Accurate statistical tests for smooth classification images.

Alan Chauvin, Keith Worsley, Philippe Schyns, Martin Arquín, Frédéric Gosselin

To cite this version:


HAL Id: hal-00806723
https://hal.archives-ouvertes.fr/hal-00806723
Submitted on 2 Apr 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Accurate statistical tests for smooth classification images

Alan Chauvin
Département de Psychologie, Université de Montréal, Montréal, QC, Canada

Keith J. Worsley
Department of Mathematics and Statistics, McGill University, Montréal, QC, Canada

Philippe G. Schyns
Department of Psychology, University of Glasgow, Glasgow, United Kingdom

Martin Arguin
Département de Psychologie, Université de Montréal, Montréal, QC, Canada

Frédéric Gosselin
Département de Psychologie, Université de Montréal, Montréal, QC, Canada

Despite an obvious demand for a variety of statistical tests adapted to classification images, few have been proposed. We argue that two statistical tests based on random field theory (RFT) satisfy this need for smooth classification images. We illustrate these tests on classification images representative of the literature from F. Gosselin and P. G. Schyns (2001) and from A. B. Sekuler, C. M. Gaspar, J. M. Gold, and P. J. Bennett (2004). The necessary computations are performed using the Stat4Ci Matlab toolbox.

Keywords: classification images, reverse correlation, Bubbles, random field theory

Introduction

In recent years, vision research has witnessed a tremendous growth of interest for regression techniques capable of revealing the use of information (e.g., Ahumada, 1996; Eckstein & Ahumada, 2002; Gosselin & Schyns, 2004b). Reverse correlation, one such technique, has been employed in a number of domains ranging from electroretinograms (Sutter & Tran, 1992), visual simple response time (Simpson, Braun, Bargen, & Newman, 2000), single pulse detection (Thomas & Knoblauch, 1998), vernier acuity (Barth, Beard, & Ahumada, 1999; Beard & Ahumada, 1998), objects discrimination (Olman & Kersten, 2004), stereopsis (Gosselin, Bacon, & Mamassian, 2004; Neri, Parker, & Blakemore, 1999), letter discrimination (Gosselin & Schyns, 2003; Watson, 1998; Watson & Rosenholtz, 1997), single neuron’s receptive field (e.g., Marmarelis & Naka, 1972; Ohzawa, DeAngelis, & Freeman, 1990; Ringach & Shapley, 2004), modal and amodal completion (Gold, Murray, Bennett, & Sekuler, 2000), face representations (Gold, Sekuler, & Bennett, 2004; Kontsevich & Tyler, 2004; Mangini & Biederman, 2004; Sekuler et al., 2004) to temporal processing (Neri & Heeger, 2002). Bubbles, a related technique (Gosselin & Schyns, 2001, 2002, 2004b; Murray & Gold, 2004), has revealed the use of information for the categorization of face identity, expression, and gender (Adolphs et al., 2005; Gosselin & Schyns, 2001; Schyns, Bonnar, & Gosselin, 2002; Smith, Cottrell, Gosselin, & Schyns, 2005; Vinette, Gosselin, & Schyns, 2004), for the categorization of natural scenes (McCotter, Sowden, Gosselin, & Schyns, in press), for the perception of an ambiguous figure (Bonnar, Gosselin, & Schyns, 2002), and for the interpretation of EEG signals (Schyns, JentzsCh, Johnson, Schweinberger, & Gosselin, 2003; Smith, Gosselin, & Schyns, 2004).

Both the Bubbles and the reverse correlation techniques produce large volumes of regression coefficients that have to be tested individually. As we will shortly discuss, this raises the issue of false positives: the risk of accepting an event that occurred by chance. Surprisingly, few classification image researchers have taken this into account (for exceptions, see Abbey & Eckstein, 2002; Kontsevich & Tyler, 2004; Mangini & Biederman, 2004). Here, we argue that two statistical tests based on random field theory (RFT) satisfy this need for smooth classification images. The core ideas of RFT are presented. In particular, the main equations for the tests are given. Finally, the usage of a Matlab toolbox implementing the tests is illustrated on two representative sets of classification images from Gosselin and Schyns (2001) and Sekuler et al. (2004). But first, in order to identify the critical properties of the proposed statistical tests, we shall discuss some limitations.
of the two statistical tests that have already been applied to classification images.

