Brief Communications

# Temporal Resolution for the Perception of Features and Conjunctions

**Clara Bodelón, Mazyar Fallah, and John H. Reynolds**
Systems Neurobiology Laboratory, The Salk Institute for Biological Studies, La Jolla, California 92037

The visual system decomposes stimuli into their constituent features, represented by neurons with different feature selectivities. How the signals carried by these feature-selective neurons are integrated into coherent object representations is unknown. To constrain the set of possible integrative mechanisms, we quantified the temporal resolution of perception for color, orientation, and conjunctions of these two features. We find that temporal resolution is measurably higher for each feature than for their conjunction, indicating that time is required to integrate features into a perceptual whole. This finding places temporal limits on the mechanisms that could mediate this form of perceptual integration.

*Key words:* psychophysics; temporal; integration; color; orientation; perception

## Introduction

A central observation that has driven modern thinking about vision is that the visual system decomposes stimuli into their constituent features, encoded by neurons that differ in their feature selectivity, distributed across functionally specialized visual areas (Hubel and Wiesel, 1968; Ungerleider and Mishkin, 1982; Sincich and Horton, 2005). The question of how these neuronal signals relate to the perception of unitary objects stands among the most formidable challenges in modern neuroscience and has been debated since Ramón y Cajal introduced the neuron doctrine over a century ago (Treisman and Gelade, 1980; Wolfe et al., 1989; Wolfe and Bennett, 1997; Ghose and Maunsell, 1999; Riesenhuber and Poggio, 1999; Shadlen and Movshon, 1999). A leading proposal is that features are bound together automatically by the successive elaboration of progressively more complex receptive field properties (Ghose and Maunsell, 1999; Riesenhuber and Poggio, 1999; Shadlen and Movshon, 1999), a proposal that is a direct extension of the simple-cell to complex-cell hierarchy postulated by Hubel and Wiesel (1965). This elaborated receptive field model is supported by neurophysiological studies that have revealed neurons with increasingly complex response properties at successive stages of visual processing (Gross et al., 1969; Maunsell and Newsome, 1987; Zeki and Shipp, 1989). Under this proposal, the color and orientation of a stimulus would be encoded by two different populations of neurons and then bound together at a later stage.

Holcombe and Cavanagh (2001) developed a new psychophysical method for quantifying the temporal resolution of the visual system for conjunctions of color and orientation. They presented sequences composed of colored oriented gratings and measured the maximum presentation rate at which observers could perceive how the features were conjoined. Observers could reliably perceive conjunctions when each grating appeared for only ~25 ms. This approaches the range of human flicker fusion thresholds, allowing almost no time for feature integration to occur. This raised the possibility that conjunctions of color and orientation are perceived the moment their constituent features are perceived, without requiring a time-consuming integrative computation. To test this, we quantified the temporal resolution of the human visual system for color, orientation and conjunctions. We find that color and orientation are reliably perceived at frequencies beyond the temporal resolution for conjunctions. These findings demonstrate that conjunctions are processed more slowly than features and impose tight constraints on models of feature integration.

