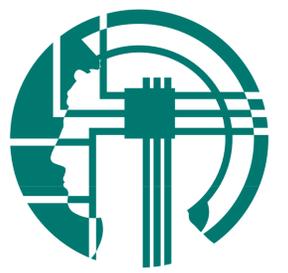




# Evaluating Architectural Interiors with Everyday Terms

Gerald Franz, Markus von der Heyde & Heinrich H. Bülthoff  
Max Planck Institute for Biological Cybernetics



## 1 Introduction

Predicting or evaluating the quality of interiors is a major unsolved difficulty in the discipline of architecture, because of the lack of generally accepted quality criteria and objective measuring methods. This study approaches the problem by analyzing the intersubjective part of colloquial language judgments on architectural interiors.

Beyond a statistical investigation of conciseness and intersubjectivity of colloquial ratings of interiors in general, this study quantifies assumed connections between subjective judgments and features of the scenes. The existence of regular patterns is exemplarily proved by demonstrating correlations between measurable image features and ratings in different denotative and connotative categories.

## 2 Methods

In a preparatory brainstorming session with 24 participants most common and subjectively important categories to characterize interiors were collected.

Subsequently, pictures of 15 different interiors were rated in 12 categories by 42 (15 female) subjects. Eleven (6 female) of the subjects had a professional background in interior design ("architects") and 31 (9 female) did not ("laypersons").

Each category was represented by a catchphrase and a pair of oppositional adjectives. The terms represented the most frequent words of the preceding brainstorming session. Additionally, dynamics and complexity (terms from architectural theory) were used.

The experiment used a novel internet based questionnaire technique derived from the classic semantic differential, a survey method that allows subjects to differentiate their judgment on a nine step scale between the extremes. The web interface presented rating categories and images of the scenes side by side.

The scenes were analyzed for their average color value, brightness, saturation and edges, measured using a basic image processing technique (Laplacian of Gaussian filter).

## 3 Results

The brainstorming analysis yielded only about 16-20 subjectively different categories. Three main groups were evident: mainly physical, mainly emotional, and somehow amalgamated and at a first glance fuzzy predicates.

In the web experiment, conciseness and intersubjectivity differed from category to category ( $std_{min} = 1.45, std_{max} = 2.23$ ). But, on average, ratings showed clear correlations ( $r^2_{max} = 0.82$ ).

Some categories were rated similarly over all scenes, e. g. **coziness**, **warmth** and **personal appreciation** appeared to be interrelated ( $r_1^2 = 0.82, r_2^2 = 0.62$ ), as well as **simplicity**, **negative complexity**, and **negative dynamics** ( $r_1^2 = 0.51, r_2^2 = 0.55$ ).

The data also showed correlations between ratings and basic image features: Most prominently, **warmth** of the ambience correlated with red RGB value ( $r^2 = 0.66$ ). Likewise, the correlations to the detected edges are remarkable: Rate of **enclosure**  $r^2 = 0.63$ , **appreciation**  $r^2 = 0.48$ , and **simplicity**  $r^2 = 0.39$ .

Furthermore, significant group specific differences were found: Architects generally responded more consistently. In addition, they tended to prefer more austere designs, most evident in the contrary tendencies in **personal appreciation**.

## 4 Conclusions

Rating architectural interiors using everyday terms provided qualitatively and quantitatively meaningful results. Systematic comparisons to image features suggest that even judgments in fuzzy categories of everyday language provide regular patterns potentially useful for the design of interior space.

It should be mentioned, however, that due to the small number of evaluated scenes the observed correlations certainly cannot claim universality. Nonetheless, the presence of several strong trends justifies the general method.

Further studies will investigate correlations of ratings with three dimensional properties of interiors. In combination with an extended scene database, this may be a promising way towards predicting and quantifying the character and ambience of rooms by their physical features.

