Journal of Environmental Psychology (1982) **2**, 275–284

THE RELATIVE CONTRIBUTION OF VISUAL AND AUDITORY CUES TO ENVIRONMENTAL PERCEPTION

ROBERT GIFFORD AND CHEUK FAN NG
University of Victoria, British Columbia, Canada

Abstract

This study assesses the relative dominance of vision and hearing in judgments of environments, and examines whether the sound sensitivity of the judge alters relative sensory dominance. Six sets of four composite (auditory-plus-visual) videotaped environmental displays were created from unisensory visual and auditory components with known levels of evoked affect. These displays, in a 2 (high, low) × 2 (high, low) Pleasure and/or Arousal format, were presented to 70 participants. The results indicate that visual cues dominate auditory cues in judgments of Pleasure, but that the degree of dominance is highly variable. Arousal judgments were not consistently related to visual or auditory cues. Sound sensitivity does not appear to influence the relative dominance of vision.

Introduction

Craik (1968) pointed out that the site visit, as a mode of presenting environments, has the advantage of conveying information about the environmental display through all sense modalities. Most everyday person–environment transactions occur on-site; perception and evaluation of settings usually occurs when information from all sense modalities is available. However, many studies of environmental perception have restricted participants to visual-only access to the setting. These visual stimuli most often are slides (e.g. Calvin *et al.*, 1972), photographs (e.g. Peterson, 1967), drawings (e.g. Canter, 1969) or models (e.g. Sorte, 1975).

Reviews of simulation techniques (e.g. Mandel *et al.*, 1978, p. 382) have cautioned that such visual-only presentations of setting may lack external validity; judgments obtained from such unisensory presentation may not agree very well with multisensory, on-site judgments. However, empirical studies (e.g. Seaton and Collins, 1972) have reported that various visual-only presentations do represent on-site judgments quite well. One inference from the Seaton and Collins data is that visual cues so dominate auditory cues that little distortion of judgments based on visual-plus-auditory displays will occur when visual-only displays are offered. The present study investigates the relative contribution of visual and auditory cues to the perception of the environment.

The tendency of visual information to dominate perceptual judgment has been well-documented in traditional laboratory studies (e.g. Pick *et al.*, 1969). When visual and auditory signals are presented simultaneously, subjects generally respond to the visual input and are often not even aware of the occurrence of the auditory input (Colavita, 1974). Both these traditional findings and those of Seaton and Collins

Reprints are available on request from the senior author, at the Department of Psychology, University of Victoria, Victoria, British Columbia, Canada V8W 2Y2

0272–4944/82/040275 + 10 $03.00/0

suggest that visual cues largely outweigh other sensory cues in determining judgments of settings.

One promising approach to the problem systematically examines the effect on judgments of multisensory stimuli that have been constructed from separate unisensory stimuli with known effects. In an example from the social-nonverbal literature, Mehrabian and Ferris (1967) investigated the decoding of consistent and inconsistent communication of attitude in facial and vocal channels. Three degrees of attitude (positive, neutral and negative) communicated visually, i.e. in facial expressions, were combined with the same three degrees of attitude communicated aurally, i.e. in vocal expressions. The results appeared to indicate that attitude inferred from composite (facial-plus-vocal) attitude communication is a linear function of the attitude communicated in each component, with the visual component having approximately one and a half times the influence of the auditory component.

Attempts to study the contribution of various sense modalities to environmental perception are scarce, but in a study by Faletti and Wellans (1979), evaluative judgments of residential environments were elicited by presenting subjects with visual-only and auditory-only cues. Subsequently, positively and negatively evaluated visual-only cues (levels of housing density) were combined with positively and negatively evaluated auditory-only cues (levels of mechanical sounds) to form four composite multisensory stimuli in a 2 × 2 matrix. Subjects were asked to judge the composite stimuli; the data analysis was directed toward determining whether judgments of the composites were more influenced by the change in visual cue pleasantness or the change in auditory cue pleasantness. Faletti and Wellans claim their results show that visual information contributes more to the overall impression than does auditory information, and set the ratio of influence at 0·85 to 0·15.

