

# A Perceptive Evaluation of Volume Rendering Techniques

Christian Boucheny\* Georges-Pierre Bonneau\*  
LJK - Grenoble University - INRIA

Jacques Droulez †  
LPPA - CNRS - Collège de France

Guillaume Thibault‡ Stéphane Ploix‡  
eDF

## Abstract

The display of space filling data is still a challenge for the community of visualization. Direct Volume Rendering (DVR) is one of the most important techniques developed to achieve direct perception of such volumetric data. It is based on semi-transparent representations, where the data are accumulated in a depth-dependent order. However, it produces images that may be difficult to understand, and thus several techniques have been proposed so as to improve its effectiveness, using for instance lighting models or simpler representations (e.g. Maximum Intensity Projection). In this paper we present two perceptual studies that question how DVR meets its goals, in either static or dynamic context. We show that a static representation is highly ambiguous, even in simple cases, but this can be counterbalanced by use of dynamic cues, i.e. motion parallax, provided that the rendering parameters are correctly tuned.

**CR Categories:** I.3.3 [Computer Graphics]: Picture/Image Generation—Display Algorithms; I.3.8 [Computer Graphics]: Applications; J.2 [Physical Sciences and Engineering]

**Keywords:** Direct Volume Rendering, perception of transparency, static and dynamic cues

## 1 Introduction

Dense three-dimensional datasets represent a challenge for the visualization community. Ideally, one would like to see at a glance all the data that fill the space, and have a clear picture in mind of the spatial layout, of the organization within the volume of the displayed physical fields. However, such representations are not natural, as the human visual system mainly experiences surfaces that hide one another through occlusion. Thus, early representations of scalar fields were based on the extraction of characteristic surfaces that could be displayed, for instance surfaces of isovalues. But these methods require the viewer to mentally reconstruct a volume from a sequence of pictures, which is a very complex cognitive task.

To overcome this limitation, it has been suggested to consider dense scalar data fields as volumes of particles with emissive and absorptive properties, like drops of water that modify the light rays traversing a cloud. This led to the development of Direct Volume Rendering methods that highly rely on transparencies. These techniques have been thoroughly developed and improved since their birth in the late 80's. However, it is surprising to notice that they do not

seem to be intensively used by final users outside the medical community, as we could observe, despite their implementation in most visualization softwares. As was argued by [Hibbard 2000], "all 3D graphics suffer from ambiguity, with many 3D points projected onto each point of the 2D screen (or of the user's 2D retina). This ambiguity is strongest in volume rendering, where the color of each screen pixel is determined by a (theoretical) continuum of semi-transparent 3D points", which could be one of the main reason why scientists do not find volume rendering to be as accurate as other techniques. Produced images may look nice, but if no recognizable structure is present in the data it can be really hard to extract spatial relationships. Another reason for the lack of interest of scientists in volume rendering consists in the difficulty to tune the Transfer Function (TF), which is the determinant of the final rendering but often proves complex to adjust even for expert users. But strikingly, when [Kaufman and Mueller 2005] give a thorough review of research in volume rendering, none of the near 300 references are devoted to evaluation and validation based on user studies.

In this paper, we address from the human visual system point of view the question of the perceived depth order of semi-transparent objects rendered by means of DVR. The focus is on depth perception, as it is a very important aspect of volume rendering that can be quantified; we do not deal with other perceptive issues such as pattern recognition or matching. To the best of our knowledge, this paper represents the first insight on the perceptive problems of DVR. We do not intend to formulate definitive conclusions, but rather wish to orient future works on volume rendering. Computational costs related to volume rendering of large datasets often lead to quasi-static representations, whereas motion is a strong cue to depth perception. We thus carried out two experiments, in static and dynamic contexts. Transparent renderings seem to generate depth order ambiguities, also found in point-based rendering. Our main hypothesis is that if we are able to perceive transparency per se, such as in clouds or water, understanding the spatial layout of data in a translucent material cannot be done in static monocular presentations, but requires dynamic depth cues such as motion parallax. In this paper, monocular will relate to representations without any binocular disparity, like when one looks at a basic computer screen.

The paper is organized as follows: first, we propose an overall review of Direct Volume Rendering in computer graphics, and present some well-known results about the perception of transparency. Then, in a first experiment, we examine how human subjects are able to discriminate in depth two static objects displayed with DVR technique, when color transparencies and accumulations are the only sources of information. To avoid perspective biases, we use orthographic projections. Then, a second experiment focuses on the perception of depth in rotating semi-transparent volumes.

## 2 Volume rendering in computer graphics

A complete overview of volume rendering is out of scope for this paper. The reader can refer to [Kaufman and Mueller 2005] for a thorough review. In this section we give the main ideas and milestones, and cite the few papers concerned with perception, validation or user studies.

\*e-mail: {christian.boucheny, georges-pierre.bonneau}@inrialpes.fr

†e-mail: jacques.droulez@college-de-france.fr

‡e-mail: {guillaume.thibault, stephane.ploix}@edf.fr

The idea to use transparencies to represent volumetric data has come from initial works on clouds rendering in computer graphics ([Kajiya and Herzen 1984], [Max 1995]). It has been proposed to use the same raytracing techniques to produce images of a volumetric scalar field, considering the field as a density emitter with single level of scattering ([Sabella 1988]). Such representations aim at considering all the data for the display, contrary to what is done in surfacic methods such as isosurface (data thresholding) or cross-section (using cutting planes) renderings that require the viewer to mentally reconstruct the spatial structure of the field seen across a sequence of images.

