Available online at www.sciencedirect.com
SCIENCE DIRECT®

Vision Research 44 (2004) 461–470

Vision Research

www.elsevier.com/locate/visres

Troubles with bubbles

Richard F. Murray [a,*], Jason M. Gold [b]

[a] *Center for Perceptual Systems, University of Texas at Austin, 1 University Station A8000, Austin, TX 78712-0187, USA*
[b] *Department of Psychology, Indiana University, 1101 East 10th Street, Bloomington, IN 47405, USA*

Received 18 December 2002

Abstract

The bubbles method is a recently developed variant of reverse correlation methods that have been used in psychophysics and physiology. We show mathematically that for the broad and important class of noisy linear observers, the bubbles method recovers much less information about how observers process stimuli than reverse correlation does. We also show experimentally that the unusual type of noise used in the bubbles method can drastically change human observers' strategies in psychophysical tasks, which reduces the value of the information that is obtained from a bubbles experiment. We conclude that reverse correlation is generally preferable to the bubbles method in its present form, but we also give suggestions as to how the bubbles method could be modified to avoid the problems we discuss.

1. Introduction

We can learn a great deal about a signal processing system, be it a single neuron or a human observer, by observing how the system responds to stimuli in noise. In a typical application of *reverse correlation* methods, the input to the system under study is one of two signals in additive Gaussian white noise, the response of the system is an attempt to identify the signal, and the result of the experiment is a *classification image*, which shows the correlation between the noise contrast at each stimulus location and the system's responses. In effect, a classification image shows how each spatial location of the stimulus contributes to the system's attempts to identify the signal. Several types of reverse correlation methods have been developed for use in psychophysics and physiology (Ahumada & Lovell, 1971; Marmarelis & Marmarelis, 1978).

The bubbles method is a recently developed variant of reverse correlation, which differs from previous methods in two ways (Gosselin & Schyns, 2001). First, in a typical bubbles experiment, the input to the system is one of two signals windowed through a noise field that consists of a number of randomly placed Gaussian blobs. That is, the input is just a few small fragments of a signal, rather than the whole signal. Second, the result of the experiment is a *bubbles image*, which is proportional to the expected value of the windowing noise on trials where the system gives the correct response. Thus, a bubbles image shows which stimulus locations help the system to identify the signal correctly.

The bubbles method is an interesting variation on reverse correlation that addresses the question of what stimulus regions help a system to give the correct response, as opposed to the question of what regions influence the system's responses in any way at all. However, in its present form it has two serious shortcomings. First, the bubbles method has never been described or analyzed in any sort of rigorous signal processing framework, and consequently many important properties of bubbles images are simply not known. How does a bubbles image depend on response bias, or on the proportion of correct responses? What are the statistics of bubbles images? Indeed, how is a bubbles image even related to the parameters of any class of models that we might wish to use to study a system? At present, we do not know the answers to these questions, so it is difficult to use the bubbles method to test hypotheses quantitatively. In Section 2, we make a first step towards remedying these problems, by showing what information a bubbles image recovers about a broad and important class of systems, namely noisy linear amplifiers. A second problem with the bubbles

[*] Corresponding author. Address: Department of Psychology, University of Pennsylvania, 3815 Walnut Street, Philadelphia, PA 19104-6196, USA. Tel.: +1-215-898-7300; fax: +1-215-746-6848.
E-mail address: murray@psych.upenn.edu (R.F. Murray).

0042-6989/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2003.10.006

method is that showing small, isolated fragments of a stimulus will often change a system's behaviour compared to when the stimulus is shown intact, and this greatly reduces the value of the information learned from a bubbles experiment. In Section 3, we report an example of a psychophysical task where the bubbles method drastically changes human observers' strategies. We suggest that using a different kind of windowing noise may make the bubbles method less likely to change observers' strategies.

2. Troubles in theory

The *linear amplifier model* (LAM) embodies a simple theory as to how observers perform shape discrimination tasks. The LAM is a useful first-order approximation that accounts for many aspects of human performance, and serves as a starting point for more complex models (Burgess, Wagner, Jennings, & Barlow, 1981; Green & Swets, 1974). In this section, we will compare the information that reverse correlation and bubbles methods recover about LAM observers.

Consider a task where the observer views one of two signals, I_X or I_Y, and reports which signal was shown. According to the LAM, an observer identifies the signal by cross-correlating the stimulus with a template T, adding an internal noise Z, and responding 'X' when the resulting decision variable s meets or exceeds a criterion a, and responding 'Y' otherwise. If we represent the observer's responses as a random variable R that takes value +1 when the observer responds 'X' and −1 when the observer responds 'Y', then we can describe a LAM observer with the following equations:

$$s = I_{\{X,Y\}} \otimes T + Z \tag{1}$$

$$R = \operatorname{sgn}(s-a) = \begin{cases} +1 & \text{if } s \geqslant a \\ -1 & \text{if } s < a \end{cases}$$

Here $\otimes$ is cross-correlation (i.e., for matrices F and G, $F \otimes G = \sum_{ij} F_{ij}G_{ij}$). Cross-correlation is a linear operation, so the LAM states that the observer's responses are based on a linear function of the stimulus, contaminated by noise. [1]

In a typical reverse correlation experiment, signals are shown in Gaussian white noise, so the stimuli are $I_X + N$ and $I_Y + N$, where N is a noise field. With two signals and two responses, there are four stimulus–response classes of trials: XX, XY, YX, and YY. The classification image C is defined as:

$$C = (\overline{N}_{XX} + \overline{N}_{YX}) - (\overline{N}_{XY} + \overline{N}_{YY}) \tag{2}$$

Here $\overline{N}_{\mathrm{SR}}$ denotes the average of the noise fields over a stimulus–response class of trials, e.g., $\overline{N}_{XY}$ is the average noise field over all trials where the signal was I_X and the observer responded 'Y'. For a LAM observer, the expected value of a classification image can be shown to be proportional to the observer's template T:

$$E[C] = kT \tag{3}$$

Thus a classification image completely characterizes how a LAM observer combines information from different spatial locations of a stimulus to decide on a response (Ahumada, 1996; Murray, Bennett, & Sekuler, 2002; Richards & Zhu, 1994). That is, a classification image tells us everything there is to know about a LAM observer, apart from the power of the internal noise.