However, there are a number of problems with the Faletti and Wellans study. First, they misquote Mehrabian and Ferris (1967) in saying that the latter used a simple weighted averaging formula in which the 'auditory and visual components weighted 0·41 and 0·59, respectively' (pp. 250–1). Next, they proceed to report their own 'analogous formulation' of the relationship between visual and auditory information. This formulation, the derivation or computation of which is not explained, was the vehicle used to arrive at the 0·15/0·85 ratio between auditory and visual information. Third, no matter how the ratio was developed, it cannot be correct since both Mehrabian and Ferris (1967) and Faletti and Wellans (1979) used within-subjects designs. The mean-square error in repeated-measures designs is different for each factor in the ANOVA. Thus, comparing the total sum of squares associated with each factor is meaningless. Finally, although Faletti and Wellans recognize the problem in a footnote, the ratio reported also cannot be correct because the relationship they found was not linear; there was a significant interaction between the auditory and visual effects. In sum, after Mehrabian and Ferris (1967) incorrectly analyzed their data, Faletti and Wellans misquoted the results of that error, repeated the mistake of using a within-subjects design and performed an 'analogous' but unexplained analysis of the data which cannot be correct.

In another approach to the sensory dominance problem in environmental perception, Merrill and Baird (1980) asked participants to tour a campus and its adjacent downtown area. They were instructed to make three judgments: an overall (composite) rating on aesthetic quality of a location, and two separate ratings based on visual stimulation alone and auditory stimulation alone. High correlations among the ratings were reported (visual-composite $r = 0{\cdot}99$; auditory-composite $r = 0{\cdot}90$; and visual-auditory $r = 0{\cdot}84$). It is doubtful whether the subjects in this study could make distinct judgments on the visual or the auditory components alone while simultaneously being exposed to sensory input from all modalities. The high correlations found in this study may well be artifactual.

The relative influence of visual and auditory stimuli on environmental perception remains an unresolved question. The present study attempts to overcome some of the problems inherent in past research in order to provide an indication of the relative role of visual and auditory cues as determinants of environmental perception. Despite the specific limitations discussed earlier, the general paradigm employed by Mehrabian and Ferris and Faletti and Wellans seems preferable to the Merrill and Baird paradigm in that effects due to each sense may be more clearly assessed.

A factor not considered by any of these studies is the possible mediating influence of individual differences. Should present and future research confirm that visual cues dominate auditory cues, the question remains as to whether the ratio is the same, or nearly so, for all individuals. For example, it is possible that persons who are especially sensitive to sound, pleasant or unpleasant, have a tendency to weight auditory cues more strongly than do sound-insensitive individuals. The present study examines whether individuals use different visual-audio weighting ratios as a function of their sensitivity to sound.

Despite the limitations of past studies in environmental perception, evidence from other judgment studies (for a review of nonverbal and person perception work, see Schneider *et al.*, 1979, pp. 129–30) indicates a greater role for visual than for auditory cues. The present hypothesis is, therefore, that visual cues are more influential than auditory cues. The goals of the present study are to (a) confirm or disconfirm the superiority of visual cues, (b) provide an accurate estimate of the ratio of influence between visual and auditory cues, or to document the variability of the ratio and (c) investigate whether individual differences in auditory sensitivity affect the relative weight assigned auditory and visual cues in forming an impression of the environment.

Method

Dependent measures

Ratings of overall aesthetic quality, general preference and the evaluative dimension of the semantic differential were used as dependent measures in the Faletti and Wellans (1979) and Merrill and Baird (1980) studies. However, recent work by Russell and his colleagues (Mehrabian and Russell, 1974; Russell and Pratt, 1980; Ward and Russell, 1981) incorporates and supercedes these older approaches. In the affect-based Russell system, environments are located in a circumplicial space which has two primary dimensions, Pleasure and Arousal. Environments [and other stimuli (Russell, 1980)] evoke emotional states which may be located anywhere in a two-dimensional space described by the Pleasure and Arousal dimensions, which are presumed to be orthogonal.

