difference in performance among the two unimodal and bimodal conditions with respect to accuracy and precision on the texture-matching task (Experiment 2) and with respect to rate of growth and precision on a magnitude estimation task (Experiment 5).

Jones and O'Neil (1985) also compared unimodal (vision only, haptics only) and bimodal judgments of surface roughness. Subjects performed two-alternative, forced-choice (Experiments 1 and 3) and same/different (Experiment 2) tasks. In keeping with the earlier results obtained by Lederman and Abbott (1981), subjects tended to be about equally accurate on Experiments 1 and 2, regardless of the modality used during exploration. Moreover, using two sources of information about surface roughness did not improve accuracy compared with either of the unimodal conditions. With respect to speed, vision was faster than touch, with the bimodal condition falling midway between the two unimodal conditions. Thus, vision was more effective than touch with respect to speed of responding, and served to improve unimodal tactile performance, whereas the addition of touch slowed down unimodal visual performance. The relatively greater temporal efficiency of vision is not surprising, because the eye can process a wide spatial area simultaneously without movement, while the hand must move to sample the surface. In Experiment 3, subjects were more accurate and faster using unimodal vision than unimodal haptics, while the bimodal condition lay between the two.

Heller (1982) required his subjects to perform three-alternative, forced-choice relative “smoothness” tasks using a similar range of abrasive surfaces to those used by Lederman and Abbott (1981) and Jones and O'Neil (1985). In Experiment 1, subjects chose the smoothest of three abrasive surfaces, with texture trials crossed from the set of abrasive surfaces. Subjects participated in three modality conditions: unimodal vision, unimodal haptics and bimodal haptics plus vision. Although subjects performed very well in all three modality conditions, the bimodal condition resulted in greater accuracy than either of the unimodal haptics or vision conditions, which were not significantly different from each other (Fig. 7.4; see also Table 7.2). The added benefit of the bimodal inputs was considered as well in Experiment 2, which confirmed Heller’s earlier findings, namely, the equivalent accuracy in the two unimodal conditions and greater accuracy in the bimodal condition.

In summary, unlike either Lederman and Abbott (1981) or Jones and O'Neil (1985), Heller found that performance improved when subjects used bimodal (i.e., vision and haptics) exploration (see Table 7.2). He found no benefit from the auditory cues that were also available, in keeping with an earlier study by Lederman (1979) that will be discussed in detail later in the chapter. Experiment 2 also confirmed that subjects in Experiment 1 did not capitalize on the better visual illumination cues available only in the bimodal condition of that experiment. (In Experiment 1, subjects had been allowed to handle the objects with the gloved, nonexploring hand and thus to move them closer for better viewing.) Therefore, in Experiment 3, Heller extended his examination of the role of vision in bimodal texture judgments, using the same set of surfaces as in Experiment 2. Subjects participated in one of three experimental conditions. One group was required to judge the relative smoothness of abrasive surfaces using available visual and haptic cues to texture, along with sight of their hands. A second group was permitted to use normal haptic texture cues and subjects could see their hands; however, they were denied visual cues to texture. In this condition, subjects manually explored the abrasive surfaces while viewing their hands through a plastic stained glass filter. The visual filter served to eliminate use of visual texture cues while retaining sight of the moving hand for purposes of visual guidance. If subjects used visual texture cues in their bimodal judgments, then Group 1 should have performed better.
than Group 2. However, there was no difference in performance, leading Heller to conclude that subjects lost sight of their hand movements, not the visual texture cues, to make their bimodal estimates. Group 3 served to confirm that visual judgments of texture were indeed impossible when the surfaces were viewed through the visual filter. In this condition, both the haptic cues to texture and sight of the moving hand were eliminated as potential sources of information.

Heller suggested that the superior bimodal performance on the smoothness discrimination task resulted because the two modalities cooperated with one another, with touch providing the relevant texture information and vision providing complementary information about the hand movements. However, his data do not suggest that subjects used texture inputs from both modalities to derive a compromise judgment about relative smoothness.