In a bubbles experiment, signals are multiplied pointwise by a windowing noise that consists of a number of randomly placed Gaussian blobs (bubbles), so the stimuli are $I_X \circ W$ and $I_Y \circ W$, where W is the windowing noise and $\circ$ is pointwise multiplication (i.e., $(F \circ G)_{ij} = F_{ij}G_{ij}$). The bubbles image is defined as the sum of the windowing noise W over all trials where the observer gives the correct response (i.e., trial types XX and YY), divided by the sum of W over all trials:

$$B = \frac{\sum_{XX,YY} W}{\sum_{XX,XY,YX,YY} W} \tag{4}$$

Here the division is pointwise (i.e., $(F/G)_{ij} = F_{ij}/G_{ij}$). In Appendix A we show that for an unbiased LAM observer, the expected value of a bubbles image recovers the observer's template T, multiplied pointwise by the difference image of the two signals, $I_X - I_Y$, blurred twice by the bubble b that is used to create the windowing noise:

$$E[B] = u + v \cdot b * b * (T \circ (I_X - I_Y)) \tag{5}$$

Here u and v are constants that are determined by such factors as the observer's proportion correct and internal noise power, * is two-dimensional convolution, and $\circ$ is pointwise multiplication. The constants u and v are of secondary interest, and the key result is that the bubbles image essentially recovers $b * b * (T \circ (I_X - I_Y))$.

Eq. (5) confirms a number of properties that we would intuitively expect of a bubbles image. First, the equation shows that a bubbles image has larger values at locations where the pointwise product of the observer's template and the difference image $I_X - I_Y$ are positive than at locations where the pointwise product is negative. This is sensible, because the difference image $I_X - I_Y$ is the ideal template for the task of discriminating between I_X and I_Y in Gaussian white noise (Green & Swets, 1974). A bubbles image is greater at stimulus locations that help the observer to give the correct response, and for a LAM observer these are the locations where the

[1] In some formulations of the LAM, the internal noise is added before the cross-correlation. This modification makes no difference to the conclusions in this paper, and adding the noise after the cross-correlation simplifies the derivation in Appendix A.

observer's template is similar to the ideal template, e.g., has a pointwise product that is positive, not negative. Second, Eq. (5) shows that the expected value of the bubbles image involves a double-convolution with the bubble b used to generate the windowing noise, from which it follows that the size of the bubble determines the level of detail that can be resolved in the bubbles image. This also makes sense intuitively, although without a careful analysis one might not realize that the level of detail is determined by a *double*-convolution with the bubble. [2]

Most importantly, Eq. (5) shows that a bubbles image does not completely recover an observer's template, but only the parts that correspond to nonzero locations in the ideal template. On the other hand, a classification image does completely recover the template, and furthermore the windowing bubble and the ideal template are known exactly, so from a classification image we can calculate the bubbles image corresponding to any given bubble, using Eq. (5). That is, a reverse correlation experiment recovers all the information about a LAM observer that a bubbles experiment does, and more.

Gosselin and Schyns (2002) discuss the fact that reverse correlation and bubbles methods recover different information about an observer, and argue that the two methods are complementary. They give the name *represented information* (R) to template features recovered by a classification image, *potent information* (P) to features recovered by a bubbles image, and *available information* (A) to stimulus features that objectively contain information as to the correct response (i.e., features that are used by the ideal observer). They suggest that potent information is the intersection of represented information and available information, $R \otimes A \approx P$. (Here the symbol $\otimes$ does not mean cross-correlation, but some type of intersection operation that is not clearly defined.) This conceptual relationship is made precise by our finding that, apart from double-blurring by the bubble, the bubbles image recovers the pointwise product of the observer's template and the ideal template, $B \sim T \circ (I_X - I_Y)$.

At this point, the advantages of understanding these methods in terms of a rigorous signal processing framework become clear. On the one hand, if the LAM is a valid model of the system under study, then it is always preferable to carry out a reverse correlation experiment rather than a bubbles experiment, because Eq. (5) shows that from a classification image we can easily determine the result of any bubbles experiment. On the other hand, if the LAM is not a valid model, then there is no reason to expect the simple relationship $R \otimes A \approx P$ (or more formally, $B \sim T \circ (I_X - I_Y)$) to hold between a bubbles image, the observer's template, and the ideal template. For instance, in a two-alternative discrimination task where landmarks that appear in the same location in both signals actually help the observer to perform the task, the landmarks will function as 'potent' information, and hence appear in the bubbles image, even though there is no 'available' information at the landmarks, because they do not appear in the ideal template. This might happen, say, in a vernier alignment task where one of the vernier lines appears at the same location in all stimuli, and so by itself provides no information as to the correct response, but helps the observer to judge the location of the other line, whose location varies from trial to trial (Beard & Ahumada, 1998). Thus in cases where the LAM is correct, the bubbles method is superfluous, and in cases where the LAM is incorrect, intuition is a poor guide as to what the bubbles method actually measures, as demonstrated by the probable failure of Gosselin and Schyns' $R \otimes A \approx P$ law in a vernier alignment task.

