Perceptual transparency from image deformation

Takahiro Kawabe¹, Kazushi Maruya, and Shin'ya Nishida

Sensory Representation Research Group, Human Information Science Laboratory, NTT Communication Science Laboratories, Nippon Telegraph and Telephone Corporation, Atsugi, Kanagawa 2430198, Japan

Edited by Brian A. Wandell, Stanford University, Stanford, CA, and approved July 7, 2015 (received for review January 22, 2015)

Human vision has a remarkable ability to perceive two layers at the same retinal locations, a transparent layer in front of a background surface. Critical image cues to perceptual transparency, studied extensively in the past, are changes in luminance or color that could be caused by light absorptions and reflections by the front layer, but such image changes may not be clearly visible when the front layer consists of a pure transparent material such as water. Our daily experiences with transparent materials of this kind suggest that an alternative potential cue of visual transparency is image deformations of a background pattern caused by light refraction. Although previous studies have indicated that these image deformations, at least static ones, play little role in perceptual transparency, here we show that dynamic image deformations of the background pattern, which could be produced by light refraction on a moving liquid's surface, can produce a vivid impression of a transparent liquid layer without the aid of any other visual cues as to the presence of a transparent layer. Furthermore, a transparent liquid layer perceptually emerges even from a randomly generated dynamic image deformation as long as it is similar to real liquid deformations in its spatiotemporal frequency profile. Our findings indicate that the brain can perceptually infer the presence of "invisible" transparent liquids by analyzing the spatiotemporal structure of dynamic image deformation, for which it uses a relatively simple computation that does not require high-level knowledge about the detailed physics of liquid deformation.

perceptual transparency | image deformation | material perception

The human visual system is able to represent the depth stratification of multiple layers at the same retinal location. The best demonstration of this ability is perceptual transparency, in which a transparent layer is seen in front of a background surface (1–4)*. The critical cue to perceptual transparency has been considered to be the pattern of luminance (and color) change at the junction of overlapped layers (1, 3)*. Such luminance cues arise in natural scenes because the light reflected from the background layer is partially absorbed or scattered by the overlapping transparent layer; thus, the cues might not work effectively when the layer is made of a highly transparent material that neither strongly absorbs nor scatters light, such as clear glass or water. This leads to the following question: Can we perceive these sorts of pure transparent layers that do not produce conventional luminance cues?

A potential alternative cue for perceptual transparency is image deformation caused by refraction. When a transparent layer has a refractive index >1 , the background image optically deforms in accordance with the 3D shape of the layer surface. The image deformation due to refraction by itself is considered an ineffective cue to the perception of a transparent layer (5), although the magnitude of the deformation could be a cue to the perception of the thickness of a transparent layer (6). Previous studies have examined the effect of the deformation cue only in static images, however.

Here we report that the dynamic image deformation of an image texture, as caused by clear water running over a textured surface, produces the vivid and compelling perception of a transparent liquid even when other cues to the presence of a liquid, such as surface specular reflections and intensity modu-

lations due to caustics, are excluded from the stimuli. This finding provides scientific evidence that the dynamic deformation is a strong cue to the perception of a transparent layer. Furthermore, it indicates that the pattern of dynamic deformation includes information sufficient for human observers to recognize the "material" of the transparent layer.

We carried out a series of psychophysical experiments to reveal the visual processing underlying the perception of a transparent layer from dynamic image deformations. We found that the transparent liquid perception requires not only a sequence of static deformations, but also low-level motion signals accompanied by dynamic deformation. In addition, we found that a transparent liquid was perceived not only when we took the pattern of dynamic deformation from a real or physics-based computer simulation of water flow, but also when we synthesized the pattern of deformation from a random source, in which we only made the spatiotemporal frequency amplitude spectrum of the image deformation close to that of real dynamic water. Our findings indicate that the brain perceptually infers the presence of transparent liquids from a rather simple analysis of low-level spatiotemporal statistics of dynamic image deformation, instead of cognitively judging it based on high-level knowledge about water.

Results

Seeing a Transparent Layer from a Pure Image Deformation. We first explored whether human observers could see a transparent liquid solely from the image deformation of an underlying pattern. Using Blender software, we rendered computer-graphics scenes (Movie S1) simulating the flow of a transparent liquid with the refractive index of water (1.33), and created video clips, each of which consisted of 90 frames and lasted 3 s. We eliminated specular reflections at the liquid surface to examine the pure effects of image deformation. In the resulting video clips, it looked

Significance

The perception of liquids, particularly water, is a vital sensory function for survival, but little is known about the visual perception of transparent liquids. Here we show that human vision has excellent ability to perceive a transparent liquid solely from dynamic image deformation. No other known image cues are needed for the perception of transparent surfaces. Static deformation is not effective for perceiving transparent liquids. Human vision interprets dynamic image deformation as caused by light refraction at the moving liquid's surface. Transparent liquid is well perceived from artificial image deformations, which share only basic flow features with image deformations caused by physically correct light refraction.

Author contributions: T.K., K.M., and S.N. designed research; T.K. and S.N. performed research; T.K. analyzed data; and T.K., K.M., and S.N. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

¹To whom correspondence should be addressed. Email: kawabe.takahiro@lab.ntt.co.jp.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1500913112/-DCSupplemental.