In a subsequent paper, Heller (1989) showed that texture perception of relatively coarse abrasive surfaces was essentially equivalent for sighted, late-blind, and early-blind subjects. He concluded that visual imagery was not necessary for texture perception. He used a two-alternative, forced-choice task to compare the accuracy with which subjects picked the smoother of two surfaces using either unimodal haptics or unimodal vision. In a second experiment, Heller required subjects to perform the same two-alternative, forced-choice relative smoothness task; however, pairs were chosen that included new very smooth surfaces in addition to the previous coarser surfaces (similar to those used in his 1982 study). The results confirmed the equivalence of touch and vision for the coarser pairs; however, touch proved to be considerably more accurate than vision for the very smooth pairs (see Table 7.2).

In summary, the studies on relative performance in conditions of unimodal vision, unimodal haptics, and bimodal vision plus haptics (see Table 7.2) have collectively tended to show that haptics and vision are relatively equal in terms of their relative accuracy, variability, and rate of growth of roughness magnitude; not surprisingly, vision is faster. With the exception of Heller (e.g., 1982), most studies have shown no benefit from using both modalities together.

Next we consider studies that have addressed relative performance involving unimanual haptics, unimanual audition, and bimanual haptics plus audition. In the previously described study by Lederman (1979), subjects also estimated the roughness magnitude of linear gratings in two other experiments. In Experiment 1, subjects used only the haptically produced sounds produced by the experimenter. In Experiment 2, subjects used either haptics or haptics plus audition together; in the latter case the sounds were produced by the subject. The stimuli set consisted of gratings that varied in ridge width (groove width was constant). Perceived roughness tended to decline with increasing ridge width regardless of the modality (or modalities) used. Moreover, the unimanual and bimanual haptics conditions were not statistically different from one another. The auditory estimates for the wider-ridge gratings differentiated the plates less than did estimates from the two haptics conditions, suggesting a slight advantage for haptics over audition. Nevertheless, neither modality was particularly effective at differentiating the set of ridge-varying gratings.

Finally, we consider recent data from roughness perception tasks that involved exploring raised two-dimensional dot patterns with a rigid probe as opposed to the bare finger. Confidence estimates were also obtained at the end of the experiment by Lederman et al. (2002, Experiment 1) from subjects in the haptics, audition, and haptics plus audition conditions. The results were suggestive: confidence in the bimodal condition was higher than the mean of the combined unimodal conditions. Because relative performance was not the primary focus of the Lederman et al. (2002) study, Lederman, Klatsky, Martin, and Tong (2003) subsequently addressed this issue directly by comparing performance on a difficult absolute texture (i.e., roughness) identification task with a rigid probe. Subjects were assigned to one of the following three exploration conditions: haptics only, audition only (haptically produced sounds), or haptics plus audition. Subjects learned to associate a man's name with each of the stimulus textures. Feedback was provided after each response. Subjects repeated the stimulus set a total of 14 times. Performance for each modality was assessed in terms of efficiency (i.e., amount learned and rate of learning) and confidence. Unimodal haptics proved to be equivalent to bimodal haptics plus audition, and both were associated with greater learning and higher confidence scores than unimodal audition. The greater learning rate for audition than haptics may seem to indicate that audition is the superior modality for this task, but that is not the case: the result actually occurred because accuracy for audition was lower than haptics at the start, and both modality conditions ultimately reached the same level of accuracy. When the experiment was repeated using an easier texture identification task, the pattern of results was similar; however, subjects were equally confident, regardless of whether either unimodal condition or the bimodal condition was used.

In conclusion, studies comparing haptics and audition in texture perception have shown that when the
barring finger is used, the rate of growth in roughness is slower for audition than for haptics, which in turn is equivalent to bimodal haptics plus audition. When touching is remote (i.e., via a rigid probe), if the task is difficult, subjects tend to show an advantage early in learning and become confident using either haptics alone or bimodal haptics plus audition than using audition alone. When the task is easier, the results are similar, but subjects are equally confident in all three modality conditions.

**Modeling multisensory texture perception**

To fully understand multisensory texture perception, it is critical that in addition to addressing issues that pertain to intersensory representation, we must also consider those relating to the processes by which those representations are created. In the following section, we describe several theoretical approaches that have been, or could be, adopted.