The basic idea in DVR techniques, whatever their technical implementation might be, consists in an evaluation of the *volume rendering integral*. The luminance corresponding to a ray  $\omega$  cast from an image pixel and crossing the volume along a segment of length  $D$  (with parametrization  $x(\lambda)$  along the ray,  $\lambda$  is the distance to the viewpoint and belongs to  $[0:D]$ ), is given by:

$$I = \int_0^D C(x(\lambda)) \cdot e^{-\int_0^\lambda \tau(x(\lambda')d\lambda')} d\lambda \quad (1)$$

where  $C(x)$  represents the voxel color at  $x$ , and  $\tau(x)$  its extinction coefficient. This equation is often evaluated in back-to-front order based on the following simplification:

$$I \approx \sum_{i=0}^n \tilde{C}_i \cdot \prod_{j=0}^{i-1} (1 - \alpha_j) \quad (2)$$

where  $\alpha_i$  and  $\tilde{C}_i$  stand for the opacity and the color of the  $i^{th}$  voxel in the ray. DVR requires the definition of a Transfer Function (TF)  $\phi$  that maps the scalar  $s(x)$  associated to voxel  $x$  onto luminance  $l$  and opacity  $\alpha$  values:

$$\phi : s \rightarrow (l, \alpha) \quad (3)$$

The Transfer Function critically determines how the data will be perceived in the final rendering. Its choice may be a very delicate problem, and some efforts have been carried out to automatize this process (for reference, see [Pfister et al. 2000]). Most often, piecewise linear functions are chosen.

Maximum Intensity Projection (MIP) and X-Ray constitute alternatives to optically-based DVR. In the former, only the maximum voxel on each ray is projected on the screen, whereas the later corresponds to an unordered sum of all voxel values along the ray. MIP is very useful for datasets with well-segmented volumes, as is the case with CT or MR medical data (e.g., angiography). MIP and X-Ray tend to generate more contrasted images than DVR, improving the visibility of substructures; but, as they are order-independent, they provide no depth cues to the viewer. Contrary to DVR, images created with X-Ray and MIP have fronto-planar symmetry, which means that volumes viewed from front and back result in the same images.

In computer graphics, a lot of work has been devoted to develop existing techniques, either to reduce the computational costs ([Lacroute and Levoy 1994]) or to better approximate the volume rendering integral and reduce image artifacts ([Engel et al. 2001]). Some works have focused on the improvement of depth perception in volume renderings. [Levoy 1988] implemented a volume rendering pipeline including lighting calculations to strengthen the perception of surfaces within the volume. [Ebert and Rheingans 2000], in their volume illustration framework, introduced NPR techniques such as silhouette enhancement and tone shading in physics-based volume rendering; [Mora and Ebert 2004] suggest to improve MIP

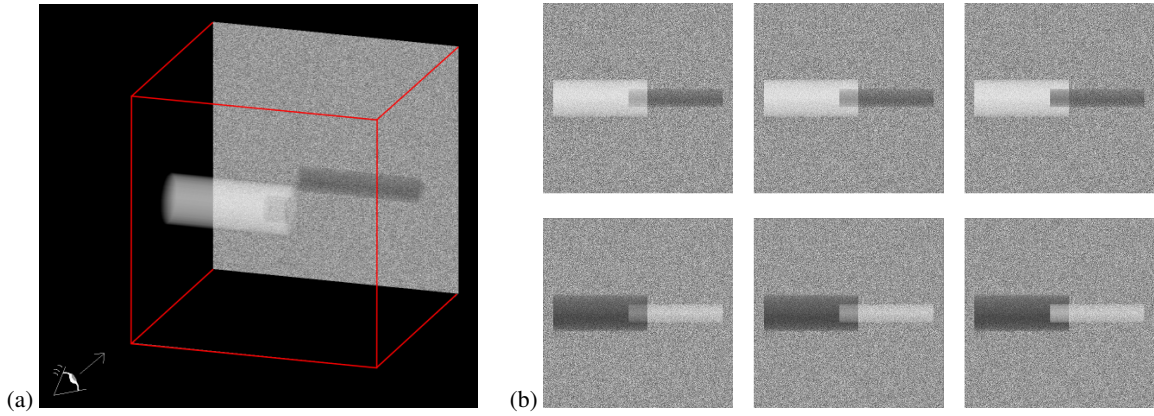
and X-Ray using gradient signals inside the volume and stereo-images. Generally, improvements are measured on the basis of the quality of produced images, compared with traditional techniques. Amongst the few user studies led to validate specific volume renderings, we can mention [Kersten et al. 2006], who use atmospheric perspective and stereo to improve X-Ray-like rendering, and [Ropinski et al. 2006] with NPR-enhancement of MIP images in angiography. Earlier, [Interrante et al. 1997] studied how textured-transparent surfaces could convey depth information, but this approach is limited to two wrapped surfaces.

### 3 The perception of transparency

Seeing dense and complex transparent media does not correspond to natural situations. We can sometimes look at solid surfaces through translucent media, such as a window or a water surface, or observe complex translucent objects such as clouds or crystals, but in this last case the layout of matter densities inside the volume does not seem obvious.

The perceptual aspects of transparency have been studied for a long time, with particular focus on the required image conditions for producing the illusion of transparent layers. [Metelli 1974] developed a first model postulating that, for a disk overlapping two vertical strips with different reflectance  $a$  and  $b$  and producing intersecting areas of reflectance  $p$  and  $q$ , the level of transparency is given by  $\alpha = \frac{p-q}{a-b}$ , with values close to 1 leading to full transparency. Interestingly, Metelli's model is firstly experimental and generative, as it corresponds to a disk (the episcotister) with a reflectance  $t$  and an open fraction of  $\alpha$  that would rotate quickly over the two background strips (leading to  $p = \alpha a + (1 - \alpha)t$  and  $q = \alpha b + (1 - \alpha)t$ ). It can be noted that Metelli's model of perceived transparency corresponds to the widely used alpha-blending equation encountered in computer graphics. Metelli's model was further developed in terms of luminance rather than reflectance by [Gerbinio et al. 1990]. More recently, Singh and Anderson [2002; 2004] developed an alternative achromatic model of transparency that relies on Michelson contrasts of adjoining areas. Such variants are required to account for the two aspects of a transparent layer: its luminance and its transmittance, whereas Metelli's model focuses only on the second aspect. The conditions and ranges of achromatic transparencies have been assessed in many studies (for instance in [Kasrai and Kingdom 2001]), but to the best of our knowledge few works focused on the perceived layout of superimposed transparent media.

