# Trade-Off Between Resolution and Interactivity in Spatial Task Performance

**Gerda J.F. Smets and Kees J. Overbeeke**
*Delft University of Technology*

Virtual reality displays usually lag far behind classical computer graphics displays in static image quality parameters, such as resolution. Both the popular press and scientific papers often stress that resolution will have to increase greatly before users can experience virtual environments as "the real thing." Nevertheless, it is already possible to do some useful work in VR environments. The point we experimentally demonstrate here is that resolution is much less important for interactive tasks that employ immersive VR, where users can explore the environment by moving their heads and bodies, than it is in classical computer graphics applications, where users can only explore by gazing at a single picture. Swartz, Wallace, and Tkacz[1] have shown, in the context of unmanned aerial vehicles, that frame rate (read: passive camera movement) is more important than resolution for target detection, recognition, designation, and tracking. They call these results "surprising."

**Experiments comparing search-and-act spatial task performance showed that image resolution is very important in static viewing, but not in immersive VR. Nor did animating the image always improve performance.**

In the experiments reported here, we investigated the relative importance of various image parameters like spatial resolution (number of pixels per video frame), intensity resolution (number of gray levels per pixel), and temporal resolution (number of frame updates per second).[2] Most experimental data concerning these resolutions come from classical psychophysics. However, experimental conditions in classical psychophysics feature stationary observers looking at short-term, point-like flashes on stationary displays, and are thus far more representative of human interaction with pictures and photographs than with highly interactive systems like those employed in virtual reality. Our senses did not develop while we were sitting still.

Pepper, Cole, and Spain[3] conducted experiments illustrating the importance of movement for depth perception. They showed that results for depth estimates under monocular movement parallax conditions compare well with those under stereoscopic movement parallax conditions. When the subject can move, one eye suffices for depth perception.

Interest in the study of our senses as perceptual systems has grown in recent years. For example, Gibson[4] initiated the approach of studying the perceptual capabilities of human observers while they explore or perform tasks involving perceptual and motor skills. These studies, recently dubbed "active psychophysics,"[5] show much more potential for measuring human capabilities and the technical requirements to support them. Furthermore, active psychophysics has more implementation potential, as we will show (see below, "Discussion").

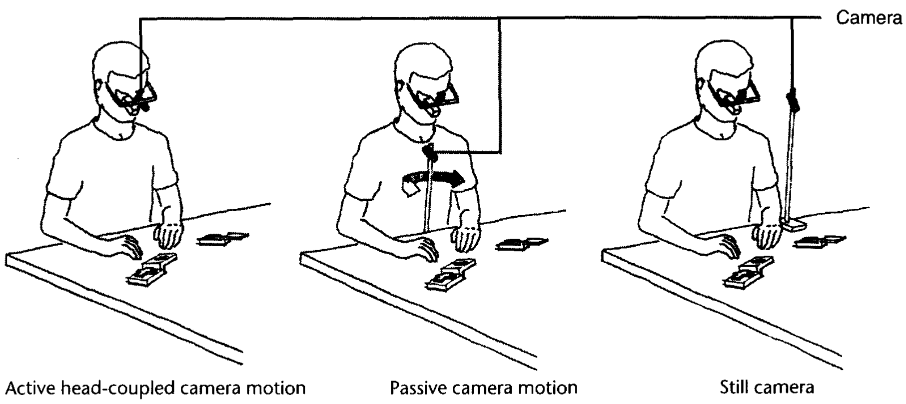
The present experiments clearly demonstrate the order of magnitude of resolutions needed in VR conditions as compared to those in static image presentation. A typical VR display contains 320 × 200 pixels. Current development emphasizes increasing image resolution almost exclusively. We want to show that image resolution is relative, by demonstrating that this variable should be considered in relation to other variables, such as movement.

## Experiment I

We compared performance in a search-and-act task for subjects whose visual information was artificially impoverished in three interactivity conditions: still camera, passive camera motion, and head-coupled camera movement (see Figure 1).