In the present study, participants judged the environmental displays on simple single-item Pleasure and Arousal scales. A magnitude estimation procedure (S. S. Stevens, 1966) was used to produce ratio-scale measures of Pleasure and Arousal.

The specific procedure is based on a magnitude estimation method developed by J. C. Stevens and Marks (1980) for judging stimuli appropriate to different sense modalities (such as audio-visual presentations of environments) on common measurement scales (such as Pleasure and Arousal).

Sound sensitivity was measured on a self-report scale* adapted and expanded from a scale developed to measure noise sensitivity (Weinstein, 1978).

Design and subjects

The basic unit of the study is a 2 × 2 between-subjects design. The study includes several replications of the basic design, each with Sound Sensitivity as a covariate. For example, one replication involves two levels of visual cues (low Pleasure, high Pleasure) and two levels of auditory cues (low Pleasure, high Pleasure). That is, subjects are presented with one of four combinations: an environmental display with high visual and high auditory Pleasure, or high visual and low auditory Pleasure, or low visual and high auditory Pleasure, or low visual and low auditory Pleasure.

The levels of visual and auditory Pleasure and Arousal are determined by a separate, previous group of raters from the same population. The question is whether Pleasure ratings of the multisensory composites depends more on changes in the level of Pleasure in the visual factor or in the auditory factor. A similar comparison is made for Arousal.

In order to assess the stability of the ratio between auditory and visual cues not only across the two primary affect dimensions (Pleasure and Arousal), but across a range of auditory and visual environmental displays, six replications were employed. In sum, the overall design allowed for a total of 12 2 × 2 comparisons of auditory vs. visual influence on environmental perception: six Pleasure comparisons and six Arousal comparisons.

The subjects were volunteer undergraduate psychology students, 18 who participated in an initial stimulus-development study and 70 who participated in the main part of the study.

Materials and procedure

The environmental displays were developed by making 14 one-minute audio tapes and 14 one-minute video tapes to serve as an initial pool of stimuli. These tapes were made in a city of 250,000 people. The audio tapes included natural as well as mechanical sounds and the visual tapes, which were made with no soundtrack, included scenes of streets and parks.

The 28 stimuli were placed, in a random order, on a single audio-visual tape for presentation. These original 28 stimuli were judged on their Pleasure- and Arousal-eliciting qualities by the initial group of 18 subjects. To ensure that they had a grasp of the concept of ratios, participants were asked to estimate the relative length of four lines of varying lengths at the beginning of the experiment. They were then told to assign the number 10 to the first stimulus on both the Arousal and Pleasure scales and to use it as the standard for comparison with subsequent stimuli. The 28 stimuli were presented to the subjects on a television screen in subdued lighting conditions. When an audio stimulus was to be presented, the tele-

* A copy of the sound sensitivity scale is available from the authors on request.

vision screen was covered, so the 'snow' would not distract them. The participants rated each stimulus immediately after its presentation.

The reliability of these ratings, based on the intra-class correlation (Ebel, 1951), was 0·85 for the whose set of stimuli and 0·84 for the subset of 8 stimuli later chosen for use in the main part of the study. This indicates that there was quite strong agreement among the raters about the level of Pleasure and Arousal elicited by the stimuli.

To create the composite (audio-plus-visual) stimuli for the main study, the two-dimensional array of stimulus medians (see Figure 1) was examined.