The "**Modality Appropriateness**" Hypothesis. This hypothesis offers an early qualitative theoretical approach to the basis of intersensory integration (e.g., Friede, 1974; Welch & Warren, 1980). According to the general form of the modality appropriateness hypothesis, observers choose to weight the various modality inputs according to their relative unimodal performance capabilities with respect to the given task (form, size, texture, etc.). Several relative performance measures have been considered: accuracy, response time, and precision or variability. Even relative availability of different information sources has been proposed; however, just below we consider whether it is as appropriate a dependent measure as the preceding ones. The modality appropriateness interpretation contrasts with that used in earlier research, which argued for the universal dominance of vision in perception (e.g., Rock & Victor, 1984). According to the modality appropriateness hypothesis, however, vision should strongly dominate touch, which should dominate audition, on spatial tasks, because vision is spatially best and audition is spatially worst (egocentric location, e.g., Haas, Piek, & Ikeda, 1965; object shape and size, e.g., Rock & Victor, 1964; orientation, e.g., Oer, 1966). In contrast, audition should strongly dominate vision (with touch in the middle) on temporal tasks, because people perform such tasks best with audition and worse with vision (e.g., Myers, Cotton, & Hilp, 1981; Nazarro & Nazarro, 1970; Welch, Dutton-Hart, & Warren, 1980).

As noted earlier, texture perception is a multidimensional task that can rely on extensive spatial or temporal processing (or both), depending on the particular property in question (e.g., roughness/smoothness vs. spatial density of the surface elements) and the gross-mal stimulus cues used to derive the percept (e.g., spatial deformation vs. vibration). Indeed, we saw that subjects could change their belief about equality when they were asked to judge the perceived "texture" of abrasive surfaces (Lederman & Abbott, 1981) using a bimodal discrepancy paradigm. Because the term texture was deliberately left undefined, subjects could have been performing this task by focusing on different surface properties with vision and touch. For example, subjects might have processed roughness via touch and spatial density via vision. After all, the haptic system is equal to or better than vision at discriminating surface roughness in terms of accuracy and precision (Heller, 1985; Lederman & Abbott, 1981), whereas vision is superior to haptics in processing fine spatial details. As interpretation is supported by the results of Lederman et al. (1985) study. When subjects were explicitly instructed to judge the spatial density of a discrepant pair of textured surfaces, they weighted the visual cues more strongly; in striking contrast, when asked to judge the roughness of the same surfaces, subjects weighted the haptic cues more strongly.

The results that pertain to the relative weighting of auditory versus haptic information on roughness perception tasks may also lend themselves to a modality appropriateness interpretation, which assumes that people process information in a modality according to the value it possesses, based on long-term experience about equality, Lederman (1979) showed that subjects ignored the touch-produced sounds when using their bare finger to explore. She argued that such sounds are typically of very low amplitude, and therefore are often masked by other exogeneous environmental sounds. As these touch-produced sounds are relatively unavailable during common auditory experience, observers may simply choose to ignore them. However, contrary to this interpretation, when texture-related sound cues were generated instead with a rigid probe (Lederman et al., 2002), subjects weighted them more strongly in their judgments of roughness magnitude. That is, they no longer ignored the auditory information, although that would be appropriate to their common experience with that modality. We met note, however, that probe-produced sounds are typically considerably louder than those produced with a bare finger, and therefore would be relatively more accessible. In hindsight, such logic implies that the appropriateness of a modality per se depends to some extent on transient characteristics of the input rather than on its general utility. This was probably not the intent of either.
Friedes (1974) or Welch and Warren (1980) when they proposed the modality appropriateness hypothesis.