0272-17-16/95/$4.00 © 1995 IEEE

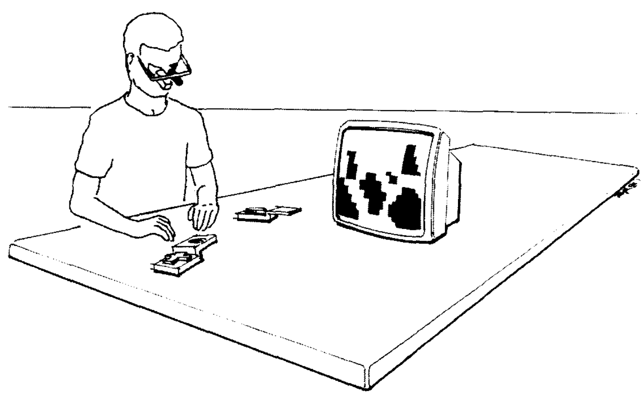



**1 Image-viewing conditions varying in level of interactivity (left to right: active, passive, still).**

### *Apparatus*

Subjects were fitted with a head-mounted system containing a display and, depending on the experimental condition, a microcamera, as illustrated in Figure 2. This system fed the camera image through a videoprocessor to manipulate aspects of the image stream, for example, the number of gray values, pixels, and frames per second. The system operated in real time. The camera was a Panasonic WV-CD1 with a camera-head diameter of 17 mm, length of 48 mm, and weight of approximately 20 g. The PAL 625-line video signal was manipulated with a Panasonic WJ-MW10 production mixer.

The display was an electronic viewfinder of a Sony Video Hi-8 camera type CCD-V900 E. The screen size measured 11 mm × 8.2 mm and weighed about 75 g. The helmet-mounted system, including camera, viewer, and helmet, weighed 350 g. The visual angle obtained by the combination of the camera and the ocular lens was 40 degrees monocular (the left eye being patched). The enlargement factor was 0.78. This meant that visual input did not entirely fill the field of view and that the subject's hands appeared to be further away than they really were. Although we used the same lenses throughout Experiment I, we changed the lens setup for Experiment II.



**2 Apparatus.**

**3 Puzzle in Experiment I.**

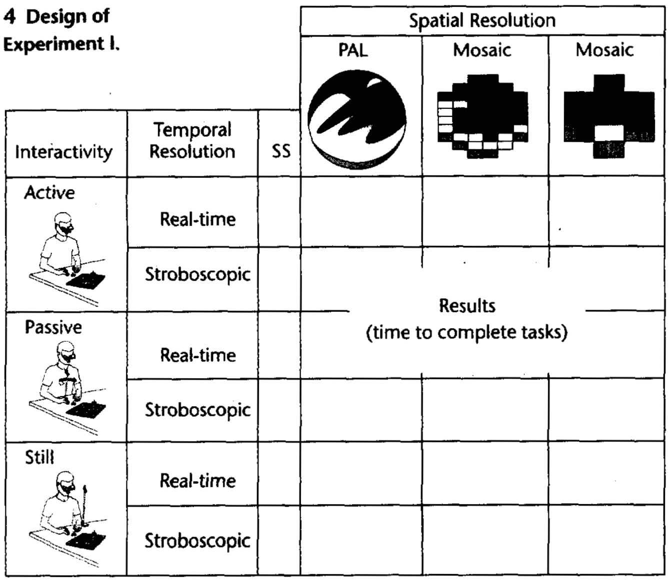
### *Procedure*

The subjects had to complete a jigsaw-like puzzle for four-year-old children, as shown in Figure 3. The puzzle pieces consisted of a hole on one side and an elevated figure on the other. The subjects had to make a chain by clicking a hole over a corresponding elevated figure. This easy task did not require any training. Subjects received explicit instructions to touch the blocks only by the side to prevent tactile exploration. The dependent variable was the time needed to complete this puzzle as measured with a stop watch. If the subject had not combined two puzzle blocks after 10 minutes, the trial was stopped and a zero result recorded.

### *Subjects*

Subjects were four junior staff members who had no previous experience in psychological experiments. None had an uncorrected visual problem.

**4 Design of Experiment I.**




SS = sum of squares

### *Design*

Figure 4 shows the experimental design for Experiment I. Under all conditions, we used four gray values: black, white, and two intermediate grays. The subjects went through all conditions. Conditions were counterbalanced to avoid order effects. As Figure 4 shows, there were three independent variables: spatial resolution, temporal resolution, and interactivity. Spatial resolution had three levels: a PAL 625 video image, an image consisting of a 36 × 30 mosaic, and an image of a 18 × 15 mosaic. Temporal resolution had two levels: real-time PAL (25 Hz) and stroboscopically sampled (5 Hz).