A fundamental condition for phenomenal transparency - namely the perception of the transparency of a surface - to occur consists in the presence of contour junctions that correspond to luminance steps between adjoining regions. The shape of these junctions (X-, T- or I-junctions) and the layout of luminances around determine if one is likely to perceive transparency, the objects perceived as transparent and thus the spatial layout of the objects ([Anderson 1997]). X-junctions, in which the borders of two objects cross in an 'X' fashion, are considered as the most informative regarding the perception of transparency, as they provide more information about contrast variations between superimposed layers, but T-junctions can also elicit a distinct illusion of transparency ([Masin 2006]). The aforementioned perceptual studies focus on achromatic transparency, i.e. regard objects with a luminance and a transmittance components, but color components can be added in the equations, and conditions for chromatic and achromatic transparencies can be jointly studied, like in [Fulvio et al. 2006].



**Figure 1:** The two-cylinder stimulus of experiment 1. Figure (a) shows a 3/4 perspective view of the two filled cylinders at different depths, with the large one being empty at its central end. The six main stimuli subjects see during the experiment are illustrated on Figure (b), with the large cylinder being either bright (top) or dark (bottom); for each brightness condition, the large cylinder stands in front (left column) or in the back (right column) of the small one, or the two cylinders are positioned at the same depth (central column).

## 4 Experiment 1: static perception in DVR

### 4.1 Experimental setup

In this first experiment, we question our ability to perceive the spatial layout of volumetric objects rendered through texture-based DVR techniques. We conduct a three-alternative forced-choice experiment in which participants are asked to decide how semi-transparent cylinders are organized in depth. The stimulus we use is intentionally simple, in order to promote transparencies as the unique cue to depth.

#### 4.1.1 Stimulus description

The scene stimulus (Figure 1) is composed of two cylinders with different radii and same length, the larger one (LC) being twice as thick as the smaller one (SC). In frontal view, the two cylinders lie horizontally, vertically centered; one stands on the left side of the display, the other on the right side. They overlap in image space in a small region that looks like a square-shape intersection area in the middle of the screen. Three dispositions in depth are possible: large cylinder in front of the small cylinder (depth condition  $Depth_{large}$ ), small cylinder in front of the large cylinder ( $Depth_{small}$ ), and the two cylinders at same depth ( $Depth_{intersect}$ ). In this last case, the small cylinder partially extends within the large one.

The two cylinders are filled with a constant scalar value  $s$ , making them appear either bright ( $s = 1$ ) or dark ( $s = 0.2$ ). All voxels outside the cylinders are set to  $s = 0$ , and the part of the LC intersected by SC in frontal view is extruded in all cases, with  $s = 0$ , so that the intersection image always correspond to the same amount of matter accumulated in depth, whatever the geometric organization of the cylinders might be. Preliminary tests pointed out the need to reinforce the volume appearance of the cylinders when filled with scalar value  $s = 0.2$ . To achieve this, we simulated pseudo-lighting in the volumes: the scalar value  $s(x)$  of a voxel  $x$  in a cylinder is given by:  $s(x) = s_0 + aR(x) \cdot L$ , where  $R(x)$  is the normalized direction of  $x$  from cylinder axis and  $L$  the direction of a pseudo-light ( $s_0 = 0.2$  or  $s_0 = 1$ ,  $a = 0.1$ , light coming from above). Values of  $s(x)$  are clamped to  $[0; 1]$ .

There are thus six predefined position conditions for the cylinders: two for laterality (large on the left or on the right) and three for

ordering in depth, and two conditions of luminance: LC bright and SC dark, and the converse. Volume data were precomputed on a regular grid, and the required configuration was loaded on a 3D texture before each trial, and displayed in real-time using OpenGL shaders. For rendering, 100 planes regularly spaced out in the scene cube were blended in back to front order. The Transfer Function we use is linear in luminance and constant in alpha:  $\phi(s) \rightarrow (s, 0.25)$  (see Section 2).

As mentioned in introduction, we choose to use orthographic projection; this is important, as otherwise the cylinders would vary in apparent size in the different depth configurations, due to linear perspective. Besides, that permits to limit important contour information to a few junctions, which constitute fundamental conditions for phenomenal transparency (see Section 3), and to remove potential 3D mental reconstructions interpolated from apparent sizes and intersections configurations. In our images, only X-junctions and luminance information of the intersection area provide relevant information for decision. Also, in an attempt to limit the influence of the contribution of the background color, we place a white-noise image at the rear; the noise is centered at a value corresponding to the average of screen intensities within dark and bright cylinders (white-noise defined in range  $[0.2 : 1]$ ).

#### 4.1.2 Experimental procedure

Subjects are asked for each stimulus to decide how the cylinders are organized in depth. Proposed answers are: "Left cylinder in front", "Two cylinders intersecting", "Right cylinder in front".

During the experiment, subjects' gaze fixations are monitored with an EyeLink<sup>®</sup> II eyetracking system, that also ensures they are correctly looking at the screen. They sit 0.8 m from a 19-inch LCD screen (resolution 1280x1024), head reposing on a chin-rest. Stimulus image subtends approximately  $17^\circ$  of visual angle, and the intersection of the cylinders  $2^\circ$ . For each participant, a calibration of the eyetracker is performed, then the experiment can start.

An experimental session consists of 5 presentations of each of the 12 possible configurations, the full sequence being randomized. In each of the 60 trials, a red cross first appears in the center of a blank screen, during 1 to 3 seconds (random time). The subject is asked to fixate the cross, then the stimulus is presented for two seconds, followed by the answer screen; gaze positions on the screen are monitored during stimulus display. The subject makes his choice

by mouse-clicking on one of the three answers that are displayed on the screen. Then, the next trial starts. At the end of the experiment, subject's reactions and impressions about the task are recorded.

Finally, a small test is carried out, to determine if the difficulty to perform the task can be related to a non-discriminable difference of contrasts between the intersection areas in the different trial configurations. For each of the two luminance conditions (LC dark or LC bright), the three images corresponding to the square intersection area in the separate depth conditions are extracted, and the subject is asked to order these images from brighter- to darker-looking. As these images correspond only to the cylinders intersection area, and thus do not contain any transparency cues, subjects are not disturbed by the depth judgments they have made previously. All subjects answered immediately and correctly.