Here we consider a new interpretation of the results offered by Lederman and her colleagues (Lederman, 1979; Lederman et al., 2002). The factor they addressed, relative cue availability, relates more to transient modality effects than to those due to the long-term characterization of the modality per se (e.g., accuracy, precision, speed of response). Lederman et al. (1996) have recently proposed that long-term appropriateness and current (and therefore short-term) accessibility be treated as independent factors that might both affect the relative cognitive salience of different sources of information. The transient changes in the relative modality weights may be primarily determined by what they have called relative "perceptual accessibility" of information to multiple modalities that may vary across perceptual conditions. When perceptual accessibility of a cue to one modality is considerably higher than to another modality, it could mask more long-term effects of modality appropriateness that favor the latter. This appears to be the case for probe-produced sounds, which, having relatively high amplitude, override a more general tendency to ignore auditory cues produced when feeling textures. A conceptual model that reflects the contribution of these factors is presented in Figure 7.5, using haptics and audition as examples. The stronger weighting of haptics than audition is indicated by the relative thickness of the arrow representing the generation of the haptic weight.

Regardless of how we choose to treat the availability of cues to different modalities, the modality appropriateness hypothesis does not allow quantitative predictions about the relative modality weights. Nor does it specify what particular factors affect the process or processes that determine the differential weightings (e.g., long-term modality appropriateness or immediate stimulus attributes).

In summary, many human studies using the perceptual discrepancy paradigm have offered strong behavioral support for the bimodal percept of a single object as involving some compromise (9%-10%) between the two unimodal inputs—for spatial location (e.g., Hay et al., 1965; Pick, Warren, & Hay, 1969), for shape (e.g., Rock & Victor, 1964; Miller, 1972), for size (e.g., Fishkin, Fishkin, & Stahl, 1975; Klein, 1986); for orientation (e.g., Oer, 1966; Singer & Day, 1966), and for texture (e.g., Heller, 1989; Lederman & Abbot, 1981; Lederman et al., 1986). This can result in bimodal performance being inferior to performance with a unimodal condition.

However, there is also behavioral evidence of information summation, both within and between the senses (e.g., within vision, Meese & Williams, 2000; for smell and taste, Dallos, Doolittle, Nagata, & Bessin, 2000; for audition & vision, Mulligan & Shaw, 1990). With this

![Figure 7.5](image)

**Figure 7.5.** A conceptual model of intersensory integration. Oval boxes represent processors, square boxes represent outputs, and arrows represent process.

LEDERMAN AND KLATZKY: MULTISENSORY TEXTURE PERCEPTION 117
outcome, the bimodal condition results in superior performance relative to either unimodal input.

In addition, there now exists a substantial neurophysiological literature that is relevant to multisensory integration. This work indicates that the responses elicited by simultaneous inputs from two modalities may be greater than the responses elicited by either modality on its own, and may well be significantly greater than the sum of the unimodal responses found within and across many sensory domains (e.g., visual–auditory, auditory–somatosensory, and somatosensory–visual in primate superior colliculus [SC]; Wallace, Wilkinson, & Stein, 1996; for cat SC, Meredith & Stein, 1986; see also Calvert, Campbell, & Brammer, 2000, for superadditivity in humans). There is one major difference between discrepancy studies on intersensory texture perception and those on multisensory integration in the SC. The former use models of intersensory discrepancy to monitor changes in perception, whereas the latter do the reverse, namely, they match bimodal inputs in space and/or time. Clearly, much more needs to be done before we can resolve the apparent conflict between the two literatures on multisensory integration. How, for example, can additive models be applied to the superadditive response gains observed in electrophysiology?

**LINEAR INTEGRATION MODELS** A more quantitative approach to the nature of multisensory texture perception has involved the use of additive and weighted averaging models, although no explicit mechanisms are proposed. The general form of such models is one in which $R_{1}$ and $R_{2}$ are associated with different weight parameters to represent the level of compromise between the two modalities, as follows:

$$R_{m} = w_{1}R_{1} + w_{2}R_{2}$$

where $R_{m}$ is the multisensory judgment of texture, $R_{1}$ is the corresponding unimodal response using vision alone, and $R_{2}$ is the corresponding unimodal response using haptics alone; $w_{1}$ and $w_{2}$ are the weights associated with the vision-only and touch-only responses. This is a specific version of the conceptual model in Figure 7.5, where the integration function is linear. The results of Experiments 1–4 by Lederman et al. (1986) could be explained in terms of such a general model, with vision being more strongly weighted than touch for judgments of spatial density, and touch being more strongly weighted than vision for judgments of roughness.