Our derivation showing what information a bubbles image recovers about a LAM observer is just a first step in understanding what the bubbles method reveals about human observers. Human observers do not always fit the LAM model, although this model does give a good first-order description of many aspects of performance. However, understanding what information a method recovers about a simple and well-defined class of observers is not only useful for understanding the method in relation to observers that fit the model, but also for interpreting departures from the model (e.g., Ahumada & Beard, 1999). Furthermore, the question of how to use reverse correlation to investigate nonlinear systems has been studied extensively (Nabet & Pinter, 1992; Wiener, 1958), and it should also be possible to determine what information the bubbles method recovers about more complex classes of observers. In any case, it is certainly better to investigate how a novel method is related even to simple and tractable models, rather than to forego rigorous analysis altogether, and to rely on intuition alone to guide our use of the method.

3. Troubles in practice

A second and more serious problem with the bubbles method is that showing only small fragments of a stimulus will often change an observer's behaviour compared to when the stimulus is shown intact. Previous studies have documented such effects, e.g., Schwartz, Bayer, and Pelli (1998) found that observers used different stimulus regions to identify faces, depending on

[2] It has been noted that the bubbles method seems to require fewer trials than reverse correlation, but this may simply be due to the very low spatial resolution of bubbles images that results from this double-blurring effect. With lowpass-filtered noise, reverse correlation may require just as few trials.

which regions were covered by Gaussian white noise. [3] Thus a bubbles image, which is calculated from responses to small fragments of a stimulus, may not only provide an incomplete characterization of a system's behaviour, but a misleading one.

Gosselin and Schyns (2001) acknowledged this possibility, and addressed it in a control experiment. In their main experiment, they used the bubbles method to determine what stimulus regions helped observers to correctly identify faces. In their control experiment, they compared face identification performance in three conditions: the ORIGINAL condition, where the stimuli were whole faces; the DIAGNOSTIC condition, where the stimuli were faces windowed to show only the parts that the main experiment had found to be helpful to observers; and the NONDIAGNOSTIC condition, where the stimuli were faces windowed to show only the parts that the main experiment had found to be unhelpful. The results were that over a range of stimulus durations, proportion correct was approximately the same in the DIAGNOSTIC and ORIGINAL conditions, and much lower in the NONDIAGNOSTIC condition. Gosselin and Schyns concluded that observers that the DIAGNOSTIC stimulus showed precisely the stimulus regions that observers used to identify the intact faces in the ORIGINAL condition.

Unfortunately, the results of this experiment are inconclusive. One technical problem is that the stimuli in the three conditions were normalized to have the same total contrast energy. As a result, in the DIAGNOSTIC condition all contrast energy was concentrated in helpful image regions, whereas in the ORIGINAL condition it was distributed over both helpful and unhelpful regions. Gosselin and Schyns pointed out that this was probably why performance was actually slightly but consistently *better* in the DIAGNOSTIC condition than in the ORIGINAL condition. This suggests that if the stimuli had been designed so that the helpful image regions had the same local contrast in the DIAGNOSTIC and ORIGINAL conditions, then performance would have been worse in the DIAGNOSTIC condition, and perhaps much worse. This would certainly undermine the claim that only the image regions shown in the DIAGNOSTIC condition helped observers to identify intact faces.

A second, more crucial problem is that even if performance was similar in the DIAGNOSTIC and ORIGINAL conditions with appropriately matched contrasts, this would still be weak evidence that the DIAGNOSTIC stimulus showed precisely the stimulus regions that observers normally use to identify intact faces. One can imagine several plausible alternative explanations. For example, in a stimulus with several redundant, informative features, it may be that observers normally use only one or two of these features, whereas the bubbles method forces observers to use different features on different trials, because only small fragments of the stimulus are shown on any given trial. If so, the DIAGNOSTIC stimulus would show many such features, and give a misleading impression of the observer's strategy, but proportion correct would quite plausibly be the same in the DIAGNOSTIC and ORIGINAL conditions, because both would show several of the redundant image features that observers could use. Another alternative explanation is that observers might use broader image regions in the ORIGINAL condition, but integrate across space less efficiently. Or, observers might benefit in some ways from the additional information available in the ORIGINAL stimuli, but suffer in other ways from masking and lateral interactions between neighbouring image regions in the ORIGINAL stimuli. These scenarios are all consistent with similar performance in the DIAGNOSTIC and ORIGINAL conditions, and in fact, this is the very reason why reverse correlation and related methods are so appealing: it is often difficult to make very general conclusions about how observers perform a task, from just a few measurements of proportion correct.

To demonstrate that the bubbles method *can* drastically change observers' strategies in some tasks, we measured bubbles images in a task that has recently been studied with reverse correlation (Gold, Murray, Bennett, & Sekuler, 2000). The stimuli were Kanizsa-square like patterns (Fig. 1, first two rows). Two conditions, the *illusory* condition and the *fragmented* condition, were run in separate blocks. In the illusory condition, the Kanizsa inducers (i.e., the clipped circles) faced inwards so as to produce illusory contours, and in the fragmented condition, they all faced downwards and to the right. In both conditions, the inducers were rotated slightly from horizontal–vertical, to produce 'fat' and 'thin' patterns, and on each trial the observer judged whether the pattern was fat or thin. The ideal templates for these fat–thin discrimination tasks (Fig. 1, row 3) are the difference images between the fat and thin stimuli (Green & Swets, 1974), and they show that the informative stimulus regions lie along the straight edges of the Kanizsa inducers. Gold et al.'s classification images (Fig. 1, row 4) showed that to discriminate between fat and thin illusory Kanizsa squares, observers used one or two whole vertical sides of the square, including the

[3] Gosselin and Schyns (2001) have also suggested a variant of the bubbles method in which only narrow spatial frequency bands of a stimulus are presented on any given trial, rather than small spatial regions. Thus it is worth noting observers' strategies can also be changed by presenting only small ranges of spatial frequencies. In letter identification experiments, Gold, Bennett, and Sekuler (1999) found that when only a narrow band of spatial frequencies were presented on any given trial, observers could use whichever frequency range was presented to identify the letter, whereas Solomon and Pelli (1994) found that when intact letters were presented, observers used only a narrow band of the broad range of available frequencies.