Stimulus presentations last 2 seconds each. A pilot study showed that longer times (we tested 5 seconds duration) bias answers. From our observations, subjects elaborate more complex per- and between-image cognitive strategies. We thus reduced presentation times to 2 seconds, in an attempt to reinforce the part of perception against conceptual reasoning in subjects' decisions, as we focus on the ability of the human visual system to perceive depth-organization of semi-transparent objects.

### 4.1.3 Participants

Ten subjects, ranging in age from 22 to 39, participated in the experiment. They were all naive as to the goals of the study, and only two of them had experience with volume rendering. All had normal or corrected-to-normal vision. Before they went through the true trial sequence, they read the experimental protocol, practiced on a few trials without head-tracking (trials order differed from the real experience) and then were shown again top-front three-quarter views of some of the stimuli (Figure 1, (a)). This ensured they correctly understood the task, and would keep in mind all the geometric and volumetric properties of the stimuli.

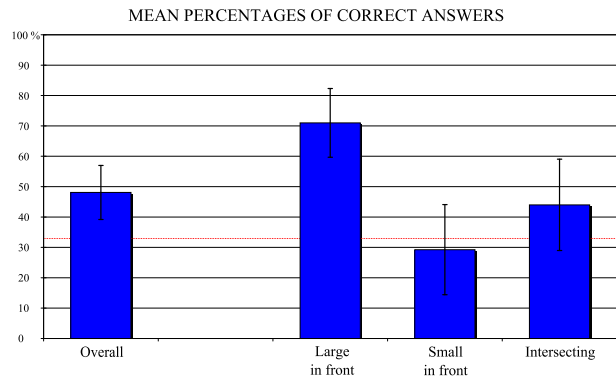
## 4.2 Results

Overall and per-condition mean performances are computed for each participant. Four of the 10 subjects passed a second experimental session with a different trials order after a few minutes break. In this case, we compute the average scores for the two sessions considered as one; we thus end with 10 result sheets.

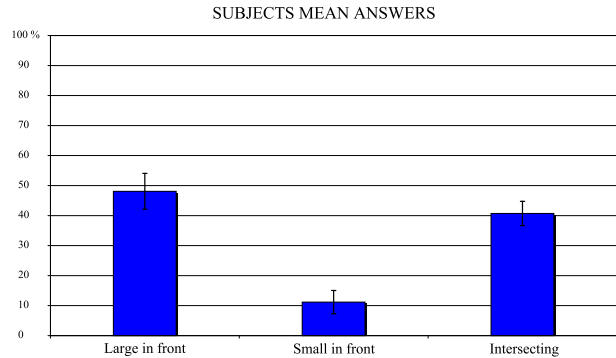
Averages are illustrated in Figure 2. In this depth-ordering task, subject performances are low, with only 48,1% average correct answers, but clearly above chance level ( $p < 0.01$ ). If we analyze the results with respect to depth configuration, we notice levels of 71% for case  $Depth_{large}$ , 29.3% for  $Depth_{small}$  and 44% for  $Depth_{intersect}$ . These results are above chance-level when the large cylinder is presented in front of the small ( $p < 0.01$ ) or when the two intersect, but not when the small lies in front of the scene.

A first explanation of these global results can be found in the mean reported answers (Figure 3). In the experiment, participants are biased toward seeing the large cylinder in front of the display (48.4% of all answers) or the two cylinders intersecting (40.4%), and reluctant to imagine the small cylinder lying closer to the eye (only 11.2% of answers). We also investigated if the respective luminosity of the two cylinders have any influence on subjects' choices, but performances when the large cylinder appears dark or bright look very similar (Figure 4).

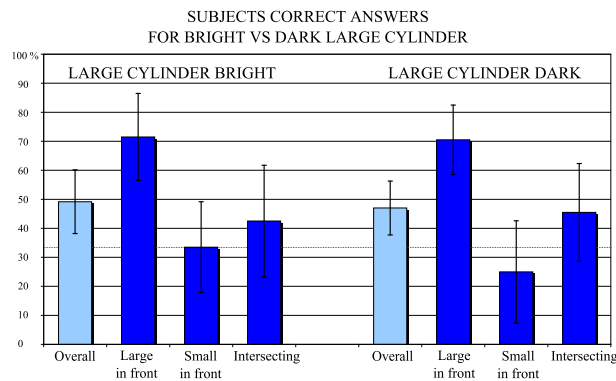
It could be argued at first sight that the globally low level of answers giving the small cylinder closer to the eye can be explained by the fact that the large cylinder, occupying more screen space



**Figure 2:** Mean correct answers averaged over all participants and detailed for the three depth configuration in experiment 1. Error bars represent 95% confidence intervals. Participants overall performances were low, only slightly above chance level. More correct answers are recorded when the large cylinder lies in front, whereas guesses for the small cylinder in front do not differentiate from chance. Intersecting cylinders were detected with little more than chance level.

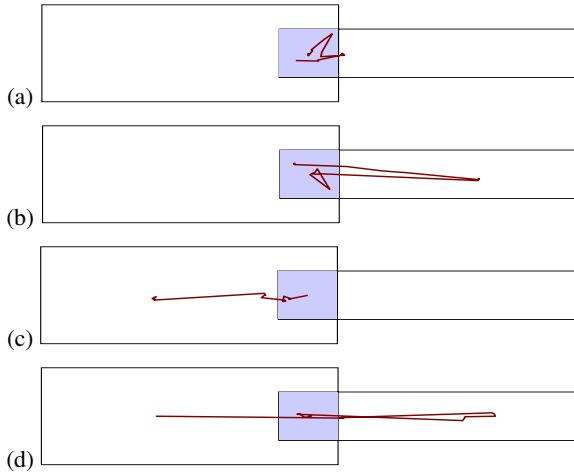


**Figure 3:** Participants mean answers for the three depth configurations in experiment 1. Error bars represent 95% confidence interval. An important bias towards seeing the large cylinder in front can be noticed, whereas subjects are less keen on placing the small cylinder closer to the eye.