Jones and O'Neil (1985: their Equation 2) have derived a complete dominance process from the general model just presented, in which less efficiently processed information from one modality is totally ignored. The following equation is based on their forced-choice task. Hence $P$ replaces the more general $R$ in Equation 1 to represent probability responses:

$$P_{UV} = \max P$$

where $max P$ is the larger of $P_{1}$ or $P_{2}$. Again, this is a specific version of the conceptual model in Figure 7.5, but in this situation, there is linear averaging where the weight for the lesser of $P_{1}$ or $P_{2}$ is set to zero and the weight for the larger is set to 1. The only empirical work that could be viewed as supporting a total dominance model of multisensory roughness perception is by Lederman (1979) and by Heller (1982, Experiment 3). In Lederman's study, when estimating the magnitude of perceived roughness, subjects ignored the sounds produced by touching textured surfaces with their bare fingers and focused solely on the associated haptic cues. In Heller’s study, subjects ignored the cues to visual texture in the bimodal condition.

Jones and O'Neil (1985, their Equation 3) also proposed a very simple version of the general model (Equation 1) in which the bimodal response $P_{UV}$ is modeled as the arithmetic mean of the two inputs:

$$P_{UV} = \frac{(P_{1} + P_{2})}{2}$$

Again, this is a linear averaging model where the weights are $1/N$, with $N$ corresponding to the number of modalities. The results of all three experiments in the Jones and O'Neil study (1985) are well described by such a model.

Note that Jones and O'Neil's model of multisensory integration of texture based on data obtained using a texture discrimination paradigm. Anderson (e.g., 1974) has offered a different experimental paradigm for formally evaluating the integration of two sources, known as the functional measure- ment approach. The general model used to represent the integration of two information sources is given by:

$$R_{m} = w_{1}R_{1} + w_{2}R_{2}$$

where $R_{m}$ represents the mean rating scale response to the combination of two stimuli, which may or may not be the same. In a more restrictive variant model, the weights are constrained to sum to 1. A simple empirical test of the averaging model requires subjects to produce a combined judgment of two sensory inputs with a rating scale. As described earlier under Empirical Findings, the sensory inputs are obtained from pairs of stimuli that vary on two dimensions and are presented in all factorial combinations.

Jones and O'Neil (1985) suggested that Anderson's method might also be extended to the study of
maxisensory integration of roughness by two modalities. In Experiments 5 and 6, Lederman et al. (1986) formally adopted Anderson’s empirical test (described earlier) as a complementary method for studying multisensory texture perception. The results converged with those obtained in Experiments 1–4 and provided support for weighted averaging models of bimodal roughness perception and of bimodal spatial density perception. The weights for touch and vision in both tasks summed to 1, thus supporting a weighted averaging model as opposed to an additive model. Notably, however, the relative magnitude of the two weights reversed with perceptual task. That is, the touch weight was larger than the vision weight in the roughness task, but larger for vision than for touch in the spatial density task. The details concerning how the relative weighting between haptics and vision was calculated in the functional measurement analysis are beyond the scope of this chapter but are detailed in Lederman et al. (1986).

For summary purposes, we have indicated in Table 7.1 which modality obtained the stronger weighting.

Weighted averaging models quantitatively represent the bimodal response as a compromise between the two modalities; however, they do not specify how the relative weighting that defines the nature of the compromise is derived (for further consideration, see, e.g., Mulligan & Shaw, 1980; Shaw, 1982).

In the remaining subsections, the models are not simple linear models.

**Maximum-Likelihood Integration Model.** Very recently, Ernst and Banks (2001) hypothesized that the observer determines the modality weights by operating as a threshold integrator to minimize uncertainty. Such a process would produce a multisensory response with a variance that is lower than either unimodal estimator, assuming both unimodal estimators are finite. The maximum-likelihood integration model successfully predicted visual dominance of haptics on a task that required subjects to judge the height of a virtual block that was presented visually, haptically, or bimodally. Variances were obtained in the two unimodal conditions and used to predict the bimodal response based on a maximum-likelihood integration solution. Dominance of the haptic system by vision was predicted on the basis of lower visual than haptic variance. The magnitude of visual dominance also declined progressively, with increasing variability in the visual responses that was effected through deliberate experimental manipulation. This model may predict the relatively equal weighting between vision and touch in the multisensory texture discrepancy experiment conducted by Lederman and Abbott (1981). Note, however, that this model is constrained by the need to use perceptual tasks that permit the calculation of a relative threshold (i.e., just noticeable difference).