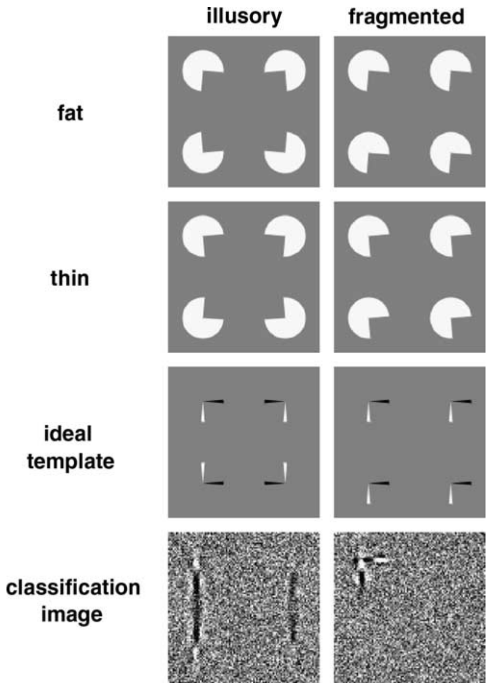

Fig. 1. Rows 1 and 2: stimuli from the fat–thin Kanizsa square discrimination task. Row 3: ideal templates for discriminating between fat and thin stimuli. Row 4: average classification images for the fat–thin task, from Gold et al. (2000).

straight edges of the Kanizsa inducers *and* the illusory contours connecting them. To discriminate between fat and thin fragmented Kanizsa squares, observers used the edges of only a single inducer. We suspected that using the bubbles method to study observers' strategies in this task would lead to very different results. If only small pieces of the stimulus are shown at a time, observers will not perceive strong illusory contours, and this may weaken their tendency to use whole sides of the square in the illusory condition. Furthermore, when only small pieces of the stimulus are shown, observers may use whichever piece appears on a given trial, and in the fragmented condition several inducers may appear in the bubbles image, rather than just one.

3.1. Method

3.1.1. Participants

Three undergraduate students at the University of Texas at Austin participated for payment. All had normal or corrected-to-normal Snellen acuity, and none were aware of the purpose of the experiment.

3.1.2. Stimuli

The stimuli were fat and thin, illusory and fragmented Kanizsa squares (Fig. 1). The Kanizsa inducer radius was 0.50° of visual angle (deg), and the inducers were spaced 2.0° apart, vertex-to-vertex. The inducers were rotated ±10° from horizontal–vertical. Peak Weber contrast, before windowing through the bubbles, was 30%. The stimuli were windowed through a number of randomly placed Gaussian blobs with a peak value of 1.0 and a standard deviation of 0.1°. (That is, the contrast at each location was multiplied pointwise by a field of unit-amplitude Gaussian blobs.) Stimuli were shown on a grey background of luminance 40 cd/m^2, on a Trinitron Multiscan E540 monitor (pixel size 0.478 mm, resolution 800 × 600 pixels, refresh rate 120 Hz). Observers viewed the stimuli binocularly from a distance of 1 m.

3.1.3. Procedure

Observers participated in two or three one-hour sessions. Each session had 1200–1500 trials, divided into 300-trial blocks that showed either illusory or fragmented stimuli. Each trial began with a 400 ms fixation interval, followed by the 200 ms stimulus, followed by a response interval in which the observer pressed one of two keys to indicate whether the stimulus was fat or thin. Auditory feedback indicated whether the response was correct. The number of bubbles varied across trials according to a one-up, two-down staircase in order to maintain approximately 71% correct performance, and the mean ± standard deviation was 30 ± 12 bubbles.

For the sake of completeness, we will point out some differences between the stimuli in this experiment and in Gold et al.'s experiment. In this experiment the inducers were white, had a radius of 0.50°, and had an inducer angle of ±10° from horizontal–vertical, whereas in Gold et al.'s experiment the inducers were black, had a radius of 0.35°, and had an inducer angle of ±1.75°. We have replicated Gold et al.'s results with stimuli in which the inducers were white, had a radius of 0.50°, and had inducer angles of up to ±8°, so we do not believe that these are crucial differences between the two experiments (Murray, 2002). In this experiment we maintained threshold performance by varying the number of bubbles from trial to trial according to a staircase, so task difficulty varied slightly from trial to trial. In Gold et al.'s experiment, threshold performance was maintained by varying the signal contrast from trial to trial, and the QUEST procedure that was used to set the contrast quickly converged to the observer's 75% threshold, so task difficulty typically did not vary much from trial to trial (Watson & Pelli, 1983). Again, we do not believe that this is a crucial difference between the two experiments.