## Materials and Methods

*Stimuli.* Stimuli were presented on a Sony (Tokyo, Japan) Multiscan E500 monitor running at 160 Hz (graphics card: ATI Radeon 9600 XT). Eight sequences were constructed, each of which was composed of repeated presentations of two gratings (Fig. 1*A*). The colors of the gratings in each sequence were selected from one of two color pairs (either red and green or gray and yellow). These color pairs were designed to ensure that, at high presentation frequencies, both pairs were perceptually indistinguishable. This was accomplished by selecting pairs of colors for the light stripes of the gratings such that in the three-dimensional space of cone photoreceptor excitations, each pair had a common midpoint (Fig. 1*A*). The same procedure was used to select dark colors in each grating. In addition, for each pair of colors, the line in cone color space that connects the dark version of each color to the light version of the other color had a common midpoint, and this midpoint was constant across color pairs. Luminances for the light stripes were equated across colors, as were luminance values for the dark stripes. The monitor was calibrated with a PR650 spectrophotometer. The CIE values for dark red were $x = 0.383$,
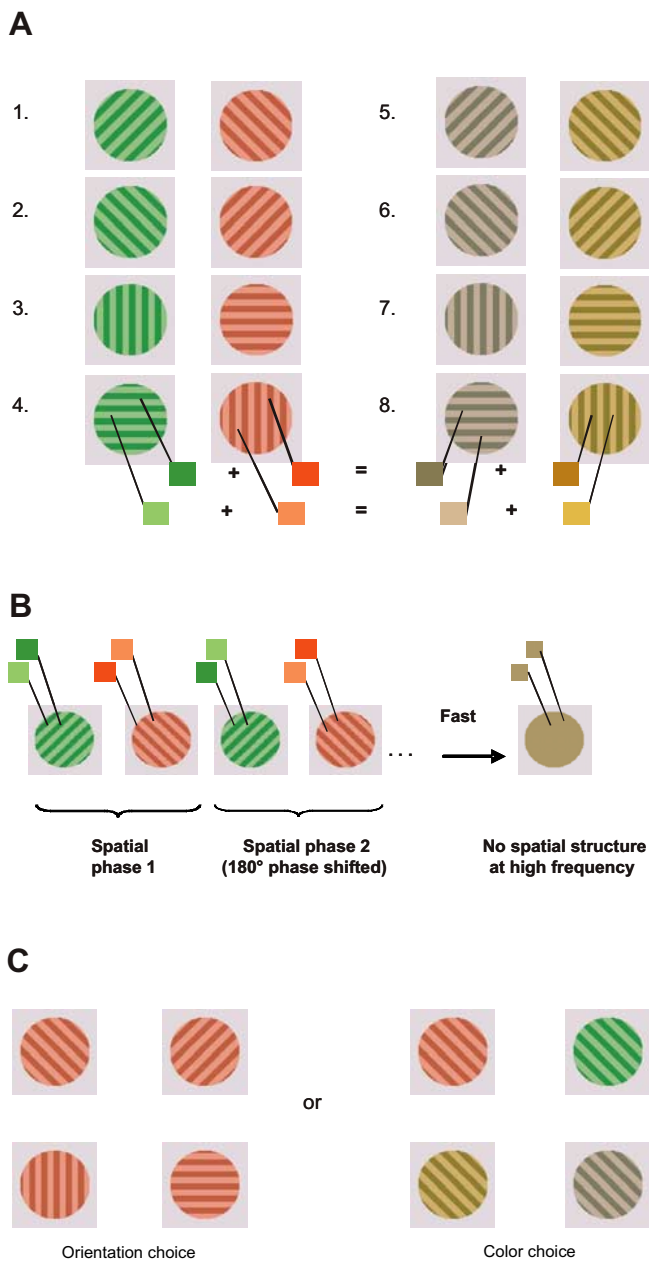
**A**



**B**



**C**



**Figure 1.** Experimental paradigm. **A**, The eight sequences, each of which was composed of two gratings. Colors were selected so that at sufficiently high presentation frequencies, the two possible color pairs combined to form the same color. **B**, The relative spatial phases of each successive instance of a given grating were offset by 180°. At sufficiently high presentation frequencies, this yielded a spatially uniform stimulus with no orientation information. **C**, Examples of orientation choice and of color choice.

$y = 0.304$, luminance (lum) $= 46$ cd/m$^2$; for light red were $x = 0.361$, $y = 0.315$, lum $= 84$ cd/m$^2$; dark green were $x = 0.273$, $y = 0.358$, lum $= 46$ cd/m$^2$; for light green were $x = 0.301$, $y = 0.344$, lum $= 84$ cd/m$^2$; dark yellow were $x = 0.362$, $y = 0.358$, lum $= 46$ cd/m$^2$; for light yellow were $x = 0.348$, $y = 0.344$, lum $= 84$ cd/m$^2$; dark gray were $x = 0.307$, $y = 0.304$, lum $= 46$ cd/m$^2$; and for light gray were $x = 0.318$, $y = 0.315$, lum $= 84$ cd/m$^2$. The background luminance was equal to the mean luminance of the dark and light stripes (65 cd/m$^2$) and had chromaticity coordinates of $x = 0.332$, $y = 0.329$.