*Adelson EH, Anandan P. Ordinal characteristics of transparency. Proceedings of the AAAI-90 Workshop on Qualitative Vision, July 29, 1990, Boston, MA, pp 77–81.

as if a transparent liquid were flowing over a static background pattern. **Movie S1** shows an example of these. We presented these video clips to human observers and asked them to rate the strength of their transparent liquid impression on a five-point scale (1, no impression of a transparent liquid, to 5, vivid impression of a transparent liquid). The video clips received a very high score (4.48, “original video clip” condition in Fig. 1C), indicating that the impression of a transparent liquid was clear and compelling to them.

Importance of Dynamic Image Deformation in Seeing a Transparent Layer. To analyze the critical stimulus characteristics of this phenomenon, we tested three other conditions. One of these was a “static image” condition, in which we randomly chose a single video frame from each video clip and presented it to the observer for 3 s. Another was a “no-motion” condition, in which we sampled one video frame every four frames from the original video clips. Each sampled frame was presented for 33 ms, followed by a uniform gray frame of 100 ms. Such a large inter-frame interval is known to impair low-level motion detection (7, 8). Unlike for the static image condition, the observers were shown more examples of deformation patterns present in the original video clips. The duration of the video clips in the no-motion condition was 3 s. The third condition was a “shuffled” condition, which was the same as the original video clips condition except that the order of video frames in an original simulation video clip was randomized in each trial, which destroyed the temporal structure of the originals.

We performed repeated-measures one-way ANOVA on the ratings averaged across observers, with the four experimental conditions as factors. The main effect was highly significant [$F(3, 21) = 15.084$; $P < 0.0001$]. Multiple comparison tests showed that the original video clips condition was significantly different from

the static image condition ($P < 0.0001$) and the no-motion condition ($P < 0.0001$), but not from the shuffled condition ($P > 0.05$). The shuffled condition was significantly different from the static image condition ($P < 0.0001$) and the no-motion condition ($P < 0.0001$).

The low rating score for the static image condition suggests that a static deformation image cannot provide a clear impression of transparent liquid, in agreement with a previously reported suggestion (5). A low score for the no-motion condition suggests that multiple deformation images cannot provide a clear impression of transparent liquid either, and that low-level motion signals detected by low-level motion detectors play a critical role in the perception of transparent liquid. Finally, a high rating score for the shuffled condition indicates that randomization of frame order did not hamper a transparent liquid impression. Thus, it seems that the transparent liquid perception is not very sensitive to the precise temporal structure of deformation in actual transparent liquid flows. We consider the implication of this finding in more detail below.

Analyzing Spatiotemporal Frequencies of Image Deformation. To explore the image information contributing to the recognition of a transparent liquid, we analyzed the spatiotemporal pattern of image deformation. Specifically, we examined whether and how the amplitude spectrum of image deformation represents the relevant image information. As shown in Fig. 24, we first computed optical flow fields between two successive frames in the simulation video clips using a previously reported algorithm (9). The calculated optical flows were decomposed into horizontal and vertical vector components. For each video clip, we performed 3D fast Fourier transformation (FFT) of each of the decomposed vector components using built-in functions in Matlab, and calculated the amplitude spectra of spatiotemporal frequencies of deformation (Fig. 2B). We then averaged the amplitude spectra across eight under-liquid scenes. The outcome analysis showed that the distribution of deformations was spatially low-frequency dominant while temporally broadband. Interestingly, the pattern of spatiotemporal frequencies of image deformation remained similar even when the order of video frames was randomized (Fig. 2C). This suggests the possibility that human perception of transparent liquid perception may be based on this characteristic spatiotemporal amplitude spectrum of image deformation.

Analysis of Real Liquid Flow. To this point, we had used video clips of computer-simulated scenes including a water-like liquid as a stimulus; however, the computer simulation only approximately reproduced image deformations caused by refraction at the surface of a transparent liquid in the real world. Thus, we then analyzed the image deformation of scenes produced by real water flow. The surface of real water in a container was stirred with a plastic stick that moved as a pendulum to produce the flow of water (Fig. 2D). A servomotor controlled the stirring speed at five levels (1, 3, 5, 10, and 15 Hz) while keeping the spatial amplitude of the stirring constant. We shot videos of the water flows from above the surface of the water. Using optical and image processing solutions (details in *Methods*), we removed specular components and caustics from the simulation and successfully isolated dynamic image deformation owing to refraction. Not surprisingly, the resulting dynamic deformation could produce a clear and vivid impression of a transparent liquid (see below).

We calculated the spatiotemporal amplitude spectrum of the image deformation of the real flow and found that, irrespective of stirring speed, the distribution of deformations was spatially low-frequency dominant while temporally broadband (Fig. 2E). These results indicate that the spatiotemporal amplitude spectrum of image deformation produced by the real flow is similar to that produced by the simulated flow.

Fig. 1. The stimuli and results of experiments using computer-graphics simulation video clips as stimuli. (A) Screenshot of the geometrical scene simulated in Blender. (B) Eight static images of under-liquid scenes, which are deformed owing to refraction at the surface of a flowing liquid. It is difficult to see a transparent layer with single static video frames. (C) Results of the first two experiments. Error bar denotes ± 1 SEM.