3.2. Results and discussion

The first three rows of Fig. 2 show individual observers' bubbles images. Note that in the fragmented condition, all the bubbles images peaked at the locations of two or three Kanizsa inducers, indicating that all observers used two or three inducers to perform the task. In contrast, Gold et al. (2000) found that all three

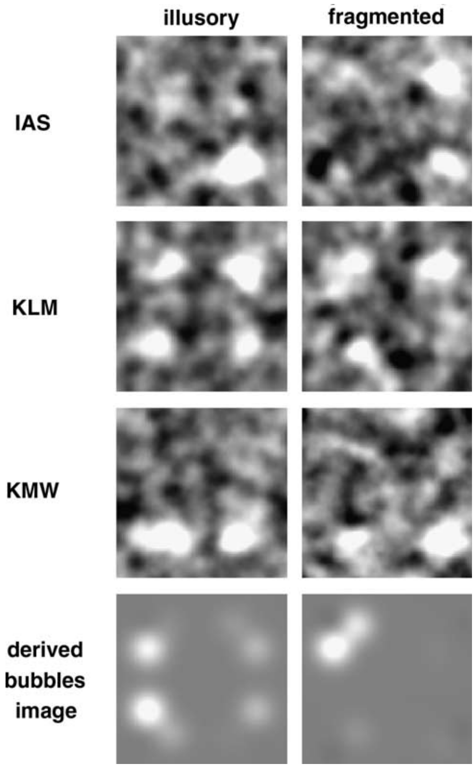

Fig. 2. Rows 1, 2, and 3: bubbles images for individual observers. Row 4: Hypothetical bubbles images obtained by applying Eq. (5) to the Gold et al.'s classification images (Fig. 1, row 4).

of their observers used only a single inducer in the fragmented condition, and Murray (2002) confirmed this result with five more observers. Thus the bubbles method seems to have led observers to use a very different strategy than they would use with the intact stimulus: when the inducers were shown in small pieces, observers used whichever piece happened to appear on any given trial. The resulting bubbles images give the misleading impression that observers normally use several inducers to perform the fat–thin task in the fragmented condition.

The bubbles images from the illusory condition give further evidence that the bubbles method can change observers' strategies: observer IAS used only a single inducer in the illusory condition, whereas all observers in the reverse correlation experiments used one or two whole sides of the illusory square, i.e., one or two pairs of aligned inducer edges. When only a few small pieces of the stimulus are presented at a time, observers do not perceive strong illusory contours or a coherent perceptual organization, so it is understandable that their strategies might differ from when the whole stimulus is shown.

The bubbles images from the illusory condition also illustrate the fact that the bubbles method does not completely recover observers' templates. The most interesting aspect of Gold et al.'s classification images in the illusory condition was that they showed that observers actually used the empty regions between the inducers, along the illusory contours, to perform the fat–thin task. Bubbles images from the illusory condition are necessarily empty between the inducers, because zero-contrast stimulus locations windowed through bubbles can obviously neither help nor hinder observers in giving a correct response. From the bubbles images, we would never guess that observers used illusory contours to judge the shapes of Kanizsa squares.

As another way of comparing our bubbles images to Gold et al.'s classification images, we took the classification images in Fig. 1 as estimates of Gold et al.'s observers' templates, and we used Eq. (5) to calculate the corresponding bubbles images. Specifically, we calculated $b * b * (T \circ (I_X - I_Y))$, where b was the Gaussian bubble we used in our experiments, T was a classification image from the fourth row of Fig. 1, and $I_X - I_Y$ was the corresponding ideal template from the third row of Fig. 1. The results (Fig. 2, bottom row) are the bubbles images that one would expect from Gold et al.'s observers in a bubbles experiment, assuming that the bubbles method would not disrupt their strategies, and assuming the LAM as a framework for translating classification images into bubbles images. These hypothetical bubbles images show that the informative parts of the stimulus that observers used were the vertical edges in the illusory condition, and both edges of the top left inducer in the fragmented condition. Again, this strategy is very different from the strategies revealed by the bubbles images that we measured in the present experiments, indicating that the practice of showing small pieces of stimuli in the bubbles experiment disrupted observers' usual strategies.

Could it be that the Gaussian noise in reverse correlation experiments disrupts observers' strategies, rather than the windowing noise in bubbles experiments, so that bubbles images actually reflect observers' normal strategies more accurately? We think this unlikely, for three reasons. First, it is intuitively clear why windowing stimuli through bubbles might change observers' strategies: when only small parts of a stimulus are shown on any given trial, observers may be forced to use stimulus features that they would not use if the whole stimulus was presented. Second, a great deal of psychophysical and physiological evidence shows that even under noiseless viewing conditions, observers' performance in threshold tasks is limited by internal noise, so by adding external noise we are probably not presenting observers with a task that is qualitatively different from a noiseless threshold task (Green & Swets, 1974). Third, and most convincingly, observers' contrast energy thresholds have been found to be an approximately linear function of external noise power in practically every task in which this relationship has been tested, including discrimination of fat vs. thin Kanizsa squares, and this is strong evidence that observers use the same strategy at all levels of external noise, from negligible levels to high levels of noise (Murray, 2002; Pelli, 1990).

The result of this experiment should not be a surprise. In tasks where observers are unable to use several redundant features simultaneously, we should expect that they will be able to use the features one at time when they are shown in isolation, and this simple fact implies that the bubbles method will often give a misleading impression of observers' strategies. In fact, it is easy to contrive tasks where the bubbles method would change observers' strategies even more drastically, e.g., an identification task in which the stimuli consist of several redundant letters scattered at locations where they are difficult to identify simultaneously. We chose the fat–thin task for this experiment in order to show how the bubbles method would affect strategies in a task that had actually been discussed in the literature, rather than a task that was designed to maximize the disruptive effect of showing only small fragments of a signal.