*Task.* At the beginning of each trial, a black fixation point appeared at the center of the screen. Observers maintained fixation within a 1° radius square fixation window. Eye position was continuously monitored throughout the trial using an ISCAN (Burlington, MA) Model ETL-400
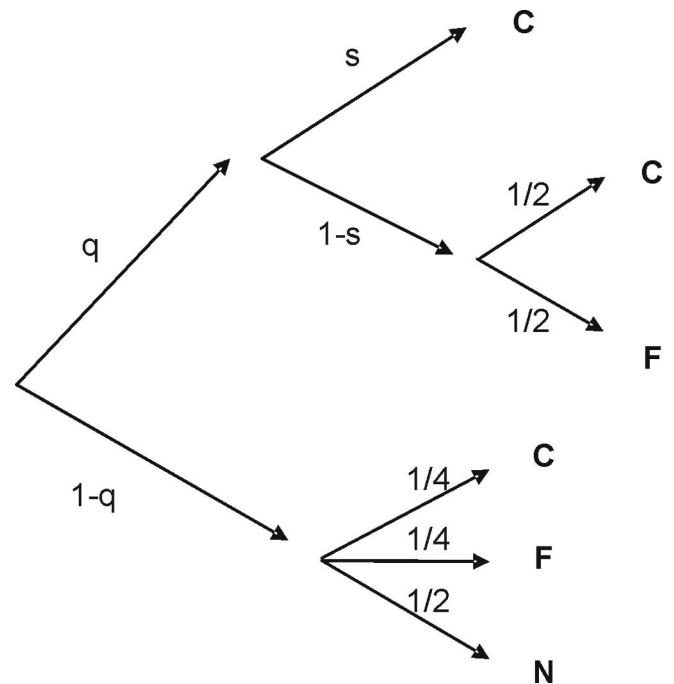


**Figure 2.** Probabilistic model for the analysis of data. For details, see Materials and Methods.

infrared eye-tracking system, operating at a sampling frequency of 120 Hz. Trials were aborted if gaze left the fixation window. After achieving fixation, a masked sequence of square-wave, 4.5° diameter, circular gratings appeared at the center of gaze. The spatial frequency of the gratings was 0.9 cycles/degree. Sequences were 1 s in duration. Masks (250 ms) appeared immediately before and subsequent to each sequence. Masks were dynamic patterns of random dots whose colors and luminances were selected at random from all colors and luminances present in the gratings.

After the presentation of the sequences, subjects were presented with an array of four gratings and indicated which of these four gratings had appeared in the sequence. The four choices could differ from one another either in orientation (with color held constant, selected at random from the four possible colors) or color (with orientation held constant, selected at random from the four possible orientations) (Fig. 1*C*).

*Subjects and training procedure.* Twelve naive observers participated in the experiment. Before starting, subjects were trained for a single session in which they ran 300 trials at three frequencies: 250 ms/grating (the slowest frequency presented in the experiment), 93.75 ms/grating, and 62.5 ms/grating. The slowest frequency was used to verify that they understood instructions. The two faster frequencies were used to give them practice performing the task as the frequency of presentation varied, without training them on the remaining frequencies that would be used in the main task.

*Quantitave analysis.* A probabilistic analysis technique known as multinomial modeling (Riefer and Batchelder, 1988) was used to estimate each observer's temporal resolution for color, orientation, and conjunctions. Multinomial modeling provides a means of estimating the probabilities of correctly perceiving a sensory event on the basis of patterns of errors in a forced-choice paradigm. For each of the two four-alternative forced choices (4AFCs), let $C$ be the choice with the correct conjunction, let $F$ be the choice that misconjoins features that were present in the sequence, and let $N$ represent the two choices containing features not present in the sequence. Consider the decision tree model appearing in Figure 2. Let parameter $q$ be the probability of perceiving the feature that varied across the four choices appearing in the 4AFC question (either color or orientation). The observer fails to perceive this feature with probability $(1 - q)$. On these trials, the observer is at chance selecting among the four choices (lower branch of the tree). Let parameter $s$ be the conditional probability of perceiving the conjunction given