**Multiple comparisons**

In a typical classification image experiment, an observer has to classify objects partially revealed by additive (reverse correlation) or multiplicative (Bubbles) noise fields. The calculation of the classification image amounts quite simply to summing all the noise fields weighted by the observer’s responses (Ahumada, 2002; Murray, Bennett, & Sekuler, 2002). By doing this, the researcher is actually performing a multiple regression on the observer’s responses and the noise fields (see Appendix). A statistical test compares these data against the distribution of a random process with similar characteristics. Classification images can thus be viewed, under the null hypothesis, as expressions of a random \( N \)-dimensional process (i.e., a random field). The alternate hypothesis is that a signal—known or unknown—is hidden in the regression coefficients.

So far, researchers have used two statistical tests to achieve this end: the Bonferroni correction and Abbey and Eckstein’s (2002) Hotelling test. We will argue that these tests are not adapted to some classification images. The former is too conservative when the elements of the classification images are locally correlated, and the latter is not suitable in the absence of a priori expectations about the shape of the signal hidden in the classification images.

**Bonferroni correction**

Consider a one-regression coefficient \( Z \)-scored classification image (see Appendix). If this \( Z \) score exceeds a threshold determined by a specified \( p \) value, this regression coefficient differs significantly from the null hypothesis. For example, a \( p \) value of .05 means that if we reject the null hypothesis, that is, if the \( Z \) score exceeds a threshold \( t_Z = 1.64 \), the probability of a false alarm (or Type I error) is .05. Now consider a classification image comprising 100 regression coefficients: the expected number of false alarms is \( 100 \times 0.05 = 5 \). With multiple \( Z \) tests, such as in the previous example, the overall \( p \) value can be set conservatively using the Bonferroni correction: \( p_{BON} = p(Z > t_{BON})N \), with \( N \) the number of points in the classification image. Again, consider our hypothetical 100-point classification image. The corrected threshold, \( t_{BON} \), associated with \( p_{BON} = .05 \), is 3.29. Such high \( Z \) scores are seldom observed in classification images derived from empirical data. In a classification image of 65,536 data points (typical of those found in the literature, like the \( 256 \times 256 \) classification images from Gosselin & Schyns, 2001, reanalyzed in the last section of this article), \( t_{BON} \) becomes a formidable 4.81! For classification images of low (or reduced) dimensionality such as those of Mangini and Biederman (2004) or Kontsevich and Tyler (2004), the Bonferroni correction prescribes thresholds that can be (and have been) attained.

**A priori expectations**

Two possibilities should be considered: either these classification images really do not contain anything statistically significant (which seems unlikely given the robustness of the results obtained with no Bonferroni correction; e.g., Gosselin & Schyns, 2001; Schyns et al., 2002, 2003), or the Bonferroni correction is too conservative. Do we have a priori expectations that support the latter and can we use these expectations to our advantage? Abbey and Eckstein (2002), for example, have derived a statistical test far more sensitive than the Bonferroni correction for classification images derived from a two-alternative forced-choice paradigm when the signal is perfectly known. Although we often do not have such perfect a priori knowledge about the content of classification images, we do expect them to be relatively smooth.

The goal of Bubbles and reverse correlation is to reveal clusters of points that are associated with the measured response; for example, the mouth or the eyes of a face (Gold et al., 2004; Gosselin & Schyns, 2001; Mangini & Biederman, 2004; Schyns et al., 2002; Sekuler et al., 2004), illusory contours (Gold et al., 2000), and so on. In other words, it is expected that the data points of classification images are correlated, introducing “smoothness” in the solutions. The Bonferroni correction, adequate when data points are independent, becomes far too conservative (not sensitive enough) for classification images with a correlated structure.