**Figure 4:** In experiment 1, mean correct answers are coherent across the two brightness conditions (large cylinder bright or dark).

(twice as much as the small does), is likelier to attract viewer’s attention, which could force “large in front” answers. However, a visual analysis of participants’ gaze trajectories during the trials reveals that in average more saccades fall into the small cylinder than into the large. Four types of ocular trajectories are mainly observed: fixations limited to the intersection area ( $EM_{in}$ ), center-to-small cylinder exploration ( $EM_{small}$ ), center-to-large cylinder exploration ( $EM_{large}$ ) and between-cylinders exploration ( $EM_{both}$ ). Examples of each category are illustrated on Figure 5. For all subjects but one,  $EM_{small}$  dominate over  $EM_{large}$ , with an approximate 3:1 average ratio ( $\overline{EM}_{small} = 15.1$  and  $\overline{EM}_{large} = 5.2$ , with 60 trials per experiment).  $EM_{large}$  and  $EM_{both}$  are globally equal in size, but with noticeable between-subjects disparities. Gaze fixations limited to the central intersection area represent approximately half of the trials; they generally correspond to a fixation of either the lateral borders of the cylinder, either one of the two X-junctions present in the stimulus.



**Figure 5:** Examples of the main gaze profiles recorded across trials in experiment 1. Subjects would focus on the intersection area (a), or visit the small (b) or the large (c) cylinder exclusively, or jump from one to another (d).

### 4.3 Discussion

In this experiment, we try to evaluate our visual ability to perceive the correct organization in depth of semi-transparent objects presented statically and rendered through DVR. Such static presentations are likely to occur for large datasets in scientific visualization, due to the computational costs associated with DVR. The stimulus we use is very simple and results in images that clearly differentiate in regard to contrast, but overall performances prove to be relatively poor, although above chance-level. Results collected in this experiment show that our perception of the organization in depth of overlapping semi-transparent objects is weak, and influenced by factors other than the sole combination of luminances.

Useful visual information is limited to contrast levels at objects borders and intersections, mainly at the central ends of both cylinders. Orthographic projection eliminates all potential geometric cues. The stimulus was chosen as simple as possible to reduce the influence of complex cognitive strategies on the results. Scenes built from more complex objects may reveal more borders, but the relative spatial continuity of the data often leads to a blurring of these information that weakens even more depth perception.

Thus, traditional static transparent renderings may not appear as

effective candidates to provide a clear understanding of the spatial structure of a volumetric dataset, even if the rendering equation takes into account the depth ordering of the data, as DVR does.

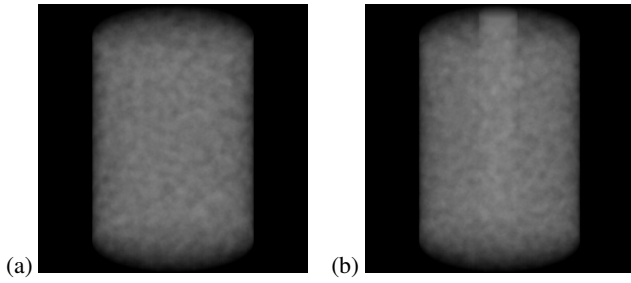
Some factors might explain at least part of our results. Firstly, the stimulus we use is highly asymmetric, due to the difference in size of the two cylinders. We tried to compensate for this point by alternatively presenting the large cylinder on the left and on the right of the display. On the other hand, this asymmetry is a desired artifact, as it provides viewers with two X-junctions, a very important cue to transparency (see Section 3). These X-junctions correspond in our case to the intersections between the vertical end of the large cylinder and the horizontal limit of the small cylinder. Subjects may be biased toward seeing the X-junctions as the unique transparency border, whereas the central end of the small cylinder is also a transparency border itself. An overview of gaze trajectories limited to the intersection area make us think this is not the case, but we cannot be totally confident on this point. If subjects find it easier – consciously or not – to interpret the image as containing only one transparent cylinder, the fact that at the X-junction the horizontal border of the small cylinder is larger than the vertical border of the large cylinder might facilitate the perception of LC as more transparent than SC. Also, the large cylinder offers a wider surface through which more background details can be seen, which might strengthen this effect. And if only one cylinder is perceived as transparent, then it will definitely appear in front of the other.

By reducing stimulus presentation time to only two seconds, we wished to limit as much as possible the influence of more cognitive approaches in the decision process. However, these cannot be totally rejected, as post-experiment introspective questions showed. For instance, many participants reported that one stimulus often influenced the decision they made for the following, which was possible when consecutive stimuli were close enough (only depth condition modified). On the other hand, all but two participants also stated that they were highly uncertain about the answers they gave, and that consecutive stimuli could as well disturb them about previous answers they gave. The limited presentation time also restricts the possibility to build firmly anchored strategy based on progressive classification of the luminance configurations. We cannot reject the assumption that short-term memory plays a role in the present experiment, but its effect is not expected to modify consistently the recorded performances.

Answer times were monitored during the experiment, but not taken into account for the analysis. Subjects were however instructed to answer as fast as possible, which was generally the case (average answer time: 1.9 seconds), but no mechanism was introduced to force quick answers. However, mean reaction times show no linear correlation with overall performances ( $R^2 = 10^{-4}$ ). We observed a limited bias during experimental sessions, with participants sometimes positioning the mouse on the answer corresponding to “small in front”, and then switching to another choice.

## 5 Experiment 2: importance of dynamic information in different DVR techniques

Experiment 1 focuses on the perception of depth order of semi-transparent objects displayed statically. In this second experiment, we look at how volume rendering techniques can convey information about the spatial layout of the supported data in a dynamic context. More precisely, we tested the influence of the Transfer Function that associates an opacity and a luminance to every voxel scalar value in order-dependent DVR (see Section 2), and compared performances to order-independent methods, namely X-Ray and MIP. We tried to measure the conditions for volume rendering techniques to provide non-ambiguous depth cues in a dynamic



**Figure 6:** The rotating cylinder filled with Perlin noise used in experiment 2, seen from top, with DVR rendering (a). To choose acceptable rendering parameters, we required that a sub-cylinder filled with maximal value be always visible (b).

display. MIP and X-Ray can be faster techniques, as they don't require any ordering of data in depth, but as such they don't provide any depth cues, which produces strong ambiguities in the display.

