Integrating Multimodal Information about Surface Texture via a Probe: Relative Contributions of Haptic and Touch-Produced Sound Sources

S.J. Lederman, R.L. Klatzky, T. Morgan & C. Hamilton

Dept of Psychology, Queen's University, Kingston, Ontario Canada
Dept. of Psychology, Carnegie Mellon University, Pittsburgh, PA, USA

Abstract

We experimentally assessed the relative contributions of tactual and auditory information to multisensory (i.e., bimodal) judgments of surface roughness using a rigid probe. Participants judged the magnitude of surface roughness and their corresponding confidence in three modality conditions: touch-only, audition-only (i.e., touch-produced sounds only) and touch+audition. The results indicated that touch cues were weighted 62% and auditory cues 38% in the bimodal judgments. Participants also proved to be more confident of their judgments in the bimodal condition. Implications for the creation of virtual roughness presented through uni- vs. multimodal interfaces is also addressed.

1. Introduction

Despite the fact that perception and performance most often involve more than one modality, until relatively recently our perceptual systems have traditionally been studied in isolation. In this paper, we focus on the extent to which perceptual events that occur in one modality are affected by the simultaneous stimulation of at least one other modality. Reviews by Welch and Warren [23], Stein & Meredith [19], and Driver and Spence [1] provide valuable summaries of the literature on intersensory interaction.

Inasmuch as multiple sources of information about an event often provide redundant cues, it is not always possible to determine the relative extent to which each modality influences the subject's judgment about that event. To overcome this ambiguity, some researchers have used the “perceptual discrepancy” paradigm -- information to one modality is artificially distorted (e.g., via optical distortion devices), thereby creating a perceptual discrepancy between two modalities. The distortion serves to “perceptually tag” them, so that it is possible to disambiguate their relative contributions. The methodology has been a favored and effective experimental technique for those investigating the relative dominance of vision, touch, audition, kinesthesia and the vestibular system with respect to judgments of spatial events. Overall, the visual system has been shown to dominate the other systems, either totally or very strongly for spatial tasks (egocentric location: e.g., [3]; object shape and size: e.g., [18]; orientation: [17]). Conversely, when judgments of temporal events are required, the auditory system tends to dominate (temporal duration: e.g., [20]; temporal rate: e.g., [14]; temporal pattern: e.g., [15]).

The focus of the current research is on the perception of surface texture, which potentially includes contributions by three different perceptual systems --touch, vision and audition. Not only can one feel and see a textured surface; potentially, one can also judge the textural properties using the concomitant sounds that are produced as a person manually explores the surface.

1.1. Touch vs. vision

A perceptual-discrepancy paradigm was used in a texture perception study by Lederman and Abbott [9]. They presented subjects with two distinctly different abrasive surfaces positioned one in front of the other, such that the bottom edge of the visually examined texture lay more distal, but adjacent, to the far edge of the haptically examined texture. Subjects were led to believe that the two surfaces actually comprised a single, homogeneous surface that extended out in front of them. They performed a match-to-sample task in which they were instructed to select from among a number of comparison surfaces the “texture” that best matched the standard texture. The task was repeated in two control conditions (vision only, touch only), involving pairs of surfaces that were in fact identical to the vision-only and touch-only stimuli presented together in the bimodal discrepancy condition. In both conditions, the same modality was used during the initial stimulus presentation and the subsequent response phase. They weighted the visual and tactual
1.3. Feeling textures with a probe

The current study addressed the same multisensory issues concerning sensory dominance and intersensory integration. However, this time the observer explored surface textures remotely using a probe as an end effector rather than the bare fingers. Inasmuch as the sounds generated by contact between a rigid probe and a rigid surface are considerably louder than those generated by a bare finger, we anticipated that the auditory-based cues to texture could well play a more significant role in judging surface texture. In this study, observers judged the roughness magnitude of raised-dot surfaces in three modality conditions: touch-only, audition-only, and touch+audition.

2. Method

2.1. Participants

Twenty-four undergraduate students from the introductory-psychology subject pool volunteered for grade credit (17 females, 7 males). The mean age was 18.8 years (SD = 0.6). All participants reported themselves as being right-handed, and as having both normal hearing and normal tactual/motoric capabilities in their hands.