Fig. 2. Spatiotemporal frequency of image deformation of transparent liquid. (A) We first calculated optical flow fields between two frames. The calculated optical flows were decomposed into horizontal and vertical vector components. The optical flow fields were computed more densely (for each pixel) than shown in this panel. Units for each of the color bars are pixels. (B) Amplitude spectra of spatiotemporal frequencies of deformation in intact video clips. (C) Amplitude spectra of spatiotemporal frequencies of deformation in the movies in which the order of video frames in a simulation video clip was randomized. (D) Photograph of the setup for taking videos of water with the surface stirred with a plastic stick attached to a servomotor. (E) Logarithmic amplitude spectra for the spatiotemporal frequency of image deformation owing to refraction at the surface of real water for each stirring speed. (F) Transparent liquid impression as a function of stirring speed.

Amplitude Spectra of Image Deformation, Not Phase, Is Critical. To address whether the characteristic spatiotemporal amplitude spectrum of image deformation is sufficient and the specific phase spectrum is not necessary for the visual system to recognize a transparent liquid, we set three conditions with video clips showing intact, phase-randomized, and amplitude-randomized deformations (Movie S2). In an intact condition, we reconstructed the image deformation of the real flow by warping the pixels (10) of an underlying pattern (a 2D 1/f noise) on the basis of the optical flow fields of the original liquid flows. In this condition, both the amplitude spectra and phase in the spatiotemporal frequencies of image deformation were intact. In a phase-randomized condition, we randomized the phase in the spatiotemporal frequencies of image deformation of the real liquid flow and warped the pixels of the underlying pattern on the basis of phase-randomized optical flow fields. In an amplitude-randomized condition, we randomized the amplitude spectra in the spatiotemporal frequency of the optical flow fields of the real liquid flow, and warped the pixels of the underlying pattern on the basis of amplitude-randomized optical flow fields.

The observers were asked to view the video clips as stimuli, and rate their transparent-liquid impression on a five-point scale.

The results show that the main effect of the condition was highly significant [$F(2, 14) = 35.822$; $P < 0.000006$] (Fig. 2F). The ratings of the amplitude-randomized condition were significantly different from those of the intact and phase-randomized conditions ($P < 0.00001$). In contrast, the ratings of the intact condition were not significantly different from those of phase-randomized condition ($P > 0.79$). No effect of stirring speed was found [$F(4, 28) = 0.247$; $P > 0.9$]. These results indicate that the human visual system uses the pattern of spatiotemporal frequency amplitude spectra of image deformation to recognize transparent liquid but, on the other hand, human transparent liquid perception is little affected by replacing a physically valid structured phase spectrum with a random unstructured one.

Critical Deformation Spatiotemporal Frequencies for Perceiving a Transparent Layer. To further narrow down the image features used by the visual system for the perception of transparent liquid from dynamic image deformation, in the next experiment we investigated whether the entire range of deformation spatiotemporal frequency was equally effective or whether there were some critical frequency ranges that were more effective than the others. To specify the critical frequencies, we used a psychophysical reverse

correlation technique. In typical reverse correlation techniques, such as Bubbles (11), stimuli are spatially filtered out to identify a diagnostic visual feature to trigger a specific categorical response. Bubbles can be applied to the spatial frequency domain, and has been used to characterize the spatial frequency component that triggers a specific categorization response (12). Previous work also used a reverse correlation technique to specify the subbands of spatiotemporal frequency recovered through a motion-form interaction (13). To specify the spatiotemporal frequency components of image deformation that contributed to the recognition of a transparent liquid (Fig. 3*A* and [Movie S2](#)), we used a technique similar to those used in those previous studies. To add a deformation to an image, we again used a pixel-warping method that was similar to the method used in a previous study (10).

As done previously (13), we calculated correlation coefficients between the binary values for the presence (1) and absence (0) of a transparent liquid impression in each trial and the presence (1) and absence (0) of amplitude spectra at each subband, and plotted the correlation coefficients as a correlogram, shown in Fig. 3*B*, where the yellow and blue cells denote the positive and negative correlation coefficients, respectively. Positive (negative) correlations meant that the subbands aided (inhibited) the perception of transparent liquids. Significant positive and negative correlations ($P < 0.05$) were observed at the lower and higher spatial frequencies of deformation, respectively. Among these, the significant positive correlations occurred mainly at middle temporal frequencies of deformation (3–10 Hz), whereas significant negative correlations occurred at broadband temporal frequencies of deformation ([Movie S3](#)). These results indicate that the visual system does not equally use the entire spatio-

temporal amplitude spectrum of image deformation, but instead mainly uses specific ranges to recognize a transparent liquid. Specifically, an image deformation may be interpreted as an optical image deformation owing to the presence of a transparent liquid when image deformation has a spectrum in the lower spatial and middle temporal bands of deformation spatiotemporal frequency and does not have a spectrum in its higher spatial frequency bands. This simple mechanism can explain why even computer-generated random image deformation could produce the impression of a transparent liquid layer.

If the human visual system indeed uses specific ranges of the spatiotemporal frequency of image deformation as diagnostic features for transparent liquid perception, observers may perceive transparent liquids even when the deformation is limited to a single narrow spatiotemporal frequency band as long as the band falls in the optimal range for transparent liquid perception. The reverse correlation experiment did not directly test this, because it simultaneously presented several spatiotemporal subbands (on average, 10% of 42 bands) in each trial. The experiment did not exclude the possibility that cross-frequency interactions (second-order or even higher-order terms) play critical roles in transparent liquid perception.