In the next section, we present two statistical tests based on RFT that provide accurate thresholds for smooth, high-dimensional classification images.

**Random field theory**

Adler (1981) and Worsley (1994, 1995a, 1995b, 1996) have shown that the probability of observing a cluster of pixels exceeding a threshold in a smooth Gaussian random field is well approximated by the expected Euler characteristic (EC). The EC basically counts the number of clusters above a sufficiently high threshold in a smooth Gaussian random fields. Raising the threshold until only one cluster remains brings the EC value to 1; raising it still further until no cluster exceeds the threshold brings it to 0. Between these two thresholds, the expected EC approximates the probability of observing one cluster. The formal proof of this assertion is the centerpiece of RFT.
Next we present the main equations of two statistical tests derived from RFT: the so-called pixel and cluster tests, which have already been successfully applied for more than 15 years to brain imaging data. Crucially, these tests take into account the spatial correlation inherent to the data set, making them well suited for classification images.

**Pixel test**

Suppose that $Z$ is a $Z$-scored classification image (see Appendix). In RFT, the subset of $Z$ searched for unlikely clusters of regression coefficients—e.g., the face area—is called the search space ($S$). The probability of observing at least one regression coefficient exceeding $t$ is well approximated by

$$P(\max Z > t) \sim \sum_{d=0}^{D} \text{Resels}_d(S) \cdot EC_d(t)$$

where $D$ is the dimensionality of $S$; $EC_d(t)$ is the $d$-dimensional EC density that depends partly on the type of statistic (for EC densities of other random fields, see Cao & Worsley, 2001; Worsley, Marrett, Neelin, Vandal, Friston, & Evans, 1996); $\text{Resels}_d(S)$ is the $d$-dimensional Resels (resolution elements), which varies with the size and the shape of $S$. The EC densities of a $D$-two-dimensional Gaussian random field $Z$ are

$$EC_0(t) = \int (2\pi)^{-1/2} e^{-u^2/2} du = p(Z > t),$$

$$EC_1(t) = \sqrt{4 \ln(2) / 2\pi} \cdot e^{-t^2},$$

and

$$EC_2(t) = 4 \ln(2) \cdot t \cdot e^{-t^2}.$$  

The Resels is given by

$$\text{Resels}_d(S) = \frac{V_d(S)}{\text{FWHM}^d},$$

where $V_d(S) = 1$ for a connected search region, $V_1(S) =$ half perimeter length of $S$, $V_2(S) =$ caliper area of $S$ (a disk of the same area as $S$ gives a good approximation and allows to derive the volumes of the lower dimension; see Cao & Worsley, 2001). The FWHM is the full width at half maximum of the filter $f$ used to smooth the independent error noise in the image. If the filter is Gaussian with standard deviation $\sigma_f$, then

$$\text{FWHM} = \sigma_f \sqrt{8 \ln 2}. $$

The filter $f$ should be chosen to give the best discrimination, or in other words to maximize the detection of signal in $Z$. There is a classic theorem in signal processing, the matched filter theorem, which states that to detect signal added to white noise, the optimum filter should match the shape of the signal. This implies that to optimally detect, say 10 pixel features, we should smooth the data with 10 pixels FWHM filter. But if for instance it was felt that larger contiguous areas of the image were involved in discrimination, then this might be better detected by using a broader filter at the statistical analysis stage (see Worsley et al., 1996).

This dependency of the pixel test on the choice of an adequate filter has led to a generalization of the test in which an extra dimension, the scale of the filter, is added to the image to create a scale space image (Poline & Mazoyer, 1994; Siegmund & Worsley, 1995). The scale space search reduces the uncertainty of choosing a filter FWHM but at the cost of higher thresholds.

**Cluster test**

The pixel test computes a statistical threshold based on the probability of observing a single pixel above the threshold. This test has been shown to be best suited for detecting focal signals with high $Z$ scores (Poline, Worsley, Evans & Friston, 1997). But if the region of interest in the search space (the mouth in a face for example) is wide, it has usually a lower $Z$ score and cannot be detected. We could improve detection by applying more smoothing to the image. The amount of smoothing will depend on the extent of the features we wish to detect (by the matched filter theorem), but we do not know this in advance.