that the observer correctly perceived the feature. Thus, if the feature was perceived, then the correct conjunction is perceived with probability $s$, resulting in a correct response ($C$). With probability $1 - s$ the conjunction is not perceived, and the observer is at chance selecting among the correct choice ($C$) and the conjunction error ($F$). The probability of selecting the correct choice is $P(C) = q \cdot s + q \cdot (1 - s) \cdot 1/2 + (1 - q) \cdot 1/4$, the probability of selecting the misconjunction is $P(F) = q \cdot (1 - s) \cdot 1/2 + (1 - q) \cdot 1/4$, and the probability of selecting a feature that was not present in the sequence is $P(N) = (1 - q) \cdot 1/2$. Suppose the observer selected the correct choice (i.e., $C$) on $n_1$ trials, the misconjunction ($F$) on $n_2$ trials, and features not present in the sequence ($N$) on $n_3$ trials. Then, the likelihood function is $L(q,s) = (n_1 + n_2 + n_3)! / (n_1! n_2! n_3!) \cdot P(C)^{n_1} \cdot P(F)^{n_2} \cdot P(N)^{n_3}$ with $P(C)$, $P(F)$, and $P(N)$.

A unique maximum exists for this type of function (Hu and Batchelder, 1994), and thus the parameters $q$ and $s$ that maximize the likelihood function are well defined and could be derived for each subject, at each presentation frequency. After determining maximum likelihood values for $q$ and $s$, Monte Carlo simulations were used to assess goodness-of-fit for each frequency presented for each individual subject. The model accounted for 97% of the variance of the data, on average. Confidence intervals for the estimated probabilities were computed using the Agresti–Coull confidence interval for a binomial proportion (Brown et al., 2001). Ninety-five percent confidence intervals for the critical rates were computed using the bootstrap method described in Wichmann and Hill (2001b).

*Test for independence of color and orientation.* Subjects viewed sequences like those that appeared in the main experiment and performed a modified 4AFC task with the four choices corresponding to the entries in a 2 × 2 contingency table in which each choice was correct or incorrect in color and correct or incorrect in orientation. For example, upon viewing a sequence composed of red/45° and green/135°, the subject might be asked to choose among the following: red-45° (C+/O+), red-vertical (C+/O−), blue-45° (C−/O+) and blue-vertical (C−/O−). Unlike the 4AFC task performed in the main experiment, identifying the individual features present in the sequence was sufficient to report (C+/O+). Therefore, this test measured the subjects' ability to report the individual features present in the sequence, not their ability to report how those features were conjoined. Eight subjects were tested with at least 48 stimulus sequences at each frequency (on average 69 repetitions, SEM 7). Additional control experiments are described in the supplemental information (available at www.jneurosci.org as supplemental material).

## Results

While maintaining fixation, 12 naive observers viewed sequences of colored, oriented gratings at the center of gaze. Each sequence was defined by two alternating colors (either red alternating with green, or yellow with gray) and two alternating orientations (either 0° alternating with 90° or 45° with 135°) (Fig. 1A). Sequences were constructed to ensure that at high presentation rates, all sequences would be indistinguishable from one another. This was accomplished for color by selecting pairs of colors that shared a common midpoint in the three-dimensional space of cone photoreceptor excitations. At high temporal frequencies, such color pairs merge to form the identical mixture color, rendering the observer unable to distinguish which colors appeared within a given sequence. To abolish orientation information at high temporal frequencies, we shifted the spatial phase of each successive grating of a given orientation by 180° (Fig. 1B). By varying presentation frequency across trials, we determined the maximum frequency at which observers could report the colors, the orientations, and the conjunctions present in a given sequence.

These critical frequencies were assessed by analyzing the pattern of correct and incorrect responses made to a 4AFC that was presented to the observer at the end of each trial (Fig. 1C). The four alternatives differed from one another either in orientation or in color. In each case, one of the four alternatives was a grating

that had appeared in the sequence. Another alternative was a misconjunction of a color and an orientation that were present in the sequence. The other two alternatives included features not present in the sequence. At sufficiently low frequencies, observers should reliably select the correct grating. At frequencies exceeding the temporal resolution for a given feature, observers should be at chance in selecting among the four alternatives. The key question is whether features are perceived at frequencies beyond the temporal resolution for conjunctions. If conjunctions are processed with lower temporal resolution than features, there should be a range of frequencies over which observers reliably reject features not in the sequence but fail to reject misconjunctions. Figure 3A shows the pattern of performance for one subject (MD) when asked to choose among gratings differing in orientation (Fig. 1C, left) or color (Fig. 1C, right). Panels are arranged according to the duration of each grating appearing in the sequence, from 6.25 ms on the left to 250 ms on the right. Within each panel, the percentage of trials on which the observer correctly reported the conjunction appears in the top left quadrant. Misconjunctions and feature errors appear in the top right quadrant and bottom half of each panel, respectively. These percentages are also indicated by gray scale, with lower percentages indicated by darker grays.