Interactivity occurred through correspondence between the observer's exploratory movements and the visual input. This was manipulated on three levels. At the first level, the camera recorded the scene from a single viewpoint (still), resulting in an overall view of the scene. At the second level, a small electrical motor moved the camera on a steady track around the scene (passive), also resulting in an overall view. At the third level, the camera was attached to the observer's head (active). In all conditions, subjects could move their heads freely. The camera positions and resulting viewing points thus differed in the three conditions. We showed in previous experiments[6] that when feeding the passive and active subjects an identical image, the active subject still outperformed the passive one: interactivity is the essential feature (see below, "Discussion").

**Table 1. Data for Experiment I.**

| | | | Spatial Resolution | | |
|---|---|---|---|---|---|
| | | | PAL 625 lines | Mosaic 36 × 30 | Mosaic 18 × 15 |
| **Interactivity** | **Temporal Resolution** | **SS** | **Time to Complete Task (in seconds)** | | |
| Active | Real-time | 1 | 39 | 114 | 200 |
| | | 2 | 42 | 353 | t.o. |
| | | 3 | 43 | 86 | 486 |
| | | 4 | 41 | 67 | 290 |
| | Stroboscopic | 1 | 48 | 102 | 302 |
| | | 2 | 57 | 291 | t.o |
| | | 3 | 47 | 372 | 481 |
| | | 4 | 57 | 94 | 157 |
| Passive | Real-time | 1 | 75 | 396 | 875 |
| | | 2 | 92 | 416 | t.o. |
| | | 3 | 62 | 367 | 368 |
| | | 4 | 55 | 230 | t.o. |
| | Stroboscopic | 1 | 147 | 414 | t.o. |
| | | 2 | 54 | 701 | t.o. |
| | | 3 | 108 | 370 | t.o. |
| | | 4 | 84 | 236 | t.o. |
| Still | Real-time | 1 | 47 | 204 | 418 |
| | | 2 | 92 | 206 | t.o. |
| | | 3 | 77 | 263 | 492 |
| | | 4 | 64 | 713 | 486 |
| | Stroboscopic | 1 | 55 | 268 | t.o. |
| | | 2 | 80 | 877 | t.o. |
| | | 3 | 47 | 350 | 469 |
| | | 4 | 100 | 495 | t.o. |

SS = sum of squares t.o. = timed out (more than 600s)

### *Hypothesis*

Self-generated optic flow, where the observer's explorative movements cause shifts in the optic array, is only present when the camera is head-mounted.[7] In this condition we expected performance to stay high, even with low spatial resolution. Therefore, we predicted a significant interaction between the independent variables. Furthermore we predicted a significant main effect for both.

### *Results*

Results appear in Tables 1 (raw data) and 2 (analysis of variance, see sidebar "Anova and Its Uses," p. 51). They include only two levels of spatial resolution, as the 18 × 15 condition was clearly too difficult. However, the data show that in the active condition, three out of four subjects solved the puzzle. The main effect of spatial resolution was significant. This, of course, is not new. The interactivity condition is also significant, indicating the importance of actively controlled visual input. The temporal resolution was not significant, perhaps because the differences in image update rate were not large enough.

As for our main concern, the

interaction between spatial resolution and interactivity was not significant at the 0.05 level, although we expected low-resolution performance to be better in the active condition. Two reasons might explain this. First, a strong learning effect developed because the same puzzle pieces were used in all conditions. Second, we excluded the third level of spatial resolution because of missing data. This level was too difficult for the subjects in some conditions.

Table 2. Analysis of Experiment I.

| Source | Sum of Squares | df | Mean Square | F-value |
|---|---|---|---|---|
| Subject | 87,713.42 | 3 | 29,237.81 | |
| Spatial resolution (SR) | 845,883.00 | 1 | 845,883.00 | 30.29*** |
| SR * Subject | 83,786.17 | 3 | 27,928.27 | |
| Interactivity (I) | 170,468.79 | 2 | 85,234.40 | 4.26** |
| I * Subject | 120,158.71 | 6 | 20,026.45 | |
| Temporal resolution (TR) | 35,752.08 | 1 | 35,752.08 | 2.57 |
| TR * Subject | 41,814.00 | 3 | 13,938.03 | |
| SR * I | 101,418.87 | 2 | 50,709.44 | 3.56* |
| SR * I * Subject | 85,396.00 | 6 | 14,232.83 | |
| SR * TR | 20,833.33 | 1 | 20,833.33 | 0.92 |
| SR * TR * Subject | 67,547.00 | 3 | 22,524.83 | |
| I * TR | 3,283.29 | 2 | 1,641.65 | 0.13 |
| I * TR * Subject | 78,335.54 | 6 | 13,055.92 | |
| SR * I * TR | 6,769.04 | 2 | 3,384.52 | 0.22 |
| SR * I * TR * Subject | 90,690.13 | 6 | 15,115.02 | |