**Other Models** In addition to weighted averaging models, Massaro and Friedman (1990) have considered a number of alternative models of information integration. Of these, fuzzy logical and constraint models offer two additional theoretical approaches to modeling multisensory texture perception. To end this section, we will consider each of them briefly.

Massaro and Friedman (1990) have described the fuzzy logical model as involving three processing operations during pattern recognition tasks: feature evaluation, feature integration, and decision. "Continuously valued features are evaluated, integrated, and matched against prototype descriptions in memory, and an identification decision is made on the basis of the relative goodness of match of the stimulus information with the relevant prototype descriptions" (p. 231). To permit comparison of the extent to which each feature matches a prototype in the feature evaluation stage, fuzzy truth values are used. These may vary between 0 (completely false) and 1 (completely true), and can therefore represent categorical or continuous information. They can also be used to represent different sources of information, as is relevant to our current discussion of multisensory integration. A value of 0.5 indicates a completely ambiguous situation. The second integration stage provides a multiplicative combination of truth values. The final decision stage involves pattern classification and requires the application of a relative-goodness rule (pp. 230–231). Because not all of the perceptual tasks used to study multisensory texture perception to date have involved pattern recognition, the third process may not always be required (e.g., forced-choice and magnitude-estimation tasks); however, relative goodness would presumably be used in tasks such as similarity sorting (Klaczynski et al., 1987; Lederman et al., 1990).

Massaro and Friedman (1990) also propose a "two-layer (i.e., input, output) connectionist" approach to modeling information integration that may be usefully applied to the problem of multisensory integration in general and, more specifically, to multisensory texture perception. Each input unit is connected to each output unit (e.g., unique roughness responses). The activation of an output unit is determined by the product of the input activation and the specific weight associated with that input-output connection. The total activation of a specific output unit (i.e., the integration stage) involves the summation of the separate activations to a
given output unit passed through a sigmoid-squashing function (Rumelhart, Hinton, & Williams, 1986). The activations at the output layer must then be mapped into a response, usually according to a relative-goodness rule. The decision follows this rule.

Masaar and Cohen (1987) have observed that although the fuzzy logical and connectionist models are expressed in different conceptual frameworks, when there are only two response alternatives, they make similar predictions (for further discussion, see Moelvann & McGeehan, 2001).

One final group of models involving signal detection theory may prove relevant to issues pertaining to how stimuli impact are combined. For example, in Ashby and Townsend (1966), Ashby and Maddox (1994), and Masaar and Friedman (1990).

Other material properties

Although this chapter has focused primarily on the multisensory perception of surface texture, the material used to construct objects and surfaces varies in other ways as well. In this section, therefore, we consider what is known about the multisensory perception of hardness, softness, and stiffness, the only other material properties that have been formally investigated. We know of no work that has addressed the multisensory processing of thermal properties such as temperature or thermal conductivity. Presumably visual cues would play a considerably lesser role under such circumstances, except under conditions of extreme heat, in which the visual cues are informative.

Softness and Hardness

Bimbs (1987) performed a study in which a number of highly trained and untrained groups of subjects ranked six different grades of wool fibers multiple times, both visually in terms of "fineness" (presumed to be determined by the apparent diameters of the fibers) and haptically in terms of "softness" (presumed to be determined by the relative softness or evenness of groups of fibers explored by hand). He found that people were remarkably proficient at this task using either modality. Very few mistakes were made; moreover, the visual and haptic judgments correlated very highly both with each other and with the professional standard ranking scale. (Unfortunately, Bimbs provided no numerical values for the correlations in this paper.)