At the slowest presentation rate (frame duration of 250 ms), the subject almost always reported the correct conjunction (97% of all trials). At the fastest presentation rate (frame duration of 6.25 ms), the subject often reported features not present in the sequence and was at chance selecting among the possible choices. At a grating duration of 25 ms, the observer rarely made errors in judging orientation (1% of all trials) or color (16.7% of all trials) but often made errors in judging how these features were conjoined. Conjunction errors were still common at a frame duration of 31.25 ms, despite the fact that the subject rarely made feature errors.

We quantified the temporal resolution of each observer for orientation, color, and conjunction using a probabilistic model analysis (Riefer and Batchelder, 1988; Hu and Batchelder, 1994) of the pattern of errors the observer made at each frequency (see *Methods*). This analysis provides the maximum likelihood estimate of the probability that the observer perceived the colors, the orientations, and the conjunctions present within each sequence, at each frequency. The results of this analysis, for two observers, including subject MD, appear in Figure 3B. The abscissa in each plot indicates grating duration. The ordinate indicates the probability that the observer perceived orientations (green), colors (blue), and conjunctions (red). Curves are maximum likelihood Weibull function fits to the psychometric data (Wichmann and Hill, 2001a). As illustrated in the top panel, subject MD almost always selected the correct conjunction at frame durations of 250 ms. This is reflected in high probabilities of perceiving of both features (blue and green points) and conjunctions (red points). At intermediate frequencies, orientation and color were perceived more reliably than were conjunctions. For example, at a frame duration of 25 ms, MD perceived orientation with probability 0.98 and color with probability 0.67 but perceived conjunctions with probability 0.26. At high frequencies, performance was near chance, as indicated by near zero probabilities of perceiving orientations, colors, and conjunctions. Psychometric functions for a second subject, NJ, appear in the panel below that of MD.