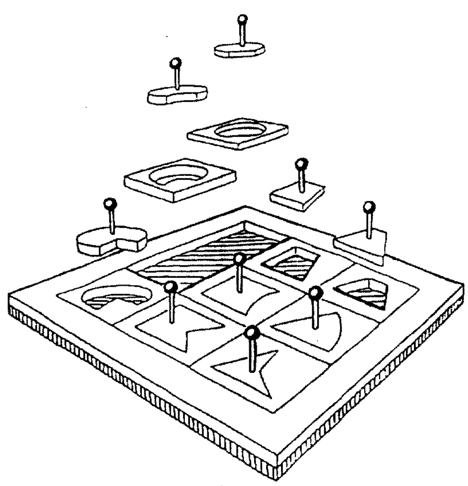
* = $p < 0.10$ ** = $p < 0.05$ *** = $p < 0.01$ df = degrees of freedom

## Experiment II

This experiment partially replicated Experiment I, but optimized to eliminate the unwanted factors identified above. We changed the lenses of the viewer/camera system to obtain a 60-degree visual angle from the combination of the camera and the ocular lens. The resulting enlargement factor was 1.00. This means that the visual input filled the field of view and that the subject's hands appeared at normal distance. In Experiment I, the subjects were hampered by the 0.78 enlargement factor, as indicated by the difficulties they had in grasping the puzzle pieces.

### *Task*

The subjects had to complete a specially designed puzzle, depicted in Figure 5. This puzzle excluded any learning by placement, since the location of each piece varied randomly throughout the trials. Subjects received explicit instructions to handle the pieces only by the pegs to prevent tactile exploration. Subjects required no training, as the task was very easy.



**5 Puzzle in Experiment II.**

### *Subjects*

The subjects were five volunteer students in industrial design engineering, who had no previous experience in psychological experiments. None had an uncorrected visual problem.

### *Design and results*

We simplified the design by leaving out the temporal resolution manipulation. The independent variables spatial resolution and interactivity were retained with identical levels. Design and results are shown in Tables 3 (raw data) and 4 (analysis of variance). The third level of spatial resolution was again excluded from the analysis. It can be seen however that in the active condition three out of five subjects still can solve the puzzle, a much better result than in both other interaction conditions (passive and still).

The main spatial effect is again significant ($p < 0.01$). Interactivity is also significant ($p < 0.05$), and returning to the point of our major concern, the interaction between spatial resolution and interactivity is significant ($p < 0.05$). With decreasing spatial resolution, observers perform better when they actively control the camera by their head movements.

## Discussion

The results of both experiments show that the added interactivity of virtual reality can compensate for losses in spatial resolution in a way that passively animated images cannot. The trade-off fits in well with Sheridan's three-factor model of telepresence,[8] where three independent factors together add up to the quality of presence realized by a teleoperator system. The factors are

1. the extent of sensory information (such as resolution),
2. the amount of control over sensors (called "interactivity" in this article), and
3. the user's ability to modify the environment.

The advantages of VR conditions also match well with results from medical prosthetics, where it was shown that babies born without arms have difficulties in developing depth perception.[9] Once these babies are equipped with simple stick-like prostheses that let them reach and touch, depth perception no longer proved a

Table 3. Design and data of Experiment II.

| Interactivity | SS | Spatial Resolution: PAL 625 lines | Mosaic 36 × 30 | Mosaic 18 × 15 |
|---|---|---|---|---|
| | | Time to Complete Task (in seconds) | | |
| Active | 1 | 33 | 73 | 242 |
| | 2 | 35 | 141 | 259 |
| | 3 | 32 | 51 | 210 |
| | 4 | 24 | 190 | t.o. |
| | 5 | 27 | 88 | t.o. |
| Passive | 1 | 56 | 143 | t.o. |
| | 2 | 51 | 596 | t.o. |
| | 3 | 41 | 153 | t.o. |
| | 4 | 48 | 531 | t.o. |
| | 5 | 101 | 600 | t.o. |
| Still | 1 | 35 | 422 | t.o. |
| | 2 | 49 | 123 | t.o. |
| | 3 | 41 | 123 | 384 |
| | 4 | 97 | 472 | t.o. |
| | 5 | 56 | 436 | t.o. |

SS = sum of squares    t.o. = timed out (more than 600 s)

Table 4. Analysis of Experiment II.