The original intention was that the distribution of 28 stimuli medians would include four auditory-plus-visual pairs which would form a neat, equally-spaced

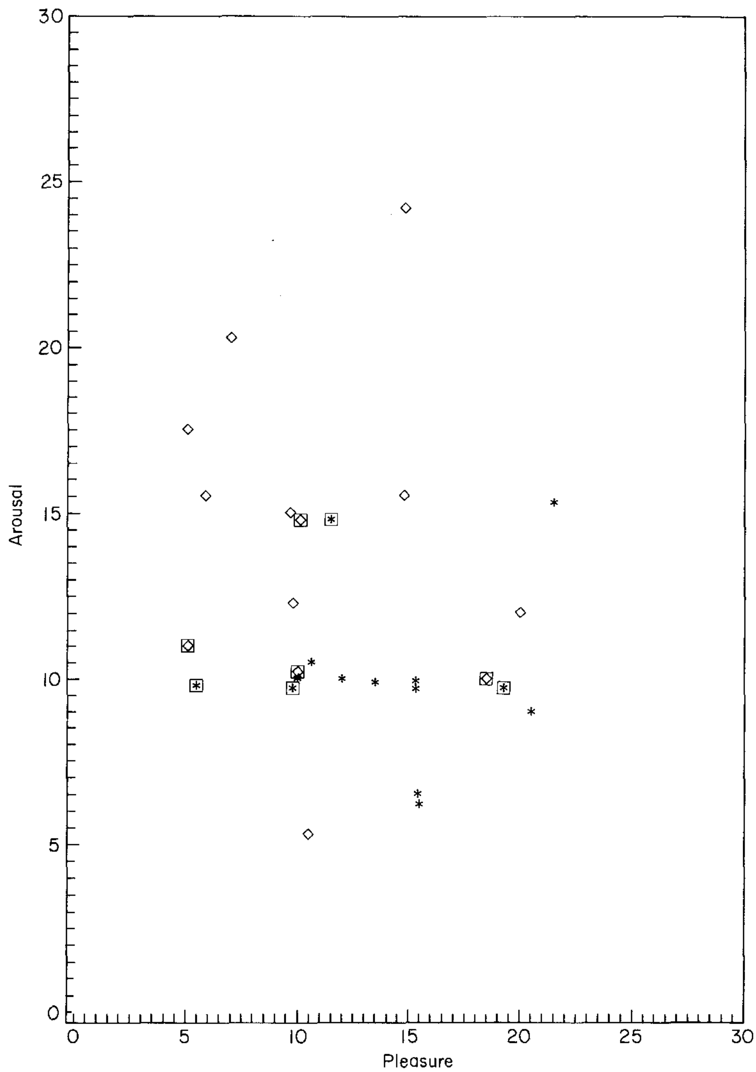

FIGURE 1. Median affect ratings of the unisensory stimuli. * Visual-only stimulus; ◇ audio-only stimulus; □ stimulus used in main study.

square in Figure 1 so that a series of 12 perfect 2 (high, low) × 2 (high, low) designs could be constructed. The medians did not emerge as neatly as desired. However, four well-matched pairs were found in the shape of an inverted T (see Figure 1), permitting five of the six planned Pleasure replications and three of the six planned Arousal replications.

Some of the replications allowed for comparison both of the relative sensory influence on both Pleasure and Arousal judgments in the same 2 × 2 matrix; in others, Pleasure was varied while Arousal was constant for all four cells, or Arousal was varied while Pleasure was constant for all four cells. In total, then, six sets of 2 × 2 composite stimuli were constructed from the 8 unisensory components on the Pleasure and/or Arousal dimensions. For example, two composites were: (a) High Pleasure Audio, Low Pleasure Visual, Low Arousal Audio, Low Arousal Visual and (b) Low Pleasure Audio, Low Pleasure Visual, High Arousal Audio, High Arousal Visual. Each composite was constructed so that the combination of unisensory elements was 'reasonable', i.e., the visual and audio components were not unrealistic when presented together.

Each of the 70 main study participants, who were studied in small groups, were shown six one-minute presentations after being administered the same length-of-line task administered to the initial group of judges. The first display was the same for every subject; it was the standard stimulus used in the preliminary phase of the study and it served the same purpose in the main part of the study. The second display was a dummy stimulus that served merely to familiarize the subjects with the task.