To examine this issue, we conducted a supplementary experiment in which we gave a single narrow-band spatiotemporal deformation to a background natural scene. For a variety of narrow-band deformations, we asked the observers to rate the transparent liquid impression. As shown in Fig. 3*C*, the rating scores in the middle range of the spatiotemporal frequency were considerably high (>3.0), significantly higher ($P < 0.05$) than those in other subbands including high temporal frequencies,

Fig. 3. Spatiotemporal frequency tuning of transparent liquid perception. (A) Schematic of the procedure for manipulating the spatiotemporal frequencies of deformation in the reverse correlation experiment. In this example, amplitude spectra of the 3D Gaussian noise pattern passed through three subbands of spatiotemporal frequencies of image deformation. (B) Results of the reverse correlation experiment. In the correlogram, a positive (negative) correlation coefficient value indicates that the presence of the subband positively (negatively) contributes to transparent liquid perception. (C) Spatiotemporal map of the rating for transparent liquid impression, obtained when we gave a dynamic image deformation with a narrow-band spatiotemporal frequency to a natural scene image.

Fig. 4. (A) Screenshot of the geometrical scene simulated in Blender. (A and B) Change in rating scores for transparent liquid impression as a function of camera-to-liquid distance. The results are in agreement with the idea that the deformation amplitude spectrum contains diagnostic image features for transparent liquid perception.

which is in general agreement with the spatiotemporal tuning suggested by the reverse correlation experiment. These results indicate that the presence of deformation in the middle spatio-temporal subband is sufficient to produce a transparent liquid impression even when there is a large difference in the deformation amplitude spectrum from the original liquid stimuli.

It should be noted that what we measured here was the spatiotemporal tuning of the perception of a transparent liquid, not that of the perception of a transparent layer. A transparent layer was visible in a wider range of stimulus conditions, including the cases where a high spatial frequency modulation suppressed the perception of liquid.

The Effect of Camera-to-Liquid Distance on Transparent Liquid Impression. To this point in our experiments, we had examined transparent liquid perception for artificial stimuli (i.e., amplitude spectrum-controlled random image modulations). Our results indicate that human observers perceive transparent liquids from dynamic image deformation not because the deformation is produced by real transparent liquid flows, but because the deformation contains diagnostic features in the spatiotemporal amplitude spectrum. Accordingly, human observers might not perceive transparent liquids from real transparent liquid flows when the deformation does not contain diagnostic features in the spatiotemporal amplitude spectrum.

To examine this issue, we conducted an additional rating experiment in which we manipulated the distance from the camera to a liquid flow surface in a computer-simulated scene (Fig. 4A). The image deformation generated by transparent liquid flows has a spatially low-pass profile under the camera-to-liquid distance that was used in previous experiments (Fig. 2A–C). When the same liquid flow is viewed from longer distances, the image deformation created by the flow should contain more energy in the high spatial frequency bands, which would effectively inhibit transparent liquid perception (Fig. 3B). Thus, the transparent liquid impression will be dependent on the camera-to-liquid distance if the observers use diagnostic features in the spatiotemporal amplitude spectrum for transparent liquid perception. On the other hand, regardless of the camera-to-liquid distance, the pattern of image deformation is always created by light refractions at the surface of flowing liquids. Thus, the observers will report a similar strength of transparent liquid impressions across variations in the camera-to-liquid distance if they see a

transparent liquid, simply because image deformations arise from transparent liquid flows.

The observers were asked to view a small patch in a liquid flow video clip (Movie S4) and rate their transparent liquid impression. The results indicate that the transparent liquid impression decreased significantly as the camera-to-liquid distance increased (Fig. 4B). Repeated-measures one-way ANOVA showed a significant main effect of the camera-to-liquid distance [$F(6, 54) = 7.523$; $P < 0.0001$]. The rating scores in the 10- and 12-m conditions were significantly lower than the scores in the 1-, 2-, and 4-m conditions ($P < 0.05$). The rating scores in the 12-m condition were also significantly lower than those in the 6-m condition ($P < 0.05$), and the rating scores in the 8-m condition were significantly lower than those in the 2-m condition ($P < 0.05$). These results indicate that the perception of transparent liquids does not stem from the physical correctness of dynamic image deformation as transparent liquid flows. Instead, the results are consistent with our suggestion that human observers use specific bands in the spatial (and temporal) frequencies of image deformation to see transparent liquid.

Discussion

Previous research on perceptual transparency has focused mainly on how the perception of a transparent layer is produced by luminance (and color) changes at the junctions and boundaries of overlapping surfaces (1, 3). It also has been shown that binocular disparities can affect the interpretation of a transparent layer (14, 15). Here we show that, without such conventional cues, human observers can still clearly see a transparent layer solely from dynamic image deformation. Furthermore, the dynamic image distortion simulating the effect of light refraction at the surface of moving water gives rise to the perception of a transparent liquid layer.

Static deformation is a very weak cue to perceptual transparency. Of course, a transparent object can be observed in a still image, but in most cases, the image also contains other cues to the transparent object, such as an object boundary and specular reflections (16–18). As shown in Fig. 5, a transparent layer is not clearly visible without specular reflections. In contrast, the dynamic deformation alone provides a vivid impression of a transparent layer.