There may be a simple remedy for this problem. As it has been used up to now, the bubbles method windows signals through a small number of Gaussian blobs. Another way of saying this is that signals are shown in multiplicative, Gaussian-blurred, sparse binary noise. The derivation in Appendix A makes it clear that the essential feature of the bubbles method that distinguishes it from reverse correlation is not this unusual type of noise, but the fact that the noise is multiplicative rather than additive. In fact, Eq. (5) is valid for many types of windowing noise. Blurred sparse binary noise obliterates all but a few small regions of the stimulus on any given trial, and this can easily change observers' strategies. If instead we showed signals windowed through multiplicative Gaussian noise with a mean of 1.0 and a small standard deviation (e.g., 0.1), then the entire stimulus would be visible on any given trial, and the contrast of individual pixels would be slightly increased or decreased by the multiplicative noise. The effect on the stimulus would be similar to windowing through Gaussian bubbles, but more subtle. We suspect that this type of noise is much less likely to change observers' strategies. Furthermore, as we discuss in Appendix A, a bubbles image measured with multiplicative Gaussian white noise would recover exactly the same information about a LAM observer as the usual bubbles method. We are currently testing this variant of the bubbles method, to see whether these theoretical predictions are borne out in practice.

4. Conclusions

The two shortcomings of the bubbles method that we have discussed can probably be fixed. The first problem is that until now, almost nothing was known about exactly what information a bubbles image actually recovers about any well-defined class of observers. Obviously, the solution to this problem is simply to study the bubbles method more rigorously, to determine what information it recovers about various types of observers. Our results in this direction show that in tasks where the LAM is an adequate model, the bubbles method is entirely superfluous, so if the method is to be at all useful, it will be in studying tasks where observers' responses are based on nonlinear functions of the stimulus. At present, nothing is known about what information the bubbles method recovers about such observers.

The second problem is that the windowing noise used in the bubbles method seems certain to change observers' strategies in many tasks, as we demonstrated in the fat–thin Kanizsa square discrimination experiment. Fortunately, it may not be necessary to use this unusual type of noise. By using a less disruptive type of noise, such as multiplicative unit-mean Gaussian white noise, we may be able to make the bubbles method less likely to drastically change observers' strategies, while recovering exactly the same information about the observer.

If these developments are successful, the bubbles method may become a useful addition to reverse correlation methods. In principle, the two methods should be complementary, as reverse correlation shows how different stimulus locations contribute to an observer's responses, and the bubbles method shows which locations help the observer to give a correct response. However, in the case of LAM observers, the results of a reverse correlation experiment allow one to fully predict the results of a bubbles experiment, but not vice versa, which suggests that reverse correlation experiments may generally be more informative. Of course, if one is interested only in what stimulus locations help an observer give a correct response, then the bubbles method is perfectly adequate (with the caveat that in its current form, it may drastically change observers' strategies). Normally, though, in psychophysical and physiological experiments we wish to characterize the system under study as completely as possible, and for this purpose, reverse correlation is more appropriate. We conclude that, until further developments resolve these problems, reverse correlation is generally preferable to the bubbles method, as it is better understood theoretically, it recovers much more information about observers than the bubbles method does, and it is less likely to disrupt observers' strategies.

Appendix A. What does a bubbles image measure?

A.1. White binary noise

A white binary noise field is a stochastic image in which each pixel is an independent Bernoulli random variable that takes value 1 with probability p_N and value 0 with probability $1-p_N$. Consider a LAM observer who classifies white binary noise fields N by cross-correlating with a template T, adding an internal noise Z, and responding '*X*' or '*Y*' depending on whether the resulting decision variable s exceeds a criterion a. If we represent the observer's responses with a random variable R that takes values ±1, we can describe the observer with the following equations:

$$s = N \otimes T + Z \tag{A.1}$$

$$R = \operatorname{sgn}(s-a) = \begin{cases} +1 & \text{if } s \geqslant a \\ -1 & \text{if } s < a \end{cases} \tag{A.2}$$

(Here $\otimes$ is cross-correlation.) On trials where the observer responds '*X*', the expected value of a single noise pixel N_i is

$$E[N_i|R=+1] = E[N_i|s \geqslant a] \tag{A.3}$$

$$= 1 \cdot P(N_i = 1|s \geqslant a) + 0 \cdot P(N_i = 0|s \geqslant a) \tag{A.4}$$

$$= P(s \geqslant a|N_i = 1)\frac{p_N}{p_{+1}} \tag{A.5}$$

$$= P\left(T_i + \sum_{j\neq i} T_j N_j + Z \geqslant a\right)\frac{p_N}{p_{+1}} \tag{A.6}$$

Here $p_{+1} = P(R=+1)$ is the probability that the observer responds '*X*'. If we use a normal approximation to the sum over j in Eq. (A.6), then the first factor in that equation is the probability of a normal random variable exceeding a criterion. Introducing the symbol $G(x,\mu,\sigma)$ for the normal cumulative distribution function, we can rewrite (A.6) as:

$$= \left[1 - G\left(a, T_i + p_N\sum_{j\neq i} T_j, \sqrt{p_N(1-p_N)\sum_{j\neq i} T_j^2 + \sigma_Z^2}\right)\right]\frac{p_N}{p_{+1}} \tag{A.7}$$

$$= G\left((1-p_N)T_i, a - p_N\sum_{j} T_j, \sqrt{p_N(1-p_N)\sum_{j\neq i} T_j^2 + \sigma_Z^2}\right)\frac{p_N}{p_{+1}} \tag{A.8}$$