Friston, Worsley, Frackowiak, Mazziotta, and Evans (1994) proposed an alternative to the pixel test to improve the detection of wide signals with low $Z$ scores (for a review, see Poline et al., 1997). The idea is to set a low threshold ($t \geq 2.3$—in the next section, we used $t = 2.7$) and base the test on the size of clusters of connected pixels above the threshold. The cluster test is based on the probability that, above a threshold $t$, a cluster of size $K$ (or more) pixels has occurred by chance that is calculated in the $D = 2$ case as follows (Cao & Worsley, 2001; Friston et al., 1994):

$$P(K > k) \sim 1 - e^{(-\text{Resels}(S) EC_1(t)p)},$$

where

$$p = e^{(\sqrt{2\pi}EC_1(t)k/(\text{FWHM}^2p(Z > t)))}. $$

**Cluster versus pixel test**

The cluster and the pixel test presented above provide accurate thresholds but for different types of signal. The pixel test is based on the maximum of a random field
and therefore is best adapted for focal signal (optimally the size of the FWHM) with high Z scores (Poline et al., 1997; Siegmund & Worsley, 1995). The cluster test is based on the size of a cluster above a relatively low threshold and therefore is more sensitive for detecting wide regions of contiguous pixels with relatively low Z scores. The two tests potentially identify different statistically significant regions in smooth classification images. Figure 1 illustrates this point with a one-dimensional classification image comprising 257 pixels convolved with a Gaussian kernel with an FWHM of 11.8 pixels. For a p value of .05, the pixel test gives a threshold of 3.1 (green line) whereas the cluster test gives a minimum cluster size of 6.9 above a threshold of 2.7 (red line).

Furthermore, the interpretation of the results following the application of the pixel and the cluster test differs drastically. On the one hand, the cluster test allows the inference that the clusters of Z scores larger than the minimum size are significant, not that the individual Z scores inside these clusters are significant. On the other hand, the pixel test allows the conclusion that each individual Z score above threshold is significant (Friston, Holmes, Poline, Price, Frith, 1996; Friston et al., 1994; Poline et al., 1997).

**Accuracy**

Since the late 1980s, RFT has been used to analyze positron emission tomography (PET) images, galaxy density maps, cosmic microwave background data, and functional magnetic resonance imaging (fMRI) data. In fact, the RFT is at the heart of two popular fMRI data analysis packages: SPM2 (Frackowiak et al., 2003) and FMRIstat (Worsley, 2003).

Not surprisingly, the accuracy of RFT has been examined extensively. An accurate statistical test must be both sensitive (i.e., high hit rate) and specific (i.e., high correct rejection rate). In particular, RFT has been evaluated in the context of so-called “phantom” simulations (Hayasaka, Luan Phan, Liberzon, Worsley, & Nichols, 2004; Hayasaka & Nichols, 2003; Poline et al., 1997; Worsley, 2005). A “phantom” simulation basically consists of generating a lot of smooth random regression coefficients, hiding a “phantom” in them (i.e., a known signal—usually a disc or a Gaussian), attempting to detect the “phantom” with various statistical tests, and deriving, per statistical test, a measure of accuracy such as a d′-prime or an ROC area. We singled-out “phantom” simulations for a reason: If we were to compare the accuracy of various statistical tests for the detection of a “phantom” template (e.g., used by a linear amplifier model) in a smooth classification images, this is exactly what we would have to do. In other words, these “phantom” simulations inform us just as much about the accuracy of RFT for fMRI data than about its accuracy for classification images.

To summarize these assessments, the p values given by RFT appears to be more accurate than those given by the Bonferroni, the Hochberg, the Holm, the Sidák, and the false discovery rate, provided that the size of the search space is greater than about three times that of the FWHM (Hayasaka & Nichols, 2003), that the FWHM is greater than about five pixels (Taylor, Worsley, & Gosselin, 2005), and that the degree of freedom is greater than about 200. Also, the cluster test is more sensitive and less specific than the pixel test.