Perceptual transparency from dynamic image deformation differs from standard motion transparency (19, 20). In motion transparency, image motion vectors are perceptually divided into two or more groups based on the difference between them and produce multiple moving layers that perceptually overlap each

Fig. 5. The role of light reflection, such as specular components, in interpreting image deformation. In A and B, an undulating transparent layer with a refractive index of 1.5 is placed in front of a pattern like stained glass. (A) Without specular components, it was difficult for the observers to see a transparent layer. (B) With specular components, the observers could perceive a transparent layer that deforms an image under it.

other. In contrast, in deformation-based transparency, the visual system does not segregate the background image motion (deformation) into multiple groups, but instead decomposes the background pattern and its motion as properties belonging to different layers. Presumably, to parsimoniously account for the complex movement in the background pattern, the visual system assumes the presence (and produces the representation) of a transparent layer that indirectly produces the movement.

Although it is only a speculation, the neural mechanisms underlying the deformation-based transparency perception are likely to include both the dorsal visual pathway responsible for motion processing (e.g., hMT+) and the ventral visual pathway responsible for object and material processing (e.g., V4 and fusiform gyrus) (19–23).

Our results indicate that human liquid perception is supported by a rather simple sensory computation that checks the presence/absence (i.e., amplitudes) of image distortion in specific spatio-temporal frequency ranges. We were able to make image deformation flows that looked like transparent liquids to human observers without taking into account physical correctness or other image factors, such as the phase spectrum, the amplitude spectrum of the whole range, divergence and curl, and higher-order statistics of image deformation. It should be noted, however, that our present data do not exclude the possibility that these image factors also contribute to different aspects of liquid perception.

The transparent liquid rating was dependent on image patterns under liquid flows. In general, the rating score was higher when natural images were used as background than when 2D 1/f noise patterns were used. Image features included in natural images but not in the noise patterns, such as edge continuity and element shape, are likely to facilitate deformation detection. How the visual system extracts image deformation remains an important research topic (10). A small difference in the spatio-temporal tuning between Fig. 3 *B* and *C* might be caused by this difference in stimuli.

Previous studies have shown that the human visual system is capable of recognizing the shape and material of nonrigid objects from image motion (24, 25). For instance, our recent study (26) found that pure visual motion flows extracted from scenes of running opaque liquids are sufficient for human observers to perceive liquids and their viscosity. In addition, the spatial smoothness of motion flow is a critical image property for liquid perception, in apparent agreement with the present finding that the liquid impression was enhanced by the deformation at a relatively low spatial frequency range. However, those previous studies examined the effects of image motion signals produced directly by the deformation of an object—i.e., motion signals produced by changes in the position of surface markers or changes in light reflection at the object surface. In contrast, in deformation-based transparency perception, image deformation is produced by light refraction at the undulating surface. The underlying optics, as well as the required computation, are totally different and likely more complex.

Scientific interest in material perception has been growing recently (27). In many cases, proper recovery of many material-related physical parameters solely from the available sensory information is computationally difficult. Thus, it has been suggested that the mechanism for material perception heuristically uses simple stimulus features correlated with the physical material properties under a range of natural environments (25, 28, 29). In agreement with this idea, our present findings indicate that the mechanism for material perception is sensitive to the presence and absence of specific ranges of spatiotemporal frequency components in the image deformation flow. Such a simple and effective computation must be beneficial for humans and animals to quickly recognize the critical material for survival—that is, water.

Methods

Observers. All observers who participated in the reported psychophysical experiments were naïve to the experiments' purpose and were paid for their participation. They reported normal or corrected-to-normal visual acuity. Ethical approval for this study was obtained from the NTT Communication Science Laboratories Ethical Committee. The experiments were conducted in accordance with the principles of the Helsinki Declaration. Written informed consent was obtained from all participants.

Apparatus. Stimuli were presented on a 21-inch CRT monitor (GDM-F500R; Sony) with a resolution of $1,024 \times 768$ pixels and a refresh rate of 60 Hz. The luminance emitted from the monitor was linearized in a range of 0–132 cd/m^2 using a photometer (OP200-E; Cambridge Research Systems). A computer (Mac Pro; Apple) controlled stimulus presentation and data collection with MATLAB and its extension, PsychToolBox 3 (30, 31).

Transparent liquid rating with computer graphics simulations.

Observers. Eight naïve persons participated in this experiment.

Stimuli. We simulated a transparent liquid flow using Blender (www.blender.org) and its physics engine as described previously.¹ The liquid flowing rightward or leftward was simulated at a velocity of 3 m/s in an imaginary box of $2 \times 2 \times 2$ m as a flow simulation domain (Fig. 1*A*). The liquid was shot by a stationary camera positioned 1.5 m above the averaged height of the liquid surface. The box boundary was not visible. The liquid inflow and outflow were set at the opposite sides of the box, and the amount of liquid in the box was kept constant by adjusting the inflow and outflow. The camera captured only the central part of the scene; thus, the captured scene included the liquid surface with stable flow, and did not include outflow and inflow boxes or the scene outside of the flow simulation domain. The depth of the liquid was ~ 10 cm. The kinematic viscosity of the liquid was set to $10^{-6} \text{ m}^2/\text{s}$, and the refraction index was set to 1.33. Both the kinematic viscosity and refraction index of the simulated liquid were identical to those of water. Specular components were removed from the scenes in rendering. The effect of caustics was not considered in the simulation.