If T_i makes only a small contribution to the template, then $\sum_{j\neq i} T_j^2 \approx \sum_j T_j^2$, and (A.8) becomes

$$= G\left((1-p_N)T_i, a - p_N\sum_{j} T_j, \sqrt{p_N(1-p_N)\sum_{j} T_j^2 + \sigma_Z^2}\right)\frac{p_N}{p_{+1}} \tag{A.9}$$

We will define $\mu = p_N\sum_j T_j$, which is the mean of the decision variable s over all trials, and $\sigma^2 = p_N(1-p_N)\sum_j T_j^2$, which is the contribution of the external noise to the variance of the decision variable. Then (A.9) can be simplified to

$$= G\left((1-p_N)T_i, a-\mu, \sqrt{\sigma^2+\sigma_Z^2}\right)\frac{p_N}{p_{+1}} \tag{A.10}$$

When $(1-p_N)T_i$ is small compared to the square root term in Eq. (A.10), which will be true when the effect of the single pixel T_i on the decision variable is small compared to the standard deviation of the decision variable, the right-hand side of (A.10) grows approximately linearly with T_i: $E[N_i|R=+1] \approx k_0 + k_1 T_i$. (This approximation is valid when the template is sufficiently large that each pixel of the template has only a small influence on the observer's responses, which is certainly true in our experiments and in Gosselin and Schyns' experiments. In tasks where observers' responses are determined by a few very small stimulus elements, a different formulation will be necessary.) The constants k_0 and k_1 are the same for all noise pixels N_i, so the expected value of the entire noise field N over all trials where the observer responds +1 recovers the template: $E[N|R=+1] \approx k_0 + k_1 T$. For later use, we will note that the Taylor expansion shows that the constant k_1 is given by

$$k_1 = g\left(0, a-\mu, \sqrt{\sigma^2+\sigma_Z^2}\right)\frac{p_N(1-p_N)}{p_{+1}} \tag{A.11}$$

Here $g(x,\mu,\sigma)$ is the normal probability density function. A similar derivation shows that on trials where the observer responds –1, the expected value of N grows approximately linearly with $-T$, i.e., $E[N|R=-1] \approx l_0 - l_1 T$, with the constant l_1 given by

$$l_1 = g\left(0, a-\mu, \sqrt{\sigma^2+\sigma_Z^2}\right)\frac{p_N(1-p_N)}{p_{-1}} \tag{A.12}$$

The key result that we will use later on is that the expected value of the noise field N on trials where the observer responds ±1 is related linearly to $\pm T$ (although there are also some technicalities relating to the constants k_1 and l_1). Consequently, any type of noise with this property can be used in the bubbles method. Gaussian white noise has this property (Murray et al., 2002), and as we discussed in Section 3, there may be advantages to using Gaussian white noise in the bubbles method, as it seems less likely to disrupt observers' strategies.

A.2. Signals windowed through blurred binary noise

If an observer uses a template T to classify a signal I_X that is windowed through binary noise blurred by a bubble b, the observer's decision variable is $s = ((b * N) \circ I_X) \otimes T + Z$, which can be rewritten as $s = (b * N) \otimes (T \circ I_X) + Z$. (Here * is two-dimensional convolution, and $\circ$ is the pointwise product.) If the bubble b is symmetric about the origin, the decision variable can be further rewritten as $s = N \otimes (b * (T \circ I_X)) + Z$. Thus the observer gives the same responses as an observer who classifies binary white noise fields using a template $T' = b * (T \circ I_X)$, and by the results of the previous section, the expected value of N over trials where the observer responds +1 is therefore $k_0 + k_1 \cdot b * (T \circ I_X)$, where

$$k_1 = g\left(0, a - \mu_X, \sqrt{\sigma_X^2 + \sigma_Z^2}\right) \frac{p_N(1 - p_N)}{p_{+1}} \tag{A.13}$$

Here we have added a subscript X to the variables μ and σ in Eq. (A.11), to emphasize that in general these values depend on the signal I_X.

Similarly, the expected value of N over trials where the observer views a signal I_Y windowed through bubbles and responds −1 is $l_0 - l_1 \cdot b * (T \circ I_Y)$, where

$$l_1 = g\left(0, a - \mu_Y, \sqrt{\sigma_Y^2 + \sigma_Z^2}\right) \frac{p_N(1 - p_N)}{p_{-1}} \tag{A.14}$$

We have assumed that the probability of a bubble occurring at any given location is independent of whether a bubble appears at any other location, and consequently the number of bubbles that appear on any given trial follows a binomial distribution. This differs from Gosselin and Schyns' (2001) formulation, in which one specifies the exact number of bubbles that appear on any given trial. The independence assumption in our formulation greatly simplifies the analysis, and because the number of bubbles normally varies from trial to trial in a staircase anyway, this minor change should not materially affect the results of a bubbles experiment.

A.3. The bubbles image

A bubbles image is defined as the sum of blurred binary noise fields over all trials where the observer gives the correct response, which we will call W_C, divided pointwise by the sum of blurred binary noise fields over all trials, which we will call W_{ALL}.