**Reanalyzing representative classification images**

In the final section of this article, we apply the pixel and cluster tests to classification images representative of the literature from Gosselin and Schyns (2001) and Sekuler et al. (2004). We give sample commands for the Stat4Ci Matlab toolbox throughout.

**Matlab implementation**

A mere four pieces of information are required for the computation of the significant regions using the pixel and the cluster tests: a desired p value, a threshold t (only used for the cluster test), a search space, and the FWHM—or, equivalently, the sigma—of the Gaussian kernel used to smooth the classification image. The main function from the Stat4Ci Matlab toolbox—StatThresh.m—inputs this information together with a suitably prepared classification image (i.e., smoothed and Z-scored), performs all the computations described above, and outputs a threshold for the pixel test as well as the minimum size of a significant cluster for the cluster test. The StatThresh.m function makes extensive use of the stat_threshold.m function, which was

![Figure 1. Regions revealed by the cluster (red) versus the pixel (green) test. See text for details.](image-url)
originally written by Keith Worsley for the \textit{FMRISTAT} toolbox.

Other functions included in the \textit{Stat4Ci} toolbox perform a variety of related computations; for example, \texttt{ReadCid.m} reads a classification image data (CID) file; \texttt{BuildCi.m} constructs classification images from a CID file; \texttt{SmoothCi.m} convolves a raw classification image with a Gaussian filter; \texttt{ExpectedSCi.m} computes the expected mean and standard deviation of a smooth classification image (see Appendix); \texttt{ZscoreSCi.m} \texttt{Z}-scores a smoothed classification image (see Equation 9 and Appendix); and \texttt{DisplayRes.m} displays the thresholded \texttt{Z}-transformed smooth classification image and outputs a summary table (see Figure 2). All of these functions include thorough help sections.

\textbf{Sekuler et al. (2004)}

Sekuler et al. examined the effect of face inversion on the information used by human observers to resolve an identification task. Four classification images extracted using reverse correlation are reanalyzed: one for each combination of two subjects (MAT and CMG) and two conditions (UPRIGHT and INVERTED). Each classification image cumulates the data from 10,000 trials. We will not further describe this experiment. Rather we will limit the presentation to what is required for the application of the pixel and cluster tests.

First, the raw classification images must be convolved with a Gaussian filter (i.e., smoothed). The choice of the appropriate Gaussian filter depends essentially on the size of the search space (for a discussion, see Worsley, 2005). We chose a Gaussian filter with a standard deviation of $\sigma_0 = 4$ pixels; its effect is similar to those of the filter used by Sekuler et al. (2004). Second, the smooth classification images must be \texttt{Z}-scored. This can sometimes be achieved analytically (see Appendix). However, if the number of trials is greater than 200—as is usually the case with classification images—the transformation can be approximated as follows:

\begin{align*}
SC_i &= \text{double}({\tt imread}('GenderCi.tiff')); \\
\sigma_{b} &= 20; \% \text{std of smoothing filter} \\
S_r &= \text{double}({\tt imread}('faceMask.tif'))/255; \\
vecSCi &= SCi(eq(S_r,0)); \\
ZSCi &= ZTransSCi(SCi,...] \\
&= \text{mean}(\text{vecSCi}(:,))\text{, std}(\text{vecSCi}(:,))]; \\
ZSCi &= ZSCi .* S_r; \\
p &= .05; \% \text{p-value} \\
tC &= 2.7; \% \text{threshold (for Cluster test)} \\
Res &= \text{StatThresh}(ZSCi,p,\sigma_{b},tC,S_r); \\
background &= \text{double}({\tt imread}('w1H.JPG')); \\
tCi &= \text{DisplayRes}(Res,background); \\
\end{align*}

where the mean and standard deviation are estimated directly from the data, preferably from signal-less regions of the classification images (e.g., regions corresponding to a homogeneous background). In the \textit{Stat4Ci} toolbox, classification image preparation can be done as illustrated in Figure 3.