Eight types of gray-scale natural images were used as under-liquid scenes (Fig. 1*B*). These images were texture-mapped on the plane beneath a liquid. The image files of the simulated scenes of a flowing transparent liquid were output and presented successively for 3 s at a frame rate of 30 Hz. Each rendered image subtended 11.5×11.5 degrees in visual angle and was presented with a spatially tapered cosine window with a diameter of 7.7 degrees for a constant section. The video clip was presented against a neutral gray background with a luminance of 66 cd/m^2 .

Procedure. The experiment was conducted in a dimly lit room. Each individual was tested separately. The observers sat 65 cm from the CRT display. To start each single trial, the observers were asked to press the spacebar on the keyboard of the computer. The task of the observers was to rate the strength of the transparent liquid impression on a five-point scale by pressing assigned keys (1, no impression of a transparent liquid, to 5, vivid impression of a transparent liquid). They were instructed to rate the impression after each video clip ended and disappeared. After they input their rating scores, the stimuli of the next trial were presented after 1 s. A video clip with one of eight under-liquid scenes was tested twice for each observer. The order of the trials was randomized across observers.

Real water analysis. To capture the image deformation of the pattern beneath real water flows, we used a glass container (20 cm high $\times$ 30 cm wide $\times$ 20 cm deep) to receive water. The water depth was 16 cm. A camera (EOS 5D mark II; Canon), which was set ~ 60 cm above the water's surface, captured images of underlying scenes that were deformed owing to refraction at the surface of water. The F-number of the camera was set at 2.8, and the ISO speed was set at 1,250. A light source (a white light), set ~ 80 cm above the water's surface, lit the water. To reduce the specular components of the water's surface, we attached polarized sheets to both the lens of the camera and the light source. A printed pattern (a 2D 1/f noise) was placed below the container. Because the container was transparent, it was possible to see the underlying pattern through it. Real water flows were generated using a plastic stick (a 5-mm-diameter straw) attached to a servomotor. Arduino Mega 2560 (Arruino), which was connected to Mac Pro, controlled the servomotor. The water was stirred by moving the plastic stick like a pendulum. The number of one-way movements of the pendulum per second

¹Thürey N, Ruede U, Free surface Lattice–Boltzmann fluid simulations with and without level sets. Proceedings of Vision, Modeling, and Visualization, November 16–18, 2004, Stanford, CA, pp 199–207.

Fig. 6. The method for removing caustics from the movie of natural liquid flow.

was manipulated at five levels: 1, 3, 5, 10, and 15 Hz, while keeping the angular travel distance of the one-way movement from vertical constant at 5 degrees.

When we first took the videos of the water, we noticed that water waves created patterns of light called "caustics" onto the pattern below the container. Because we wanted to see the pure effects of image deformation on liquid impression, we needed to remove the caustics. We did this using the following image processing technique. First, we linearized the relationship between the luminance values of the printed underlying pattern and the original RGB values of a computer-generated image for the underlying pattern in the computer. Second, we linearized the relationship between the luminance values of the printed underlying pattern and the RGB values of the image of the underlying pattern taken by the camera. Third, we placed the printed version of the 2D 1/f noise below the container. The original image for the 1/f noise was created by the intensity variation of the G channel between 64 and 192 in the computer. We added a uniform intensity (192) of the R channel to the G channel 1/f noise and printed this out (Fig. 6). Fourth, we subtracted the R channel intensity $[RI(x, y)]$ of an image $I(x, y)$ taken by the camera from the G channel intensity $[GI(x, y)]$ of the image with appropriate weights, ω (Fig. 6). Here the intensity variation of the R channel of the image included little image deformation of the underlying pattern, because the 1/f noise was created based on the intensity variation of the G channel. Thus, we successfully removed the intensity variation originating in caustics from the intensity variation of the G channel. Finally, we converted the image with the intensity of the G channel into a gray-scale image and created the video clips of image deformation owing to the real water flow. We originally obtained a $1,920 \times 1,080$ pixel video clip for each condition, but the left side of the video clip always contained specular components of the water's surface even though we tried to remove them with polarized filters. Thus, we trimmed the right side of the video clip by 512×512 pixels. We created a 3-s video clip for each of the five stirring-speed conditions. We transformed the size of the video

clip into 256×256 pixels (7.8×7.8 degrees) so as to have the same visual angle as the stimuli used in the reverse correlation experiment.

Experiment with phase and amplitude randomization.

Observers. Eight naive observers participated in this experiment. None of them had participated in the previous experiments.

Stimuli. We calculated the optical flow fields of the video clips of the real liquid flow as described previously (9). The spatial extent of the original video clips was 256×256 pixels (7.8×7.8 degrees), and the temporal extent was 73 frames (for 2.4 s). We created five shorter (32 frames/1.06 s) movie clips from the original movie clip. The start frames of the five video clips were the first, 11th, 21st, 31st, and 41st. Using a previously described method (10), we warped the pixels of an underlying pattern (a 2D 1/f noise) on the basis of calculated optical flow fields. We used the 2D 1/f noise as an underlying pattern rather than naturalistic scenes because we wanted to see whether earlier results obtained in this study could be extended to the scenes that only had general contrast statistics of natural images (e.g., ref 32). In an intact condition, an underlying pattern was deformed on the basis of intact optical flow fields.