Consider the sum over correct trials, W_C. By the results of the previous section, the expected value of $b * N$ on trials where the observer correctly responds 'X' or 'Y' is $b * (k_0 + k_1 \cdot b * (T \circ I_X))$ or $b * (l_0 - l_1 \cdot b * (T \circ I_Y))$, respectively. If the constants k_1 and l_1 are equal, and if the observer gives an equal number of 'X' and 'Y' responses, then the expected value of W_C is therefore $w_1 + w_2 \cdot b * b * (T \circ (I_X - I_Y))$ for some w_1 and w_2. Inspection of (A.13) and (A.14) shows that a sufficient condition for $k_1 = l_1$ is that (a) $P_{+1} = P_{-1}$, (b) $\sigma_X^2 = \sigma_Y^2$, and (c) $a - \mu_X = -(a - \mu_Y)$, which can be rewritten as $a = (\mu_X + \mu_Y)/2$. Condition (a) requires that the observer gives unbiased responses. Condition (b) requires that the variance of the decision variable is the same on signal-X and signal-Y trials, which is often approximately true, although exceptions have been reported (Green & Swets, 1974). Condition (c) requires that the observer's criterion lies midway between the mean of the decision variable on signal-X and signal-Y trials, and whenever an observer gives unbiased responses using a decision variable that has the same variance on signal-X and signal-Y trials, this condition will be met. That is, (a) and (b) imply (c).

Second, note that the expected value of the sum of the windowing noise over all trials, W_{ALL}, is the same at all spatial locations, because the bubble locations are uniformly distributed.

Finally, the bubbles image is W_C/W_{ALL}, where W_C has an expected value of $w_1 + w_2 \cdot b * b * (T \circ (I_X - I_Y))$, and W_{ALL} has an expected value that is constant over all locations. The central limit theorem ensures that each pixel of both W_C and W_{ALL} are approximately normal. When the means of two normal random variables are not zero, and when the standard deviations are small compared to the means, then the expected value of the ratio of the two random variables is approximately the ratio of their means. In a bubbles experiment, W_C and W_{ALL} meet both these conditions after a reasonably large number of trials. Thus the expected value of the ratio W_C/W_{ALL} is approximately proportional to the expected value of W_C, which is to say that the expected value of the bubbles image is $u + v \cdot b * b * (T \circ (I_X - I_Y))$, for some u and v, as in Eq. (5).

We should note that treating W_{ALL} as a constant neglects the fact that dividing by W_{ALL} actually helps to correct for small variations in how many bubbles appear at different locations over the course of an experiment, due to random sampling fluctuations. However, the mathematics is much simpler if we treat W_{ALL} as a constant, and after just a few hundred trials the variation in W_{ALL} from place to place is small, so the correction of dividing by W_{ALL} is also small.

References

Ahumada, A. J., Jr. (1996). Perceptual classification images from Vernier acuity masked by noise [ECVP abstract]. *Perception, 26*(Suppl.), 18.

Ahumada, A. J., Jr., & Beard, B. L. (1999). Classification images for detection [ARVO abstract #3015]. *Investigative Ophthalmology and Visual Science, 40*(4), S572.

Ahumada, A. J., Jr., & Lovell, J. (1971). Stimulus features in signal detection. *Journal of the Acoustical Society of America, 49*(6), 1751–1756.

Beard, B. L., & Ahumada, A. J., Jr. (1998). A technique to extract relevant image features for visual tasks. In B. E. Rogowitz & T. N. Pappas (Eds.), *SPIE Proceedings: Vol. 3299. Human Vision and Electronic Imaging III* (pp. 79–85). Bellingham, WA: SPIE.

Burgess, A. E., Wagner, R. F., Jennings, R. J., & Barlow, H. B. (1981). Efficiency of human visual signal discrimination. *Science, 214*(4516), 93–94.

Gold, J. M., Bennett, P. J., & Sekuler, A. B. (1999). Identification of band-pass filtered letters and faces by human and ideal observers. *Vision Research, 39*(21), 3537–3560.

Gold, J. M., Murray, R. F., Bennett, P. J., & Sekuler, A. B. (2000). Deriving behavioural receptive fields for visually completed contours. *Current Biology, 10*(11), 663–666.

Gosselin, F., & Schyns, P. G. (2001). Bubbles: A technique to reveal the use of information in recognition tasks. *Vision Research, 41*(17), 2261–2271.

Gosselin, F., & Schyns, P. G. (2002). RAP: A new framework for visual categorization. *Trends in Cognitive Sciences, 6*(2), 70–77.

Green, D. M., & Swets, J. A. (1974). *Signal detection theory and psychophysics*. Huntington, NY: R.E. Krieger Publishing Company.

Marmarelis, P. Z., & Marmarelis, V. Z. (1978). *Analysis of physiological systems: The white-noise approach*. New York: Plenum Press.

Murray, R. F. (2002). *Perceptual organization and the efficiency of shape discrimination*. Ph.D. thesis, University of Toronto.

Murray, R. F., Bennett, P. J., & Sekuler, A. B. (2002). Optimal methods for calculating classification images: Weighted sums. *Journal of Vision, 2*(1), 79–104.

Nabet, B., & Pinter, R. B. (1992). *Nonlinear vision: Determination of neural receptive fields, function, and networks*. Boca Raton, FL: CRC Press.

Pelli, D. G. (1990). The quantum efficiency of vision. In C. Blakemore (Ed.), *Vision: coding and efficiency* (pp. 3–24). Cambridge: Cambridge University Press.

Richards, V. M., & Zhu, S. (1994). Relative estimates of combination weights, decision criteria, and internal noise based on correlation coefficients. *Journal of the Acoustical Society of America, 95*(1), 423–434.

Schwartz, O., Bayer, H. M., & Pelli, D. G. (1998). Features, frequencies, and facial expressions [ARVO abstract #825]. *Investigative Ophthalmology and Visual Science, 39*(4), S173.

Solomon, J. A., & Pelli, D. G. (1994). The visual filter mediating letter identification. *Nature, 369*(6479), 395–397.

Watson, A. B., & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception and Psychophysics, 33*(2), 113–120.

Wiener, N. (1958). *Nonlinear problems in random theory*. Cambridge, MA: Technology Press of Massachusetts Institute of Technology.