The conjunction present at any moment could, in principle, be known the instant its constituent features are known. If so, assuming the probabilities of identifying the two features are statistically independent, the probability of detecting a conjunction,
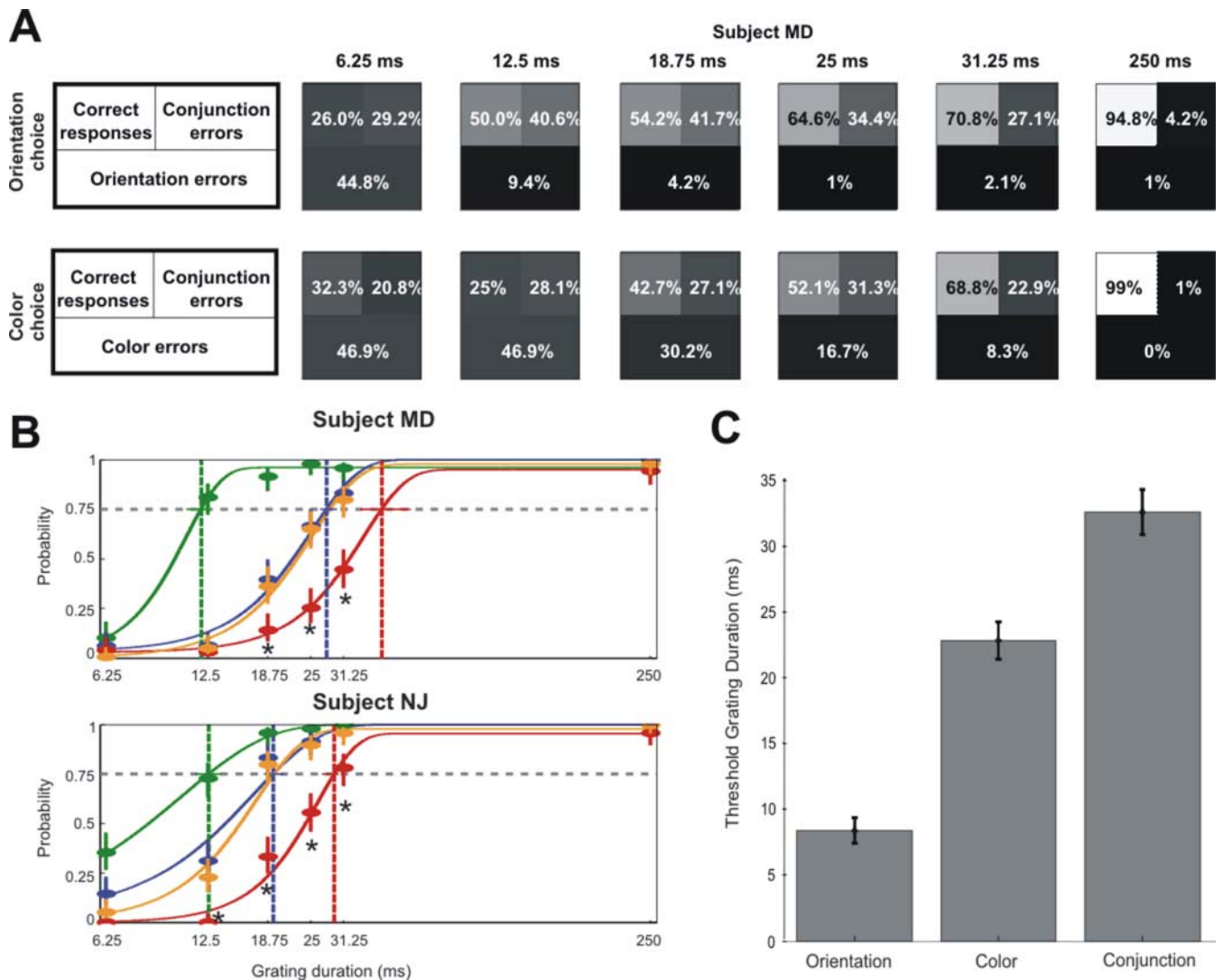
**Figure 3.** Results. *A*, Responses to the four-alternative forced choices illustrated in Figure 1*C* as a function of grating duration for observer MD. Within each panel, the top left square indicates the percentage of correct responses, the top right square indicates the percentage of conjunction errors, and the bottom quadrant indicates the percentage of feature errors. These percentages are also indicated by the gray scale in each panel, with lower percentages indicated by darker colors. This is shown for the orientation choice and for the color choice. *B*, Probability of discriminating the colors, orientations, and conjunctions present within a sequence, as a function of grating duration for two observers (top plot for subject MD and bottom plot for subject NJ). Plots show maximum likelihood estimates of the probabilities of discriminating the orientations (green points), the colors (blue points), and the conjunctions (red points) present in a sequence, as inferred from error patterns on the four-alternative forced-choice task. Vertical lines centered on each point indicate 95% confidence intervals on each estimate, as computed using Agresti–Coull interval estimation (Brown et al., 2001). Curves are Weibull function fits (Wichmann and Hill, 2001a) to the estimated probabilities for orientation (green), color (blue), and conjunction (red). The orange line is the product of the probabilities of perceiving the two features, which is the probability of discriminating conjunctions if features were independently discriminated and instantaneously integrated. Asterisks indicate points where $p_{color} * p_{orientation} > p_{conjunction}$, $p < 0.05$. Vertical dashed lines indicate the 75% threshold frequencies for orientation, color, and conjunction. Confidence intervals for thresholds were determined by bootstrap (Wichmann and Hill, 2001b) and are indicated by horizontal bars falling on the 0.75 line. *C*, Threshold grating durations for orientations, colors, and conjunctions, averaged over 12 observers. The average threshold for orientation is 8.4 ± 1.0 ms (mean ± SEM), for color is 22.9 ± 1.4 ms, and for conjunctions is 32.4 ± 1.6 ms.