Once the classification image has been smoothed and \texttt{Z}-scored, it must be inputted into the \texttt{StatThresh.m} function together with the four additional required pieces
of information: a \( p \) value (\( p \leq .05 \)), the sigma of the Gaussian filter used during the smoothing phase (\( \sigma_b = 4 \) pixels for this reanalysis), a threshold for the cluster test (equal to 2.7 for this reanalysis), and a search space (i.e., the face region).

The statistical threshold obtained using the pixel test is very low compared with that obtained using the Bonferroni correction (i.e., 3.67 rather than 4.5228; see Bonferroni correction). In fact, the \( \text{stat\_threshold.m} \) function outputs the minimum between the Bonferroni and the pixel test thresholds. Figure 4 displays the thresholded classification images for both the pixel and cluster tests. For the cluster test, only the clusters larger than the minimum size (i.e., 66.6 pixels) are shown. Red pixels indicate the regions that attained significance in the UPRIGHT condition; green pixels, in the INVERTED condition; and yellow pixels, in both. A face background was overlaid to facilitate interpretation.

**Gosselin and Schyns (2001)**

Gosselin and Schyns (Experiment 1) examined the information used by human observers to resolve a GENDER and an expressive versus not expressive (EXNEX) face discrimination task. They employed the Bubbles technique to extract two classification images per observer, one per task. Each of the two classification images reanalyzed in this section combines the data from 500 trials executed by subject FG. The classification images can be built either from the opaque masks punctured by Gaussian holes and applied multiplicatively to a face on each trial, or from the center of these Gaussian holes. The former option naturally results in smooth classification images; and the latter option calls for smoothing with a filter, just like the classification images of Sekuler et al. (2004). For this reanalysis, we used a Gaussian filter identical to the one used to sample information during the actual experiment (\( \sigma_b = 20 \) pixels). In this case, both options are strictly equivalent. These smooth classification images (see SCi in Figure 3) were Z-scored using Equation 9 with estimations of the expected means and standard deviations based on the signal-less pixels outside the search region.

Next, the Z-transformed smooth classification image (see ZSCi in Figure 3) is inputted into the \( \text{StatThresh.m} \) function with the four additional required pieces of information: a \( p \) value (\( p \leq .05 \)), the sigma of the Gaussian filter used during the smoothing phase (\( \sigma_b = 20 \) pixels for this reanalysis), a threshold for the cluster test (equal to 2.7 for this reanalysis; see tC in Figure 3), and a search space (see \( S_r \) in Figure 3). See Figure 3 for all the relevant \( \text{Stat4Ci} \) toolbox commands.

Again, the statistical threshold obtained using the pixel test is extremely low compared with that obtained using the Bonferroni correction: 3.30 rather than 4.808 (see Bonferroni correction). Figure 5 displays the thresholded classification images for both the pixel and cluster tests. For the cluster test, only the clusters larger than the minimum size (i.e., 861.7 pixels) are shown. Red pixels indicate the regions that attained statistical significance. A face background (see background in Figure 3) was overlaid to facilitate interpretation.

**Take-home message**

We have presented two statistical tests suitable for smooth, high-dimensional classification images in the absence of a priori expectations about the shape of the signal. The pixel and the cluster tests, based on RFT, are accurate within known boundaries discussed in the article. These tests require only four pieces of information and their computation can be performed easily using the \( \text{Stat4Ci} \) Matlab toolbox. We expect these tests to be most useful for researchers applying Bubbles or reverse correlation to complex stimuli.