## 5.1 Experimental setup

### 5.1.1 Stimulus description

For all rendering conditions, the same stimulus is used, based on [Kersten et al. 2006] previous work: a vertical cylinder, orthographically projected on the screen, rotates either to the left or to the right in frontal view (rotation speed:  $34^\circ/s$ ). Such a stimulus relies on [Wallach and O'Connell 1953] experiments, showing that a 3D shape projected on a screen can be correctly perceived when the object rotates around a vertical axis; however, two mirror shapes can be seen, as in the Necker cube, so that a viewer experiences an inversion of the rotation direction when he jumps from one perception to the other. The perceived rotation direction thus reveals our underlying perception of the volume, the occurrence of change reversals indicating an ambiguous stimulus.

The cylinder is filled with volume data generated from Perlin noise functions [Perlin 1985], while exterior is set to 0. The advantage of this function is to provide non-homogeneous data that vary continuously, without any identifiable structures in the volume that could cognitively bias subject's decisions. The Perlin noise value  $P(x)$  in a voxel  $x$  is defined by:

$$P(x) = \sum_{i=0}^{n-1} \frac{N(b^i x)}{a^i} \quad (4)$$

where  $N(x)$  is the basic Perlin noise function, and  $b$  and  $1/a$  define the relative frequency and persistence of the summed harmonics, respectively. We chose  $a = b = 2$  and  $n = 4$ . Volume data are stored in a 3D texture, and rendering is performed through 100 planes accumulated back to front. Animation images of the rotating cylinders are precomputed, to guarantee that the stimuli will effectively be presented with a constant refresh rate of 60 images/s.

Six rendering conditions are defined: X-Ray, MIP and four DVR with different parameters for the Transfer Function. TF tuning remains a very delicate problem in volume rendering, due to the huge number of candidates. In this experiment, we decided to test two particular cases, one Transfer Function linear in luminance (we will call it  $DVR_l$ ), the other linear in opacity ( $DVR_\alpha$ ):

$$(DVR_l) \quad l(s) = s \quad , \quad \alpha(s) = \alpha_l \quad (5)$$

$$(DVR_\alpha) \quad l(s) = 1 \quad , \quad \alpha(s) = \alpha_\alpha s \quad (6)$$

Two values are chosen for  $\alpha$  in each condition, based on a simple transparency test: we placed inside our main cylinder, tangent to its border, a small cylinder of same height but four times less thick. Scalars inside this test cylinder were given a value of 1, and values of  $\alpha$  were chosen such that this cylinder could be seen when located in front and at the rear of the main cylinder. This approximately corresponds to situations where maximal values can always be perceived, no matter where they stand. We ended with  $\alpha_l = 0.01$  and  $\alpha_l = 0.025$  in  $DVR_l$  condition, and  $\alpha_\alpha = 0.015$  and  $\alpha_\alpha = 0.025$  in  $DVR_\alpha$  condition. We will further refer to these conditions as  $DVR_l^{0.01}$ ,  $DVR_l^{0.025}$ ,  $DVR_\alpha^{0.015}$  and  $DVR_\alpha^{0.025}$ , respectively.

Each of the six rendering conditions is implemented for a cylinder rotating to the left and to the right of the display, which results in 12 conditions.

### 5.1.2 Experimental procedure

Subjects are asked to determine the perceived direction of rotation of the cylinder, by clicking with the mouse on the icon corresponding to their choice (either "to the left" or "to the right", as they could perceive it for the 'front surface' of the volume).

Experimental setup is the same as in experiment 1 (Section 4.1.2), with subject's gaze position tracked during stimulus presentation. The rotating cylinder occupies  $14^\circ$  fov vertically,  $10^\circ$  fov horizontally, resulting in a maximum speed of  $3^\circ/s$  in visual field.

An experimental session consists of 5 presentations of each of the 12 possible configurations, the sequence of trials being randomized. Each of the 60 trials starts with the red cross fixation, followed by a presentation of the rotating stimulus in one of the defined conditions. After 0.5 second, the answer screen is displayed and the subject decides the rotation direction. Then the following trial immediately starts. At the end of the experiment, subject's reactions and impressions about the task are recorded.

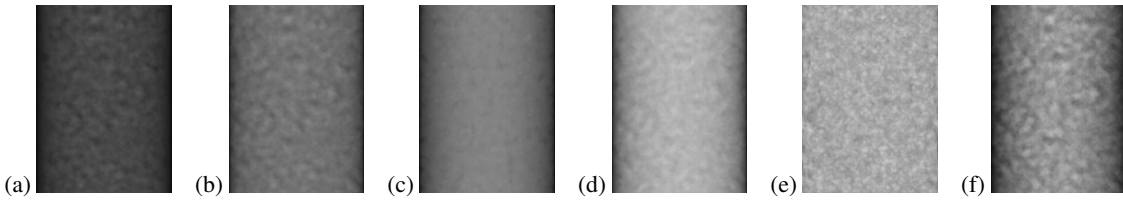
The short presentation time – half a second – limits the change reversals of rotation direction perception. This eliminates the influence of perception bias in case of ambiguities; for instance, if two directions were successively experienced for the same stimulus, then one might tend to answer more frequently "to the right" than the converse. Such a preference has been noticed in the experiment, with 65% answers corresponding to "to the right".