Displays three through six for each subject consisted of four cells from four different replications. That is, each subject saw a composite stimulus representing one cell from each of four (out of the six) replications. Thus, each one of the six replications remained a 2 × 2 between-subjects design, although subjects served in four different replications. Subjects were never presented with the same original unisensory auditory or visual component twice.

The Sound Sensitivity scale was administered after the subjects finished judging their series of four scenes on the Pleasure and Arousal scales. The scale is composed of 30 items, some of which were borrowed with slight adaptations from Weinstein's (1978) noise sensitivity measure and others were designed by the authors to tap sound sensitivity in general, as opposed to noise sensitivity in particular. The items were arranged in a six-point bipolar format, with a response of 1 indicating strong agreement and 6 indicating strong disagreement.

Results

Russell (1980) has hypothesized and demonstrated that Pleasure and Arousal are orthogonal. To determine whether the two dimensions were orthogonal in the present data, the Pearson correlation coefficient between all pairs of Pleasure and Arousal ratings in the main study was computed. The correlation was − 0.039, very nearly the non-relationship hypothesized. Because the two dependent variables were essentially independent, separate analyses of them were undertaken. (If the two dependent variables were significantly correlated, one would be well-advised to analyse them together, as in a MANOVA.)

Data analysis strategy
Partial correlation analysis found no significant correlations between the auditory-only × visual-only interaction and the ratings of the composite displays. Since the interaction was non-significant in all replications, it was not included in the following analyses.

Each of the six replications was analysed separately using a hierarchical multiple regression approach to ANOVA (Cohen and Cohen, 1975). The independent variables were based on the mean ratings of the initial sample of 18 raters, while the dependent variables are based on ratings of the experimental sample of 70 participants. An important reason for choosing the regression approach was that it allows for the actual levels of visual-component and audio-component affect (Figure 1) to be entered in the analysis, rather than dummy coding, which can artificially bifurcate the data. For each replication, the auditory and visual affect levels were entered simultaneously in the attempt to predict the affect evoked by the composite displays. The Sound Sensitivity scores were entered next. The results of the analyses are presented in Table 1.

Arousal
Among the three Arousal replications, one significant relationship between ratings of the composite and the three variables was found. The significant equation indicates that auditory cues were 4·17 times more influential than visual cues (V/A influence ratio of 0·24). However, environment-elicited Arousal was not consistently related to the level of arousal evoked by either the visual cues or the auditory cues or sound sensitivity. As seen in Table 1, a possible contributing factor is that the Arousal

TABLE 1
Composite ratings as a function of auditory and visual cues

	R^2	Visual component†	Auditory component†	Sound Sensitivity†	Ratio of effects (V/A)‡
			Arousal		
1	0·18	0·274	−0·026	−0·322*	
2	0·10	0·154	0·220	+0·110	
3	0·24*	0·218	0·445**	−0·035	0·24 (0·24)
			Pleasure		
1	0·30**	0·525**	−0·164	−0·017	10·24 (10·90)
2	0·41**	0·589**	0·188	0·091	9·82 (10·24)
3	0·65**	0·669**	0·479**	0·157	1·96 (1·96)
4	0·20*	0·357*	−0·081	0·145	19·45 (21·80)
5	0·47**	0·591**	0·188	−0·157	9·86 (8·18)

* $P < 0·05$, ** $P < 0·01$.
† Standardized regression coefficients (beta weights).
‡ The ratio of effects is computed as the ratio of percentages of variance accounted for by visual component to the auditory component, i.e. the ratio of the squares of the beta weights. The first value listed is the ratio with Sound Sensitivity included as a predictor; the second value is the ratio with Sound Sensitivity excluded. Beta weights and R values in other columns are based on regression equations which include Sound Sensitivity.

Note. Each equation is based on a sample of 40 ratings, i.e. different subsets of the 70 ratings.

levels of cues varied less than did Pleasure levels of cues; the relative restriction of range in Arousal may have hindered its prediction.