2.2 Force-control apparatus, stylus, and stimulus surfaces

To maintain applied force constant throughout the experiment, a balance apparatus was used to present the stimulus surfaces to the participants [11]. Lederman [8] employed the same apparatus in her study on the role of touch-produced sound cues in the perception of the magnitude of surface roughness with the bare fingers. The force was maintained by initially balancing the balance arm with equal weights at both ends. A separate 0.29N weight was included at the stimulus end, so that when it was removed, an equivalent upward force would be applied at that end. To achieve the desired applied finger force, the participant was instructed to continuously apply
the counter force (i.e., 0.29N) necessary to render the balance arm steady and level. Subjects sat to the left of the apparatus, with sight of the apparatus, their hand, and the stimuli all blocked by the presence of an opaque screen that extended in front of the subject along the side of the apparatus.

The stimuli consisted of a set of plastic polymer plates with raised elements in the form of truncated cones. The element height was .40 mm; the diameter was .46 mm at the top and the average base diameter was 1.05 mm (the base diameters varied somewhat with interelement spacing, due to the production process). A computer program spatially jittered the dots (radially and angularly), while maintaining the original mean interelement spacing. A total of 8 plates were used, with interelement spacing ranging from 0.500 to 3.125 mm, in 0.375 mm increments. The plates are described in detail in Lederman et al. (1999).

The probe was made of delrin plastic. It had a cylindrical shaft with a length of 110 mm and a diameter of 10 mm. The probe terminated in a sphere 3 mm in diameter. The functional contact diameter of the probe tip was measured by inking the tip and rotating it around the contact point on a surface. This functional diameter was 2mm.

In order to keep probe speed relatively constant in all 3 conditions, a computer program was used to display a dot moving 70 mm side to side across a computer monitor (both the experimenter and subject viewed the monitor during all 3 modality conditions). The dot moved at a constant speed of 114.6 mm/s. This speed was selected for purposes of comparing current results with a rigid probe to earlier results (Lederman et al., 1999). Participants were instructed to track the dot's movements with their own. They were taught how to hold the probe, and to make reversing lateral motions of about 60 mm across the plate surface.

For the touch-only condition, headphones, wax earplugs and an audiotape with background masking noise were used in order to eliminate both ambient noises in the room and the sounds produced by touching the surfaces with the probe. The specifics of the masking noise are detailed in Lederman et al. (1999).

2.3. Experimental design

A 3x8 within-subject design was used. The first factor was modality, with 3 levels (touch-only, audition-only, and touch + audition). The second factor was interelement spacing, with 8 levels corresponding to the different stimulus plates. Each interelement spacing was presented twice. The interelement spacing x repetition combinations were blocked within modality, with all 16 conditions randomly presented in the touch-only condition, then in a different random order in the audition-only condition, and so forth. The order in which the three modality conditions were presented was completely counterbalanced across groups of 6 participants.

2.4. Procedure

Participants were instructed to hold the probe between their index finger and thumb like a pencil, with the butt of the probe resting within the crook between their thumb and index finger. They were seated on a stool, to the left of the apparatus facing the front. Their vision was blocked in all conditions by an occluding screen placed between the subject and the apparatus. The stimuli were presented one at a time. Participants estimated roughness magnitude by moving the probe back and forth across the stimulus surface, while applying sufficient force to keep the balance arm steady and level.

Roughness magnitude was estimated using an absolute magnitude-estimation procedure (for details see [10]). Participants were instructed to assign the non-zero, positive number that best described the magnitude of the perceived roughness of the stimulus. They were also asked to estimate the confidence of their response on a 5-point scale, with 1 = very unconfident and 5 = very confident about their magnitude-estimation judgement. All participants received initial training with the dot-tracking program to help them learn to maintain their probe motion at a constant speed across all trials. As previously discussed, force was held constant throughout.

For the touch + audition condition, participants could both feel and hear the textured surfaces as they examined them with the probe. In the touch-only condition, participants could only feel the surfaces. They were unable to hear the touch-produced sounds because of the earplugs they wore and because of the masking noise played through headphones. In the audition-only condition, the experimenter produced the touch-produced sound cues by moving the rigid probe across the plates in the same manner as the participants. The latter were instructed to report how rough the surfaces "seemed" (touch + audition), “felt” (touch only) or “sounded” (audition only), and their associated numeric confidence ratings. After each block of trials with a given modality condition, the participant was given a brief break, followed by additional practice with the next condition before continuing with the experiment. The experiment lasted approximately 45 minutes for each participant.

3. Results

3.1. Magnitude estimates

An initial 3-factor repeated measures ANOVA was performed on the raw magnitude estimation data, with 3
factors: repetition (2 levels), modality (3 levels) and interelement spacing (8 levels). Neither the main effect of repetition nor any of the interaction with repetition was statistically significant. Therefore, the magnitude estimates were averaged across the two repetitions. Next, to control for differences in the numerical scales used, the resulting mean magnitude estimates for each participant were subsequently normalized by dividing each score by the individual participant mean, then multiplying by the grand mean (across participants). Finally, the normalized data were then logarithmically transformed (base 10) to achieve a more normal distribution.