### results of the brainstorming session

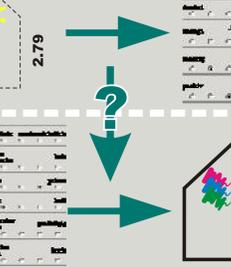
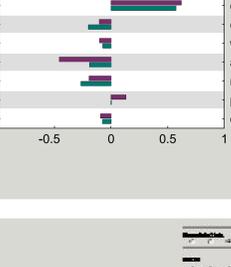
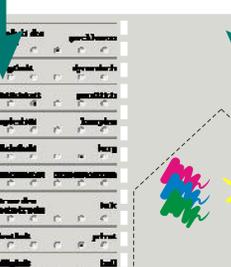
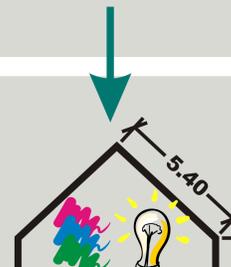
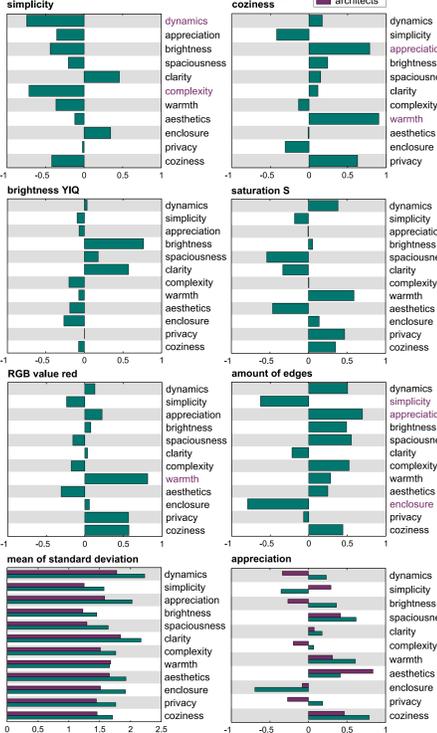
highest frequency categories	n
favor, disinclination	33
dimensions, proportions	28
rate of enclosure, transparency	26
form	25
level of detail, structure	23
brightness, color	20
warm, cold	19
naturalness	16
style	12

highest frequency adjectives	n
cool, cold	10
warm	9
open	9
bright, sunny	8
dark, gloomy	8
cozy, homey, comfortable	8
friendly, pleasant, cheerful	7
scaring, menacing, intimidating	7
sterile, clean	6

### categories and terms of the web questionnaire

category in English	category in German	German low extreme adjective	German high extreme adjective
coziness	Gemütlichkeit	ungemütlich	gemütlich
privacy	Privatheit	öffentlich	privat
rate of enclosure	Geschlossenheit des Raumes	offen	geschlossen
aesthetic quality	künstlerische Qualität	niedrig	hoch
warmth	Wärme des Raumeindrucks	kalt	warm
complexity	Komplexität	einfach	komplex
clarity	Übersichtlichkeit	unübersichtlich	übersichtlich
appropriateness of size, spaciousness	Angemessenheit der Raumgröße	beengt	großzügig
brightness	Helligkeit	dunkel	hell
personal appreciation	persönliche Wertung	negativ	positiv
simplicity	Schlichtheit	überladen	karg
dynamics	Bewegtheit	statisch	dynamisch

### correlation coefficient r to means of ratings



*Do colloquial judgments about architecture contain recurring intersubjective patterns?*

*Are they correlated to measurable features of the initial scene?*

*The bandwidth of common judgments about the ambience of architectural interiors was collected in a preparatory brainstorming session.*

*In a web experiment with a semantic differential questionnaire, 15 scenes were systematically evaluated using 12 selected categories from the first session.*

*The brainstorming revealed a limited and consistent spectrum of interior characterizations.*

*Also, certain rating categories of the web experiment (e.g., warmth, coziness, appreciation) appeared to be highly interrelated.*

*Several correlations between rating categories and basic image features of the scene were unexpectedly high.*

*Ratings of architects differed in conciseness and in appreciation categories from other observers.*

*Comparing ratings with scene features provided further insight into the underlyings of judgments.*

*An extended database may allow generalizable and quantifiable predications about the character of rooms.*