Pleasure

All five of the Pleasure analyses showed significant results at the 0·01 or 0·05 level. Inspection of the beta weights in Table 1 reveals that judgments of the visual component alone have consistently higher contributions than that of the audio component alone. The ratio of the percentages of variance accounted for by two respective components ranges from 1·96 to 19·45; the mean value is 10·3. In other words, the relative contribution of the visual component to the auditory component is roughly estimated to be about 10 to 1. The results indicate that the visual component is a consistently stronger predictor of the Pleasure ratings of the composite stimuli.

Sound Sensitivity

The internal consistency reliability coefficient of the scale is 0·73. However, inclusion of Sound Sensitivity scores neither alters the ratio of influence much (see Table 1) nor predicts most overall ratings significantly. Only one of the 8 analyses showed a significant effect of Sound Sensitivity.

Discussion

The results clearly show that the visual sense contributes more to the judgment of pleasantness of environments than does the auditory sense. The finding is consistent with Faletti and Wellans' (1979) conclusions, despite their methods, as well as those of Seaton and Collins (1972), who approached the problem less directly, and the traditional perceptionists. However, the present study also indicates that the ratio between visual and auditory influence is highly variable. In addition, it appears that individual differences in sensitivity to sound do not alter the ratios. When these findings are contemplated, however, the relatively short duration of presentation and the presentation of several combinations in sequence to participants should be recalled. The findings may apply to environmental perceptions of individuals who are moving through a succession of settings (such as commuters and tourists) rather than to those who spend considerable time in a single setting.

Visual dominance

Evidence from this study confirms the notion that humans are predominantly visual animals when an environmental display is assessed for pleasantness. This conclusion is warranted only when vision and hearing are compared (i.e., when no other senses are part of the study) and when reasonably typical amounts of visual and auditory information are available. Certainly, as Posner, Nissen and Klein (1976) have argued, when vision provides inadequate information, the influence of other senses will increase. One might extend this caveat to include extremes of *affective* value. That is, if auditory and visual cues are both present in typical quantities, but one or more of the auditory cues have extraordinary affective qualities (e.g., a cutting remark, fingernails scratching the blackboard, music one despises), then one would expect the auditory cue to be assigned a greater role. Apart from these exceptional situations, however, Posner, Nissen and Klein's (1976) contention that visual dominance is related to the weak capacity of visual inputs to alert organisms to their occurrence, necessitating habitually greater attention to them, may account for the present results.

Variability of visual dominance

The present study suggests that while vision dominates hearing, the magnitude of the dominance is variable. This variability may be due in part to measurement error, but it is possible, through small sample methods, to estimate the population standard deviation of the ratio (Alder and Roessler, 1964, pp. 123–4). For the present sample of 5 Pleasure ratios, this estimate is 6·9 (around the mean of 10·3).

One can conclude that visual dominance is hardly a constant phenomenon. Presumably, this variation depends on the specific nature of the visual and the auditory cues, apart from their unisensory levels of Pleasure. Since the percentages of variance accounted for by all three predictors (visual pleasure, auditory pleasure and sensitivity to sound) ranged from 20 to 65%, between 35 and 80% of the variance is accounted for by error and other factors.

In the case of Arousal judgments, none of the three variables were consistently significant predictors. While this is a mysterious finding, it does suggest that visual dominance holds for some kinds of judgments but not for others.

Sound Sensitivity

Our attempt to show that the dominance ratio varies with perceiver characteristics was not fruitful. Variations in Sound Sensitivity do not seem to be associated with variations in influence ratios. This could be due to the self-report nature of the measure of sound sensitivity (if one subscribes to the belief that people cannot validly describe their own tendencies). Although the scale has reasonable internal consistency, the construct validity of the Sound Sensitivity scale has not yet been assessed; it may still lean too heavily on its noise sensitivity measurement origins despite our efforts to alter it toward the measurement of more generalized sound sensitivity. Other individual difference measures may, of course, account for some of the variance in dominance ratios. If not, one would be forced to conclude that variability in dominance ratios is largely or entirely a function of incoming visual and auditory cues.