| Source | Sum of Squares | df | Mean Square | F-value |
|---|---|---|---|---|
| Subjects | 98,595.53 | 4 | 24,648.88 | |
| SR | 388,968.53 | 1 | 388,968.53 | 21.53*** |
| SR * Subject | 72,259.13 | 4 | 18,064.78 | |
| I | 140,221.07 | 2 | 70,110.53 | 5.49** |
| I * Subject | 102,151.27 | 8 | 12,768.91 | |
| SR * I | 92,785.87 | 2 | 46,392.93 | 4.06** |
| SR * I* Subject | 91,434.47 | 8 | 11,429.31 | |

** = $p < 0.05$    *** = $p < 0.01$    df = degrees of freedom

difficulty. Bach-y-Rita[10] found in his Tactile Visual Substitution System, which presented digitized camera images to a congenitally blind subject by means of an array of vibrating pins placed against the skin of his back, that the subject can "see" a spatial layout in front of him if and only if he controls the movements of the camera. Otherwise, he only feels the vibrating pins on his back.

In our own applied research, we are using the experiments described here to develop nonimmersive systems for teleoperation and surgery using the Delft Virtual Window System.[6,9] DVWS produces movement parallax by adapting the viewpoint of a real or virtual camera to match the displacements of the observer's head in front of the display (not unlike fish-tank VR). For several application areas, such as X-ray luggage inspection and medical and industrial endoscopes, this system easily outperforms a static stereoscopic display. Experiments with DVWS indicated a perceptual advantage of an active (head-coupled) observer over a passive one (noncoupled).[6] The test involved aligning wedges on local and remote objects. Both active and passive observers received identical output on their monitor screens. While the active observer was able to align a real wedge placed in front of the screen with a wedge virtually leaping out of the screen, the passive observer could not.

## Conclusions

The experiments reported here provide behavioral evidence about the relative importance of spatial and temporal resolution factors (pixels per frame and frames per second, respectively) in static, dynamic, and interactive display conditions. Although the experiments were performed using real light and cameras, the results apply equally well to computer-based display systems. Results show that especially in interactive virtual reality viewing conditions, static resolution qualities are a relatively minor concern for (some) spatial orientation and performance tasks, as compared to their prominence for static and passive animation displays. ■

## Acknowledgments

We wish to thank F. Loosen of the Leuven Catholic University for statistical advice.

## References

1. M. Swartz, D. Wallace, and S. Tkacz, "The Influence of Frame Rate and Resolution Reduction on Human Performance," *Proc. Human Factors Society 36th Annual Meeting*, Human Factors and Ergonmics Soc., Santa Monica, Calif., 1992, pp. 1,440-1,444.
2. V. Ranadive and T. Sheridan, "Video Frame Rate, Resolution, and Grey Scale Trade-Offs for Undersea Manipulator Control," *Proc. 17th Annual Conf. on Manual Control*, Los Angeles, 1981, pp. 77-88.
3. R.L. Pepper, R.E. Cole, and E.H. Spain, "The Influence of Camera Separation and Head Movement on Perceptual Performance Under Direct and TV-Displayed Conditions," *Proc. SID*, Soc. for Information Display, Playa del Rey, Calif., 1983, pp. 24:73-80.
4. J.J. Gibson, *The Ecological Approach to Visual Perception*, Houghton Mifflin, Boston, Mass., 1979.
5. J.M. Flach, "Control with and Eye for Perception: Precursors to an Active Psychophysics," *Ecological Psychology*, Vol. 2, No. 2, 1990, pp. 83-110.
6. C.J. Overbeeke and M.H. Stratmann, *Space Through Movement*, doctoral dissertation, Faculty of Industrial Design Engineering, Delft University of Technology, Netherlands, 1988.
7. D.N. Lee, "A Theory of Visual Control of Braking Based on Information about Time-to-Collision," *Perception*, Vol. 5, 1976, pp. 437-459.
8. T. Sheridan, "Musings on Telepresence and Virtual Presence," *Presence*, Vol. 1, No. 1, 1992, pp. 120-126.
9. G.J.F. Smets, "Designing for Telepresence: The Interdependence of Movement and Visual Perception Implemented," *Proc. Fifth IFAC/IFIP/IFORS/IEA Symp. on Analysis, Design, and Evaluation of Man-Machine Systems*, The Hague, Netherlands, 1992, pp. 1-7.
10. P. Bach-y-Rita, *Brain Mechanisms in Sensory Substitution*, Academic Press, London, 1972.