**Appendix: The construction of a classification image**

In a reverse correlation or Bubbles experiment, an observer is presented with a noise field on trial \( i (i = 1, \ldots, n) \) and produces the response \( Y_i \).
At a particular pixel \( v \), we suppose that some feature of the noise field, \( X(v) \), is correlated with the response. In a reverse correlation experiment, \( X(v) \) might be the added noise at pixel \( v \); in a Bubbles experiment, \( X(v) \) might be the actual bubble mask at pixel \( v \). We aim to detect those pixels where the response is highly correlated with the image feature of interest. The sample correlation at pixel \( v \) is

\[
C(v) = \frac{\sum_i (X_i(v) - \bar{X}(v)) (Y_i - \bar{Y})}{\sqrt{\sum_i (X_i(v) - \bar{X}(v))^2} \sqrt{\sum_i (Y_i - \bar{Y})^2}},
\]

where bar indicates averaging over all \( n \) trials. It is straightforward to show that if there is no correlation between image features and response, then

\[
Z(v) = \frac{\sqrt{n-2}C(v)}{\sqrt{1 - C(v)}},
\]

has a Student-\( t \) distribution with \( n - 2 \) degrees of freedom provided \( X(v) \) is Gaussian, but in any case \( n \) is usually very large so the standard normal distribution will be a very good approximation, by the Central Limit Theorem.

Provided that \( \Sigma X_i = 0 \) and \( \Sigma Y_i = 0 \), the \textit{ZTransSci.m} function from the \texttt{Stat4Ci} toolbox implements Equation A1. In this case, the numerator is simply the sum of all the noise fields weighted by the observer’s responses. The remaining term in \( C(v) \) can be approximated if \( X_i(v) \) is white noise \( W_i(v) \) (i.e. independent and identically distributed noise at each pixel) convolved with a filter \( f(v) \), that is if

\[
X_i(v) = \sum_u f(v-u)W_i(u).
\]

In the case of reverse correlation, \( W_i(v) \) is usually a Gaussian random variable and often there is no filtering, so \( f(v) \) is zero except for \( f(0) = 1 \). In the case of Bubbles, \( W_i(v) \) is a binary random variable taking the value 1 if there is a bubble centered at \( v \), and 0 otherwise. The

\[
\sum_i (X_i(v) - \bar{X}(v))^2 : n\sigma^2 \sum_u f(u)^2,
\]

where \( \sigma^2 \) is the variance of the white noise. For reverse correlation, \( \sigma^2 \) is the variance of the Gaussian white noise. For Bubbles, it is the Binomial variance

\[
\sigma^2 = \frac{N_b}{N} \left( 1 - \frac{N_b}{N} \right) \cdot \frac{N_b}{N},
\]

where \( N_b \) is the number of bubbles and \( N \) is the number of pixels.

The central limit theorem ensures that, at the limit, \( Z(v) \) is a Gaussian random field with an effective FWHM equals to the FWHM of the filter \( f(v) \). The rate of convergence toward Gaussianity depends partly on the predictive variable and partly on the total number of bubbles per Resels. Worsley (2005) has examined the exactness of the \( p \) values given by the Gaussian procedures presented in this article in function of these two factors: At 10,000 bubbles per Resels, the \( p \) values given by the Gaussian procedures depart from the true \( p \) values by less than \( \pm 0.04 \) logarithmic unit; at 500 bubbles per Resels, a figure more often encountered in practice (e.g., Gosselin & Schyns, 2001), the discrepancy can be as much as \( \pm 0.3 \) logarithmic unit. If the predictive variable has a positively skewed distribution, the Gaussian procedure is liberal; and if it has a negatively skewed distribution, as is usually the case in practice (e.g., Gosselin & Schyns, 2001), the Gaussian procedure is conservative.

Acknowledgments

This research was supported by an NSERC (249974) and a NATEQ (03NC082052) grant awarded to Frédéric Gosselin; by a NATEQ (84180) grant awarded to Martin Arguin and Frédéric Gosselin; and by an ESRC grant R000237901 to Philippe G. Schyns. We thank Allison Sekuler, Carl Gaspar, Jason Gold, and Patrick Bennett for having kindly given us access to their classification images.

Commercial relationships: none.

Corresponding author: Frédéric Gosselin.

Email: frederic.gosselin@umontreal.ca.

Address: Département de psychologie, Université de Montréal, C.P. 6128, Succursale Centre-ville, Montréal, Québec, Canada H3C 3J7.

References


Abbey, C. K., & Eckstein, M. P. (2002). Classification image analysis: Estimation and statistical inference for two-alternative forced-choice experiments. \textit{Journal...