## 5.2 Results

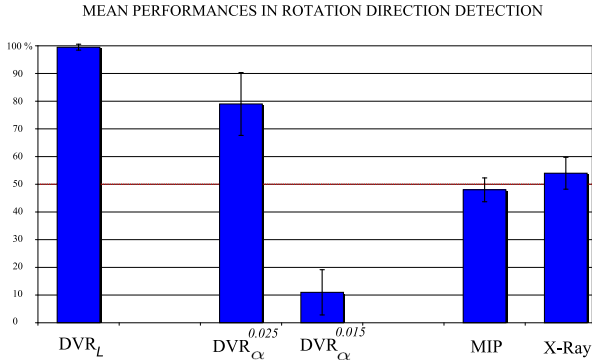
Ten subjects participated in this experiment, the same who volunteered for experiment 1.

Performances for MIP and X-Ray conditions are very close to chance-level (48% and 54% correct answers, respectively), which is not a surprise as they don't provide viewers with any depth cues, being order-independent techniques. This validates that our stimulus does not include any cue that can be used to solve the task.

DVR leads to very different performance levels, depending on the nature of the implemented TF. Condition  $DVR_l$  ( $\alpha$  constant) presents no ambiguity, with 99.5% correct answers. On the contrary, performances for an alpha linear with respect to the scalar value are much more contrasted, with 11% and 79% correct answers for conditions  $DVR_\alpha^{0.015}$  and  $DVR_\alpha^{0.025}$ , respectively. The value 0.015 shows a strong ambiguity in the visual display, as performances lie far below chance level, which can be interpreted as: "data at the rear of the volume are perceived closer to the eye as those in front". A value of 0.025 reveals more correct depth perception within the volume, but this rendering is still prone to ambiguities, as performances clearly differentiate from perfection ( $p < 0.01$ ).



**Figure 7:** The 6 cylinder renderings implemented in experiment 2, in frontal view. (a) and (b) correspond to DVR with Transfer Function linear in luminance (conditions  $DV R_l^{0.01}$  and  $DV R_l^{0.025}$ ), (c) and (d) to DVR with Transfer Function linear in opacity (conditions  $DV R_\alpha^{0.015}$  and  $DV R_\alpha^{0.025}$ ), (e) is MIP rendering and (f) X-Ray



**Figure 8:** Correct answers in experiment 2 for the different rendering conditions, with pooling over both levels for  $DV R_L$ . Only DVR with luminance-based Transfer Function provide non-ambiguous depth perception within the cylinder, whereas linear-alpha TF can be highly unreliable.

### 5.3 Discussion

Results obtained in this experiment prove that in a dynamic context DVR can lead to a strong perception of the organization in depth of volumetric data, but that this can be achieved only through a careful tuning of the Transfer Function (TF). Given the size of potential TF space (for integer scalar values in range [0:255] mapped to 1-byte luminance and 1-byte alpha, there are already  $(256 * 256)^{256}$  possibilities !), we restricted our analysis to as few as two TF subclasses, those linear in luminance and constant in opacity, and the converse. We have shown that the former produce reliable renderings, whereas the latter are more prone to elicit ambiguous perception of the spatial layout of the data. In general cases, the tuning of the TF is a very complex problem that is solved empirically, even if some attempts to propose automatic methods have been proposed (for reference, see [Pfister et al. 2000]). Today TF are getting more and more complex; non-linearities are often introduced, as well as multiple dimensions. Our results, obtained on the very simple case of linear unidimensional functions, cannot be easily generalized to all potential TF. However, we explicitly show that, even for order-dependent volume rendering, a Transfer Function resulting in a correct discrimination of objects within a volume can disturb our perception of the spatial layout of these objects.

The different techniques and sets of parameters produce renderings with different contrast ratios, that were not equalized over the conditions (see Table 1). This could be a factor modifying the perception of the cylinder motion. For instance, a moving target with very low contrast is likely to embed areas seen as flat in luminance, and

fixations to such patterns would lead to a percept close to immobility. In the present experiment, however, participants reported that they always perceived motion, and only few of them felt that speed might have varied across trials, with slower rotations in condition  $DV R_\alpha^{0.015}$ , which of course was not the case.

## 6 General discussion and conclusion

The display of space filling data remains a challenge, as it requires on the one hand to find renderings that does not hide much of the data to the viewer, which is the main drawback of surfacic techniques, while revealing without ambiguities the organization in depth of these data. Various methods have been proposed, the major class being based on physical (optical) models. However complex and informative they might appear, these renderings implicitly assume that the human visual system produces accurate perceptions of transparent volumes, whereas we seldom face such visual images in our daily life. Interestingly, we find the same physical roots in early psychological works on the perception of transparency, that were based on the experimental framework of the episcotister. But, as mentioned by [Singh and Anderson 2002], the relationship between perceptual theory and generative models is often left rather vague, and their experimental work emphasizes the deviation between perceptual transparency and Metelli's episcotister model. In a similar way, we showed that using models of light propagation in dense matter does not necessarily elicit a correct perception of the spatial layout of even simple structures through static presentations. Indeed, it appeared that the human visual system is not tuned to interpret correctly such representations, and may be strongly biased.

However, these order-dependent methods operate reasonably well in a dynamic context, provided that Transfer Functions are carefully chosen. This result underlies the need to focus on methods that generate fast renderings, with at least 15 frames per second to preserve short range motion perception, thus providing viewers with strong dynamic cues to depth. As the underlying algorithms are very costly, requiring intensive per-pixel computations, two approaches can be considered. At first, the development of programmable graphic hardware (generalized shaders) opens new prospects on the speed of computations. Several works have already been performed in this direction (e.g., [Roettger et al. 2003]). On the other hand, it is possible to reduce the amount of data displayed per image, considering that our attention has limited capacity. For instance, psychological experiments have shown that we are unable to track more than four objects at once, due to the limitation of our attentional system ([Sears and Pylyshyn 2000]). The definition of objects in a volumetric data field can be quite difficult, and it should also not reduce to classical isosurface computations. New methods have to be found to facilitate the analysis of space filling data, and this will be all the more crucial as scientific visualization will face larger and larger amounts of data, due to reduced computational costs: perceptually enhanced visualizations will become one of the major challenges of tomorrow's engineering systems.