Stiffness

In a series of studies, Srinivasan and his colleagues (DiFranco, Beauregard, & Srinivasan, 1997; Srinivasan, Beauregard, & Broek, 1996; Wu, Bastoglu, & Srinivasan, 1999) have investigated the role of cross-modal sources of sensory information pertaining to the perceived stiffness of deformable objects in virtual environments. Srinivasan et al. (1996) pitted vision against kineses in a two-alternative forced-choice task in which subjects were required to choose which of two virtual springs was "stiffer." Subjects felt the stiffness of each spring by pushing a button knob on a device that provided the subject with force profiles of virtual springs which they watched on a computer screen. The images that graphically represented the corresponding spring compressions. The discrepancy between the actual deformation displayed on the computer monitor (visual cues) was varied across trials. Values ranged from zero (i.e., the actual and visual deformation signals were in complete registration) to complete reversal (i.e., the physically softer spring looked like the harder spring, and vice versa). The results indicated that the visual system strongly dominated the kinesthetically sensed position of the hand. The authors concluded that subjects ignored the latter information about spring deformation and based their judgments on the relationship between the visual position information and the indentation force sensed tactually. Although some may argue that perceiving spring tension should not be regarded as a property of texture, the latter term is not so well defined in the literature that we feel justified in omitting this set of provocative studies.

In the next study in this series, DiFranco et al. (1997) considered the influence of auditory cues on the haptic perception of the stiffness of virtual springs. In the first experiment, subjects tapped on virtual surfaces while hearing various pre-recorded auditory feedback sounds that represented compliant and rigid surfaces (e.g., polyfoam, metal). Subjects were asked to rank the virtual surfaces in terms of perceived stiffness. They were unaware that the physical stiffness was uncorrelated with the sounds. The results indicated that in general, when the surface was paired with the sounds associated with tapping more rigid surfaces, it was ranked as being stiffer than when it was paired with sounds associated with tapping more compliant surfaces. In a second experiment, subjects were able to rank-order fairly well the perceived stiffness of five different surfaces in the absence of auditory cues. In a third experiment, a selection of five sounds from Experiment 1 were paired with each of the five haptic stiffness values used in Experiment 2. The results obtained by DiFranco et al. are in keeping with the demonstrated influence of auditory cues on the remote perception of actual surface
roughness with a rigid probe (Lederman et al., 2002). The ordinary cues inferred from subjects' judgments of stiffness, particularly for those experimentally naive subjects who did not participate in Experiment 2 (haptics only). Overall, the influence of auditory cues was somewhat less strong than that of vision. The authors suggest that the reason may be that the visual graphics cues on the monitor provided excellent spatial information about the changing position of the stylus, whereas the auditory cues did not.

Continuing their investigation of the multisensory perception of stiffness in a virtual environment, Wu et al. (1999) showed that when only haptic feedback was provided, subjects judged compliant objects that were further away as being softer than those that were presented closer to them. However, the addition of visual perspective feedback served both to reduce this form of haptic distortion and to improve resolution, thus demonstrating two advantages to representing visual stiffness bimodally (see Table 7.2).

Practical implications of scientific research on multisensory texture perception

The scientific results from the study of intersensory texture perception are critically important to those who wish to design multisensory systems for a variety of applications, including, for example, teleoperation tasks and virtual environments that can be seen as well as heard and felt. Lederman and his colleagues have highlighted the value of creating cross-modal displays to overcome the current limitations in the haptic display hardware used to deliver force and cutaneous feedback when these displays are used in virtual environments. Potentially exciting medical applications include using multisensory interfaces to display texture and other material properties (hardness/softness, stiffness, thermal properties) during remote dermatological diagnosis, with the properties conveyed to the visual and haptic systems complementing one another. Another powerful application involving the display of surface texture and other material properties includes the use of virtual training systems to teach novice surgeons complex surgical procedures as tissue compliance and texture are important diagnostic predictors. Noice dermatologists could also benefit from learning dermatological diagnostic procedures by working with virtual diseased skin models. And on the business side, e-commerce offers yet another promising application in which the user could explore the properties of fabrics and other products via multisensory interfaces.

REFERENCES