A within-subject ANOVA was conducted on these logarithmically transformed normalized magnitude estimates, with two factors: modality and interelement spacing. In describing the results of this and all future ANOVAs, only the significant main effects and interactions will usually be considered; the Greenhouse-Geisser adjusted degrees of freedom are reported.

### 3.2. Effect of modality

The main effect of modality was highly significant, $F(1.2, 26.9) = 36.7, p < 0.0001$: the touch-only condition produced higher roughness estimates than did either the audition-only or the touch + audition conditions. The touch + audition mean lay between of the two control conditions.

![Figure 1. Log$_{10}$ normalized roughness magnitude estimates as a function of log interelement spacing (mm) for three modality conditions. The functions have been fit with quadratic equations.](image)

Dependent t-tests of the means showed that all three conditions (touch-only, audition-only, and touch + audition) differed significantly from one another (audition-only vs. touch-only: $t(23) = -6.54, p < 0.001$; audition-only vs. touch + audition: $t(23) = -5.23, p < 0.001$; touch-only vs. touch + audition: $t(23) = 6.19, p < 0.001$).

The means of the log normalized magnitude estimates (averaged across participants) are plotted as a function of log interelement spacing for each modality condition (touch-only, audition-only, and touch + audition) in Figure 1.

By collapsing the magnitude estimates across participant and interelement spacing, and then comparing across modality, it was possible to calculate a measure of the relative contributions of touch and audition to the bimodal estimates of roughness. (We used the overall means for the three modality conditions, since the interaction between modality and interelement spacing was not statistically significant.) We calculated the following statistic, which indicates the percent weighting of the tactile information in the bimodal estimates. It was calculated as in (1), by determining the distance between the touch + audition condition and the audition-only control, as a percentage of the distance between the two control conditions.

$$\text{% T dominance} = \left(\frac{\text{Mean}_{\text{Touch+Audition}} - \text{Mean}_{\text{Audition-only}}}{\text{Mean}_{\text{Touch-only}} - \text{Mean}_{\text{Audition-only}}}\right) \times 100 \quad (1)$$

A value of 100% would indicate total tactual dominance; a value of 0% would indicate total auditory dominance. The relative weighting by touch was 62%; accordingly, the relative weighting by audition was 38%. Two dependent t-tests were performed on the % T dominance values with subject as the unit of observation. To determine if dominance by touch was total, the mean obtained value was tested against the null hypothesis that tactual dominance equals 100%. The null hypothesis was rejected ($t(23) = -6.45; p < 0.001$; 1-tailed); thus, touch-produced sounds statistically contributed to subjects' roughness judgments. To determine if the tactual information and the auditory information were equally weighted, the obtained value was tested against the null hypothesis that tactual dominance equals 50%. The results indicated that the difference just attained significance at the .05 level $t(23) = 2.08; p = .05$; 1-tailed); thus, tactual information was weighted more heavily than auditory information. (The results were very similar when we applied an arcsin transformation on the tactual dominance scores. However, because the latter tests involved an additional transformation, we only report the results of the former t-tests here). The % T dominance values for individual subjects were also informative: 6 subjects showed very strong tactual dominance (i.e., $\geq 80\%$), while
3 subjects showed very strong auditory dominance (i.e., % T values ≤ 20%). Of the remaining 16 subjects, 11 showed varying levels of tactile dominance (i.e., 51-79%), and 4 subjects showed varying levels of auditory dominance (21 - 49%). In summary, the data indicate that touch is weighted more strongly than the touch-produced sounds by 17 of 24 subjects. However, the touch-produced sounds also strongly contributed to the bimodal judgments of roughness.

3.3. Effect of interelement spacing

There was a significant effect of interelement spacing, $F(1.7, 39.2) = 145.8, p < 0.001$: roughness magnitude increased with increasing interelement spacing up to a peak; it subsequently declined with further increases in interelement spacing.

As we have consistently found that quadratic equations fit the psychophysical functions obtained with rigid probes better than linear equations, we applied quadratic fits. The $r^2$ values for the touch-only, audition-only, and touch + audition conditions were 0.93, 0.90 and 0.94, respectively.