To create video clips used in a phase-randomized condition, we conducted 3D FFT for the sequence of optical flow fields, and replaced the phase spectra of the original optical flow fields with the phase spectra of white noise with a spatiotemporal size ($256 \text{ pixels} \times 256 \text{ pixels} \times 32 \text{ frames}$) similar to that of the original optical flow fields. To create video clips used in an amplitude-randomized condition, we conducted the 3D FFT for the sequence of optical flow fields and replaced the amplitude spectra of the original optical flow fields with the phase spectra of white noise with a spatiotemporal size similar to the original optical flow fields. Thus, we tested 75 stimuli consisting of the five levels of stirring speed, three types of video clips (intact, phase-randomized, and amplitude-randomized), and five temporal positions of the video clips.

Procedure. Each observer viewed a stimulus video clip that was presented repeatedly throughout a trial with an intervideo-clip interval of 800 ms, and the observers rated their impression based on the extent to which they felt

the video clip contained transparent liquid flows on a five-point scale. As soon as the ratings were input, the video clip of the trial disappeared, and in 1 s a video clip for the next trial was presented. Each video clip was rated eight times.

Analysis. The rating scores were collapsed across the temporal position of the video clips, and thus each data point in Fig. 2F was drawn from 40 trials (five temporal positions $\times$ repetitions).

Reverse correlation experiment.

Observers. Eleven naive observers participated in this experiment. None of them had participated in the previous experiments.

Apparatus. The apparatus was identical to those used in the previous experiments.

Stimuli. We used a reverse correlation technique to check the tuning of a transparent liquid impression to the spatiotemporal frequencies of deformation. A schematic explanation for our manipulation of image deformation is shown in Fig. 3A. First, for vertical and horizontal deformations, we created two sets of a 3D Gaussian noise [256 pixels (for horizontal dimension) $\times$ 256 pixels (for vertical dimension) $\times$ 64 pixels (time dimension)]. Second, we used rectangular filters in the Fourier space to leave specific subbands of spatiotemporal frequency in the noises. From 42 spatiotemporal subbands [6 temporal subbands (with amplitudes at temporal frequencies of 1, 2, 3, 4–5, 6–10, and 11–15 Hz) $\times$ 7 spatial subbands (with amplitudes at spatial frequencies of 1, 2, 3–4, 5–8, 9–16, 17–32, and 33–64 cpi, corresponding to 0.13, 0.26, 0.38–0.51, 0.64–1.03, 1.15–2.06, 2.19–4.12, and 4.24–8.23 cpd, respectively)], we left 10% of the subbands on average. Which subbands were left was randomly determined in each trial. After inverse FFT, the amplitudes of the two filtered Gaussian noises were normalized in a range between -0.24 and 0.24 degrees of visual angle. The normalized vector map made from the two filtered noises was called a “spatiotemporal deformation cubic matrix.” The underwater scene was 2D 1/f noises. On the first frame, we presented an intact 2D 1/f noise that was randomly created in each trial. On the subsequent frames, we deformed the scene according to the temporal integral of the spatiotemporal deformation cubic matrix to make the deformation between the neighboring frames match the spatiotemporal deformation cubic matrix. The deformation was given to the 1/f noise by warping the pixels (10). The deformed image was presented at a frame rate of 30 Hz for 2.13 s (Movie S2). The observers were asked to report whether a transparent liquid was seen in the video clip in a yes/no manner (binary response, not multiscale rating). Each observer performed 600 trials.

Experiment using stimuli with narrow-band dynamic image deformation.

Observers. Ten naive observers participated in this experiment. None of them had participated in the previous experiments.

Apparatus. The apparatus was identical to that used in the previous experiments.

Stimuli. As in the reverse correlation experiment, we artificially deformed background images by using the amplitude of bandpassed white noises as the magnitude of image deformation. We created a 3-s sequence of white noise images and filtered out the sequence using a bandpass filter with a bandwidth of 1.5 octaves. The central spatial frequencies of the filter was 2, 4, 8, 16, 32, or 64 cpi (0.13, 0.26, 0.51, 1.03, 2.06, or 4.12 cpd), and the central temporal frequency was 0.33, 0.5, 1, 2, 5, 10, or 15 Hz. Thus, one of 42 subbands was tested independently, where the sequence of white noises was randomly generated in each trial. Here the size of background images was 512×512 pixels (15.6×15.6 degrees). Instead of 2D 1/f noises, we used one of eight natural images (Fig. 1B) as a background.

Procedure. The task of the observers was to view the repeatedly presented stimulus movies and rate transparent liquid impression on a five-point scale. Each background image was tested once, and thus each observer received a total of 420 trials. The order of trials was randomized across the observers.

Experiment investigating the effect of camera-to-liquid distance.

Observers. Ten naive observers participated in this experiment. None of them had participated in the previous experiments.

Apparatus. The apparatus was identical to that used in the previous experiments.

Stimuli. The method used to create stimulus movies containing the computer-rendered scenes of liquid flows was identical to that used in the first set of experiments except that instead of natural scene images, we used a 2D 1/f noise as background. In addition, when rendering the scenes of liquid flow, we systematically altered the camera-to-liquid distance at seven levels (1, 2, 4, 6, 8, 10, and 12 m).