## On Anova and its use

Anova (short for "analysis of variance") is a statistical technique to separate the variance ascribable to one group of causes from the variance ascribable to other groups. This is done by testing differences between variances under the *F*-distribution. Take the example of interactivity in Table 3. Does the variance calculated per condition (that is, separately for active, passive, and still) and summed exceed the variance calculated over all conditions? The terms in Tables 2 and 4 are

Source = the source of variance
Sum of squares = that is, the variance
df = degrees of freedom (number of conditions minus 1)
Mean square = sum of squares divided by df
*F*-value = mean square of source divided by its error term

Anova is a technique to test interaction effects. For example, in our case we predict that the mean performance times will be roughly the same in the PAL resolution conditions for all interactivity conditions, but will increase dramatically from the active to the passive and still conditions. We predict a noncrossing interaction. Noncrossing interactions as revealed by Anova are only valid for data on an interval scale, and not every performance time measure is necessarily on an interval scale (see Winer,[1] pp. 449-452). For the task under discussion, however, we think the assumption of an interval scale is warranted, as it often is in reaction time (RT) measurements for slower visual search tasks (but not for fast-reaction tasks). The task for this experiment consisted of a repetitive sequence of object positionings and took a minimum of 24 seconds to complete.

The question narrows down to whether there is a linear relation between performance time, as measured by the experimenter, and physical time. Most psychophysical studies reveal no linear relation. For example, brightness and light intensity is a typical case of a psychophysical law: there is no linear relation between perceived brightness and physical intensity. But we think we are measuring in a totally different realm. We are not measuring time as experienced by the subject, but the time elapsing while the subject performs the same task under different conditions.

Furthermore, alternative hypothesis testing procedures are not readily available. We see three alternatives that, although not very powerful or much used, do not depend on the assumption of an interval scale. These are conjoint measurement and two nonparametric tests, namely, the median test and the randomization test. Townsend and Ashby[2] describe conjoint measurement, which they say has a theoretical advantage over the Anova of "not requir[ing] knowledge of the true underlying numerical scales" (p. 396). However, they found this advantage of little practical use

> [because] conjoint measurement provides no basis for statistical testing its predictions . . . [and] until some technique is developed for observing the empirical processing time relation, it appears that, as a test of RT additivity, conjoint measurement has few advantages over more traditional techniques such as the analysis of variance.

Probably for that reason, applications of conjoint measurement are rarely encountered in the experimental literature. The median test also does not need the data to be on an interval scale, as it compares medians (middle values) rather than means. However, this test has very little power, requiring large amounts of data even when the effects of the independent variables are large. The randomization test (Edgington,[3] pp. 158-159) calculates the exact alpha values for the data matrix by explicitly calculating all relevant permutations of the measured data. No statistical or scale assumptions are made. This method has not been used very often because of the complexities of the combinatorial calculations.

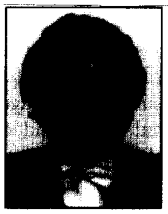
Summarizing, we are convinced that the effect we describe is sufficiently strong, that the argument's force does not reside in the statistics, that our use of the Anova is sufficiently valid for the longer time spans involved, and that the other tests do not yet offer an alternative to Anova.

### References


1. B.J. Winer, *Statistical Principles in Experimental Design*, McGraw-Hill, London, 1962 (int'l student edition).
2. J.T. Townsend and F.G. Ashby, *The Stochastic Modeling of Elementary Psychological Processes*, Cambridge Univ. Press, Cambridge, NY, 1983.
3. E.S. Edgington, *Randomization Tests*, Dekker, New York, 1980.



***Gerda J.F. Smets** is a full professor at the Faculty of Industrial Design Engineering, Delft University of Technology, where she heads an interdisciplinary group of software engineers, physicists, psychologists, design engineers, ethologists, mechanical engineers, and designers. Her research interests focus on the interaction between Gibsonian perception theory and industrial design engineering including 3D perception, virtual reality, computer-human interaction, aesthetics, and form semantics. Smets has a PhD from Leuven Catholic University, Belgium, 1971.*


***Kees J. Overbeeke** is an associate professor at the Faculty of Industrial Design Engineering, Delft University of Technology. His research interests include form semantics, 3D visual and auditory perception, computer-human interaction, and design epistemology. He has a PhD from Delft University for his work on a head-slaved system to create a 3D impression on a flat screen.*

*Readers may contact both authors at Delft University of Technology, IO, Jaffalaan 9, 2628 BX, Delft, the Netherlands, e-mail {g.j.f.smets, c.j.overbeeke}@tudelft.io.nl.*