To further analyze the effects of exploration modality on the quadratic functions, the original $a$, $b$, $c$ parameters in the quadratic function $y = ax^2 + bx + c$ were transformed into a new equation, $y = a(x-x_0)^2 + h$, in which $a$ represents the degree of curvature of the quadratic function, $x_0$ represents the position of the peak along the interelement spacing (x) axis, and $h$ represents the position of the peak along the perceived-roughness (y) axis. Parameter $a$ remains the same in both equations. In the new parameterizations, $x_0 = -b/2a$, and $h = c - (b^2/4a)$. Assuming that the underlying perceptual outcome (log values) is a quadratic function of log spacing, our data-normalization process will preserve the parameters $a$ and $x_0$, as well as the relative value of $h$ across conditions. We assume further that subjects differ by a multiplier that is applied to the output of the underlying sensory function; across different conditions, the quadratic function representing the sensory response changes, but subjects maintain the same multipliers. If the multiplier for a subject varied across conditions, $a$ and $x_0$ would be preserved, but not the relative height across conditions. The functions peaked between 2.4 and 2.75 mm along the interelement axis.

Individual ANOVAs were run on each parameter (with subject as the unit of observation) to determine how the parameters of the quadratic functions were altered as a result of the perceptual modality used to judge surface roughness. Only parameter $h$ changed significantly $F(1.1, 26.1) = 31.1, p < .001$: the touch-only function had the highest peak value on the y-axis, followed by the touch + audition function, and then the audition-only function. Pairwise comparisons confirmed that the three functions were all significantly different from one another (audition-only vs. touch-only: $t(23) = -6.0, p < .001$; audition-only vs. touch + audition: $t(23) = -4.7, p < .001$; touch-only vs. touch + audition: $t(23) = 6.6, p < .001$). The mean values for parameters $a$, $x_0$, and $h$ are listed in Table 1.

### Table 1. Mean $a$, $x_0$, and $h$ parameter values (log units) for the transformed quadratic equations.

<table>
<thead>
<tr>
<th>Modality</th>
<th>Quad. parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$a$</td>
</tr>
<tr>
<td>Touch-only</td>
<td>-1.25</td>
</tr>
<tr>
<td>Touch + Audition</td>
<td>-1.31</td>
</tr>
<tr>
<td>Audition-only</td>
<td>-1.54</td>
</tr>
</tbody>
</table>

3.4. Confidence estimates

A within-subjects ANOVA was also conducted on the confidence estimates (averaged across repetitions) with 2 factors: modality (3 levels), and interelement spacing (8 levels). Although the main effect of modality failed to reach significance, we note an interesting trend: the mean confidence ratings in the Touch + Audition condition were higher than the corresponding mean of the two unimodal controls. A non-parametric Sign test indicated that this pattern was statistically significant ($p < .05$). The effect of interelement spacing, however, was significant $F(4.40, 101.08) = 13.99, p < 0.001$: confidence ratings tended to decline with increasing interelement spacing, but particularly for the two narrowest interelement spacings; with further increases in interelement spacing, confidence leveled off.

These patterns are evident in Figure 2, in which the confidence ratings (collapsed across participants and repetitions) are plotted as a function of interelement spacing for the two unimodal and the one bimodal conditions. The two narrowest plates were judged with the greatest confidence, followed by the wider plates, which were all judged with lower confidence. Note that the confidence ratings do not replicate the quadratic trend in the magnitude estimates, indicating that confidence is not directly derived from perceived roughness.
4. Discussion

The earlier study by Lederman (1979) showed that concomitant touch-based auditory cues produced by the bare finger played no role in estimating the roughness magnitude of linear incised gratings. More specifically, the perceptual estimates of surface roughness derived from auditory, as opposed to touch, cues differed. Moreover, the touch + audition estimates were identical to the corresponding touch-only estimates. Therefore, the sense of touch dominated the sense of audition completely, i.e., by 100%.

The current study similarly assessed the relative use of concomitant touch and touch-produced auditory information to judge surface roughness. This time, however, contact was effected via a rigid probe. With rigid contact between surface and end effector, the amplitude of the accompanying sounds is usually considerably greater. We speculated that under such circumstances, observers may use the touch-produced sounds as well to judge roughness. Like Lederman (1979), we found that the tactual and auditory estimates were significantly different from one another -- the tactual roughness estimates were consistently greater than the corresponding auditory roughness estimates. Such a difference created the necessary perceptual discrepancy -- a natural one -- between the two sensory modalities that in turn allowed us to determine the relative contribution of each modality to bimodal estimates of surface roughness. Contrary to Lederman's findings with the bare finger (1979), the current results indicated that when observers explored the surfaces with a rigid probe, they used both tactual and auditory information to make their estimates. The estimate of % tactual dominance indicated that the tactual and auditory sources were weighted 62% and 38%, respectively.