References

Alder, H. L. and Roessler, E. B. (1964). *Introduction to Probability and Statistics*. San Francisco: Freeman.

Calvin, J. S., Dearinger, J. A. and Curtin, M. E. (1972). An attempt at assessing preferences for natural landscapes. *Environment and Behavior*, **4**, 447–70.

Canter, D. (1969). An intergroup comparison of connotative dimensions in architecture. *Environment and Behavior*, **1**, 37–48.

Cohen, J. and Cohen, P. (1975). *Applied Multiple Regression Correlation Analysis for the Behavioral Sciences*. Hillsdale, N.Y.: Erlbaum.

Colavita, F. B. (1974). Human sensory dominance. *Perception and Psychophysics*, **16**, 409–12.

Craik, K. H. (1968). The comprehension of the everyday physical environment. *Journal of the American Institute of Planners*, **34**, 29–37.

Ebel, R. L. (1951). Estimation of the reliability of ratings. *Psychometrika*, **16**, 407–24.

Faletti, M. V. and Wellans, A. R. (1979). From people to places: Extending a multimodal information processing paradigm. *Environmental Psychology and Nonverbal Behavior*, **3**, 248–52.

Mandel, D. (1978). Methodological approaches to environmental psychology. In P. A. Bell, J. D. Fisher and R. J. Loomis (eds), *Environmental Psychology*. Toronto: Saunders.

Mehrabian, A. and Ferris, S. R. (1967). Interference of attitudes from non-verbal communication in two channels. *Journal of Consulting Psychology*, **31**, 248–52.

Mehrabian, A. and Russell, J. A. (1974). *An Approach to Environmental Psychology*. Cambridge, Mass.: MIT Press.

Merrill, A. A. and Baird, J. C. (1980). Perception and recall of aesthetic quality in a familiar environment. *Psychological Research*, **42**, 375–390.

Peterson, G. L. (1967). A model of preference: quantitative analysis of the perception of the visual appearance of residential neighborhoods. *Journal of Regional Science*, **7**, 19–31.

Pick, H. L., Warren, D. H. and Hay, J. C. (1969). Sensory conflict in judgments of spatial orientation. *Perception and Psychophysics*, **6**, 203–5.

Posner, M. I., Nissen, M. J. and Klein, R. M. (1976). Visual dominance: an information-processing account of its origins and significance. *Psychological Review*, **83**, 151–71.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, **39**, 1161–78.

Russell, J. A. and Pratt, G. (1980). A description of the affective quality attributed to environments. *Journal of Personality and Social Psychology*, **38**, 311–22.

Schneider, D. J., Hastorf, A. H. and Ellsworth, P. C. (1979). *Person Perception*. Reading, Mass.: Addison-Wesley.

Seaton, R. and Collins, J. (1972). Validity and reliability of simulated buildings. In W. Mitchell (ed.), *Environmental Design: Research and Practice*, Volume 1. Stroudsburg, Pa.: Dowden, Hutchinson & Ross.

Sorte, G. J. (1975). Methods for presenting planned environments. *Man-Environment Systems*, **5**, 148–54.

Stevens, J. C. and Marks, L. E. (1980). Cross-modality matching functions generated by magnitude estimation. *Perception and Psychophysics*, **27**, 379–89.

Stevens, S. S. (1966). A metric for social consensus. *Science*, **151**, 530–41.

Ward, L. M. and Russell, J. A. (1981). The psychological representation of molar physical environments. *Journal of Experimental Psychology: General*, **110**, 121–52.

Weinstein, N. D. (1978). Individual differences in reactions to noise: A longitudinal study in a college dormitory. *Journal of Applied Psychology*, **63**, 458–66.

Manuscript received: 10 August 1982
Revised manuscript received: 25 October 1982