Rendering	MIP	X-Ray	$DVR_l^{0.01}$	$DVR_l^{0.025}$	$DVR_a^{0.015}$	$DVR_a^{0.025}$
Mean pixel intensity (in [0:255])	170	145	73	114	128	176
Michelson contrast	0.174	0.235	0.219	0.155	0.1245	0.113

**Table 1:** Michelson contrasts for the different renderings implemented in experiment 2.

Future works will include further analysis of our perception of depth in transparent media. In particular, it would be interesting to understand more thoroughly the bias toward seeing large objects in front of smaller ones that was observed in experiment 1. Also, we will clarify the results of experiment 2, and analyze the reasons of the reversal or attenuation of depth perception for dynamic displays with alpha-linear TF. As regards depth cues, it would be interesting to study more deeply the influence of perspective projection in volume rendering, as we limited our work to the orthographic case. Finally, it is also worth mentioning the development of easier-to-use stereoscopic displays; further experiments exploring the power of stereoscopy in DVR have to be carried out.

## Acknowledgments

This work was funded by eDF and a CIFRE grant from ANRT. The authors thank the persons who kindly gave time to participate in the experiments, and reviewers for useful comments.

## References

- ANDERSON, B. L. 1997. A theory of illusory lightness and transparency in monocular and binocular images: the role of contour junctions. *Perception* 26, 4, 419–453.
- EBERT, D., AND RHEINGANS, P. 2000. Volume illustration: Non-photorealistic rendering of volume models. In *Proceedings Visualization 2000*, T. Ertl, B. Hamann, and A. Varshney, Eds., IEEE Computer Society Technical Committee on Computer Graphics, 195–202.
- ENGEL, K., KRAUS, M., AND ERTL, T. 2001. High-quality pre-integrated volume rendering using hardware-accelerated pixel shading. In *HWWS '01: Proceedings of the ACM SIGGRAPH/EUROGRAPHICS workshop on Graphics hardware*, ACM Press, New York, NY, USA, 9–16.
- FULVIO, J. M., SINGH, M., AND MALONEY, L. T. 2006. Combining achromatic and chromatic cues to transparency. *J Vis* 6, 8, 760–776.
- GERBINO, W., STULTIENS, C. I., TROOST, J. M., AND DE WEERT, C. M. 1990. Transparent layer constancy. *J Exp Psychol Hum Percept Perform* 16, 1 (Feb), 3–20.
- HIBBARD, B. 2000. Confessions of a visualization skeptic. *Computer Graphics* 34, 3 (Aug.), 11–13.
- INTERRANTE, V., FUCHS, H., AND PIZER, S. M. 1997. Conveying the 3D shape of smoothly curving transparent surfaces via texture. *IEEE Trans. Vis. Comput. Graph* 3, 2, 98–117.
- KAJIYA, J. T., AND HERZEN, B. P. V. 1984. Ray tracing volume densities. In *SIGGRAPH '84: Proceedings of the 11th annual conference on Computer graphics and interactive techniques*, ACM Press, New York, NY, USA, 165–174.
- KASRAI, R., AND KINGDOM, F. A. 2001. Precision, accuracy, and range of perceived achromatic transparency. *J Opt Soc Am A Opt Image Sci Vis* 18, 1 (Jan), 1–11.
- KAUFMAN, A., AND MUELLER, K. 2005. *The Visualization Handbook*. ch. 7, 127–174.
- KERSTEN, M., STEWART, J., TROJE, N. F., AND ELLIS, R. E. 2006. Enhancing depth perception in translucent volumes. *IEEE Trans. Vis. Comput. Graph* 12, 5, 1117–1124.
- LACROUTE, P., AND LEVOY, M. 1994. Fast volume rendering using a shear-warp factorization of the viewing transformation. In *SIGGRAPH*, ACM, 451–458.
- LEVOY, M. 1988. Display of surfaces from volume data. *IEEE Computer Graphics and Applications* 8, 3, 29–37.
- MASIN, S. C. 2006. Test of models of achromatic transparency. *Perception* 35, 12, 1611–1624.
- MAX, N. 1995. Optical models for direct volume rendering. *IEEE Transactions on Visualization and Computer Graphics* 1, 2, 99–108.
- METELLI, F. 1974. The perception of transparency. *Sci Am* 230, 4 (Apr), 90–98.
- MORA, B., AND EBERT, D. S. 2004. Instant volumetric understanding with order-independent volume rendering. *Comput. Graph. Forum* 23, 3, 489–498.
- PERLIN, K. 1985. An image synthesizer. *Computer Graphics* 19, 3 (July), 287–296.
- PFISTER, H., LORENSEN, B., SCHROEDER, W., BAJAJ, C., AND KINDLMANN, G. 2000. The transfer function bake-off (panel session). In *VIS '00: Proceedings of the conference on Visualization '00*, IEEE Computer Society Press, Los Alamitos, CA, USA, 523–526.
- ROETTGER, S., GUTHE, S., WEISKOPF, D., ERTL, T., AND STRASSER, W., 2003. Smart hardware-accelerated volume rendering.
- ROPINSKI, T., STEINICKE, F., AND HINRICHS, K. H. 2006. Visually supporting depth perception in angiography imaging. In *Proceedings of the 6th International Symposium on Smart Graphics (SG06)*, Springer, Vancouver, 93–104.
- SABELLA, P. 1988. A rendering algorithm for visualizing 3d scalar fields. In *SIGGRAPH '88: Proceedings of the 15th annual conference on Computer graphics and interactive techniques*, ACM Press, New York, NY, USA, 51–58.
- SEARS, C. R., AND PLYSHYN, Z. W. 2000. Multiple object tracking and attentional processing. *Can J Exp Psychol* 54, 1 (Mar), 1–14.
- SINGH, M., AND ANDERSON, B. L. 2002. Toward a perceptual theory of transparency. *Psychol Rev* 109, 3 (Jul), 492–519.
- SINGH, M. 2004. Lightness constancy through transparency: internal consistency in layered surface representations. *Vision Res* 44, 15, 1827–1842.
- WALLACH, H., AND O'CONNELL, D. N. 1953. The kinetic depth effect. *J Exp Psychol* 45, 4 (Apr), 205–217.