The quadratic nature of the touch-only psychophysical function (log-log values) confirms earlier results on remote touch with a rigid probe (e.g., Klatzky & Lederman 1999; Lederman, Klatzky, Hamilton and Ramsay, 1999). We include here the new finding that a quadratic equation best describes the psychophysical functions for both the audition-only and bimodal conditions as well. That the curvature and peak position on the x-axis are the same for all three modality conditions indicates that for audition and touch the processing algorithms, based on a common vibratory input, are not fundamentally different. They differed only with respect to the relative heights of the peaks along the y-axis (parameter h) -- the touch-only peak was highest, followed in turn by the touch + audition and then the audition-only functions. In a separate experiment not reported here, we found that amplifying the touch-produced sounds, generated by participants as they touched the plates, raised the height of the bimodal psychophysical function for roughness. The relative height of the auditory and touch curves likely reflects the amplitude of the sounds as they reach the ears. We are currently investigating whether the current weighting of the auditory and tactual inputs could change if the sounds in the auditory-only condition are amplified.

4.1. Designing virtual textures for multimodal interfaces

The psychophysical data obtained in the current experiment also provide information relevant to the design of multimodal interfaces for teleoperation and virtual environment applications.

Previous studies on feeling textures through a probe in our labs (e.g., [6], [10]) have focused on the general nature of the psychophysical function that best describes the haptically derived roughness magnitudes as a function of interelement spacing (log/log scales). The results have repeatedly confirmed that a quadratic function best describes the data. The current paper has extended this work by determining the nature of the psychophysical function that best describes roughness judgments based on a second potential source of sensory information, namely the concomitant sounds that are produced by and that normally accompany remote haptic exploration via a stylus. It is significant that a quadratic function best describes both the auditory and the touch+auditory judgments as well. Identical values for all three modality conditions were obtained for curvature and for peak
position along the interelement spacing axis, peak height differed for the two unimodal conditions with the touch+audition peak lying in between. We suggest that the quadratically shaped functions and corresponding parameter values for tactile and auditory conditions are determined by the mechanical interactions between the rigid probe and the textured surface that alter the vibratory signals reaching both hand and ear. As long as the probe diameter is wider than the interelement gap, it will ride along the tops of the raised elements. As the interelement spacing increases, the probe diameter will become narrower than the gap; at this point, the probe will drop into the gap, and may even ride along the smooth surface floor. We believe the position of the quadratic peak along the interelement spacing axis is determined by the point at which the probe drops down to the base. Such information will prove critical if one aims to use a mechanical model of the contact interactions to direct the choice of algorithms for rendering vibration-based textures simultaneously available to hand and ear.

At least potentially, texture perception is quintessentially multimodal -- information about surface textures is available to three different sensory systems, haptics, vision and audition. Might therefore multimodal interfaces (in this case, haptic and auditory) effect more realistic remote texture environments than unimodal interfaces? Before answering this question, we must first determine whether observers actually attend to multisensory cues to surface texture. Previous research [8] has indicated that observers ignore the touch-produced sounds that accompany the haptic inputs obtained with a bare finger. However, this may well be due to the fact that such auditory cues are typically of relatively low amplitude; hence, they are often masked by other more intense extraneous sounds in the environment. In striking contrast, the current study has shown that when a rigid stylus is used to effect remote contact with a textured surface, observers clearly do incorporate the touch-produced auditory information into their perceptual representations of the roughness magnitude. (We will provide additional converging experimental evidence to this effect at the conference.) We have previously argued that feeling textures remotely with a spherical-tip stylus serves as an elegant yet simple model for effecting contact by means of point-contact haptic interfaces. Our data suggest that touch-produced sounds may constitute a second valuable source of vibration-based inputs for rendering virtual textured environments.

In considering whether it would be beneficial to develop algorithms for creating virtual textures based on multimodal (as opposed to unimodal) cues, it is further necessary to examine relative performance under unimodal and bimodal conditions. The current experiment was not explicitly designed to do so, although we did in fact observe an interesting trend in the estimated confidence with which subjects made their magnitude-estimate judgments of roughness. The mean confidence ratings for the bimodal condition were consistently higher than either unimodal condition. While this pattern is suggestive with respect to the design of multimodal interfaces, it is inadvisable to place too much faith in this measure, given that estimating confidence proved to be a difficult task for subjects to perform. We are therefore currently conducting experiments that will directly assess the consequences for relative performance of using bimodal versus unimodal sensory inputs to perform texture perception tasks.

5. References


Acknowledgements. We thank Gillian Ramsay for her valuable suggestions. This research was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) and by the Institute for Robotics and Intelligent Systems (IRIS), a Federal Center of Excellence of Canada.