Procedure. The task of the observers was to view the stimulus movies that were repeatedly presented and then rate the strength of the transparent liquid impression on a five-point scale by pressing assigned keys. Each level of camera-to-liquid distance was tested 20 times. Thus, each observer received a total of 140 trials. The order of the trials was randomized across the observers.

ACKNOWLEDGMENTS. This research was supported by Grants-in-Aid for Scientific Research on Innovative Areas (22135004 and 15H05915) from the Japanese Ministry of Education, Culture, Sports, Science, and Technology.

- Metelli F (1974) The perception of transparency. *Sci Am* 230(4):90–98.
- Singh M, Anderson BL (2002) Toward a perceptual theory of transparency. *Psychol Rev* 109(3):492–519.
- Anderson BL (1997) A theory of illusory lightness and transparency in monocular and binocular images: The role of contour junctions. *Perception* 26(4):419–453.
- D’Zmura M, Colantoni P, Knoblauch K, Laget B (1997) Color transparency. *Perception* 26(4):471–492.
- Sayim B, Cavanagh P (2011) The art of transparency. *i-Perception* 2(7):679–696.
- Fleming RW, Jäkel F, Maloney LT (2011) Visual perception of thick transparent materials. *Psychol Sci* 22(6):812–820.
- Baker CL, Jr, Braddick OJ (1985) Temporal properties of the short-range process in apparent motion. *Perception* 14(2):181–192.
- Braddick O (1974) A short-range process in apparent motion. *Vision Res* 14(7):519–527.
- Sun D, Roth S, Black MJ (2010) Secrets of optical flow estimation and their principles. *Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition* (Institute of Electrical and Electronics Engineers, New York), 2432–2439.
- Bex PJ (2010) (In) Sensitivity to spatial distortion in natural scenes. *J Vis* 10(2):23.1–23.15. Available at jov.arvojournals.org/article.aspx?articleid=2121192. Accessed July 20, 2015.
- Gosselin F, Schyns PG (2001) Bubbles: A technique to reveal the use of information in recognition tasks. *Vision Res* 41(17):2261–2271.
- Thurman SM, Grossman ED (2011) Diagnostic spatial frequencies and human efficiency for discriminating actions. *Atten Percept Psychophys* 73(2):572–580.
- Nishida S (2004) Motion-based analysis of spatial patterns by the human visual system. *Curr Biol* 14(10):830–839.
- Anderson BL, Julesz B (1995) A theoretical analysis of illusory contour formation in stereopsis. *Psychol Rev* 102:705–743.
- Anderson BL (2007) Filling-in models of completion: Rejoinder to Kellman, Garrigan, Shipley, and Keane (2007) and Albert (2007). *Psychol Rev* 114(2):509–527.
- Blake A, Bülthoff H (1990) Does the brain know the physics of specular reflection? *Nature* 343(6254):165–168.
- Blake A, Bülthoff H (1991) Shape from specularities: Computation and psychophysics. *Philos Trans R Soc Lond B Biol Sci* 331(1260):237–252.
- Fleming RW, Torralba A, Adelson EH (2004) Specular reflections and the perception of shape. *J Vis* 4(9):798–820.
- van Doorn AJ, Koenderink JJ (1982) Spatial properties of the visual detectability of moving spatial white noise. *Exp Brain Res* 45(1–2):189–195.
- Snowden RJ, Verstraten FA (1999) Motion transparency: Making models of motion perception transparent. *Trends Cogn Sci* 3(10):369–377.
- Orban GA (2008) Higher-order visual processing in macaque extrastriate cortex. *Physiol Rev* 88(1):59–89.
- Orban GA, Zhu Q, Vanduffel W (2014) The transition in the ventral stream from feature to real-world entity representations. *Front Psychol* 5:695.
- Hiramatsu C, Goda N, Komatsu H (2011) Transformation from image-based to perceptual representation of materials along the human ventral visual pathway. *Neuroimage* 57(2):482–494.
- Jain A, Zaidi Q (2011) Discerning nonrigid 3D shapes from motion cues. *Proc Natl Acad Sci USA* 108(4):1663–1668.
- Norman JF, Wiesemann EY, Norman HF, Taylor MJ, Craft WD (2007) The visual discrimination of bending. *Perception* 36(7):980–989.
- Kawabe T, Maruya K, Fleming R, Nishida S (2015) Seeing liquids from visual motion. *Vision Res* 109:125–138.
- Fleming RW (2014) Visual perception of materials and their properties. *Vision Res* 94:62–75.
- Motoyoshi I, Nishida S, Sharan L, Adelson EH (2007) Image statistics and the perception of surface qualities. *Nature* 447(7141):206–209.
- Marlow PJ, Kim J, Anderson BL (2012) The perception and misperception of specular surface reflectance. *Curr Biol* 22(20):1909–1913.
- Brainard DH (1997) The psychophysics toolbox. *Spat Vis* 10(4):433–436.
- Pelli DG (1997) The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spat Vis* 10(4):437–442.
- Field DJ (1987) Relations between the statistics of natural images and the response properties of cortical cells. *J Opt Soc Am A* 4(12):2379–2394.