$p_{conjunction}$, should be equal to the product of the probabilities of detecting the features, $p_{orientation}$ and $p_{color}$. Alternatively, if features must first be discriminated and then, through an additional computational step, integrated, there should be a range of frequencies over which $p_{conjunction}$ is significantly lower than the product of $p_{orientation}$ and $p_{color}$. To examine this, we compared $p_{conjunction}$ with the product of $p_{orientation}$ and $p_{color}$ across all tested temporal frequencies. The product of $p_{orientation}$ and $p_{color}$ is indicated by the orange curve appearing in each panel of Figure 3*B*. For both subjects, $p_{conjunction}$ was significantly less than the product of $p_{orientation}$ and $p_{color}$ across a range of grating durations, indicated by asterisks in Figure 3*B*, according to the McNemar test for paired proportions (Newcombe, 1998). This pattern held across all 12 observers. For every observer, there was at least one

tested frequency at which conjunctions were perceived significantly less reliably than would be expected if integration were instantaneous (see the supplemental figure, available at www.j-neurosci.org as supplemental material). In no observer was there any frequency at which the probability of perceiving a conjunction significantly exceeded the product of the probabilities of perceiving the two features. As an additional means of testing the significance of the differences in judging features and conjunctions, we performed a two-way ANOVA across subjects with judgment type (orientation, color, and conjunction) and frame duration as factors. Main effects for both factors and for interactions were highly significant ($p < 0.0001$).

To quantify temporal resolution for features and conjunctions across observers, we estimated each observer's temporal

resolution for orientation, color, and conjunction on the basis of Weibull function fits to each observer's estimated probabilities. Critical frequencies were defined to be the grating duration at which the Weibull function crossed a threshold of 0.75 (Fig. 3B). Confidence intervals for the thresholds (Wichmann and Hill, 2001b) for the subjects MD and NJ are shown in Figure 3B, indicated by horizontal colored bars falling on the 0.75 threshold line. The average temporal resolution for features and conjunctions across all subjects appears in Figure 3C. The average time required to perceive orientation was 8.4 ± 1.0 ms (mean ± SEM). Color was significantly slower, at 22.9 ± 1.4 ms. Conjunctions were slower still, at 32.4 ± 1.6 ms. As an additional means of estimating the temporal resolution across subjects, we averaged the Weibull fit parameters across subjects and determined where the Weibull function with these average parameters crossed the 75% threshold. These threshold estimates were in close agreement with the averaged thresholds (orientation, 8.8 ms; color, 22.0 ms; conjunction, 33.2 ms).

Note that the comparison of the product of $p_{orientation}$ and $p_{color}$ with $p_{conjunction}$ assumes that the probabilities of identifying the two features are statistically independent. We tested this assumption in a control experiment (see Materials and Methods). Briefly, subjects viewed sequences like those described previously (Fig. 1A,B) but performed a modified 4AFC task. The four choices in the 4AFC corresponded to the entries in a 2 × 2 contingency table for the variables color and orientation. One of the choices was correct in color and orientation (C+/O+). Another was correct in color but incorrect in orientation (C+/O−). A third choice was correct in orientation but incorrect in color (C−/O+). The fourth choice was incorrect in both color and orientation (C−/O−). Statistical dependence was assessed using Fisher's exact test, corrected for multiple comparisons. None of the subjects showed significant interactions between color and orientation at any of the tested frequencies. One can never infer from a failure to detect a significant effect that none exists. We therefore looked at the difference between percentage correct (C+/O+) and the product of (C+) with (O+) across all conditions. We found that the mean difference was small (~1.5%), indicating that any deviation from independence could not possibly account for our main effect. Therefore our assumption of independence is justified, and we conclude that the observed difference between $p_{orientation} \cdot p_{color}$ and $p_{conjunction}$ reflects the cost of integration.

## Discussion

The present findings agree with previous psychophysical studies that have found that visual stimuli can be processed rapidly (Holcombe and Cavanagh, 2001; Rousselet et al., 2002). In the most directly related study, Holcombe and Cavanagh found that observers could reliably report conjunctions of orientation and color presented for, on average, ~ 25 ms. Our estimate, 32.6 ms, is broadly comparable, although slightly longer. The difference may be attributable to differences between the two studies, including differences in the luminance contrast of the gratings (~55% in the previous study, ~30% in our case), differences in masking, and individual differences across subjects. The key advance in the present study is that by quantifying the temporal resolutions for features and conjunctions, the present study demonstrates that the temporal resolution for conjunctions is not simply a reflection of the time required to process the slower feature. For every subject, features were available to perception before being combined into a perceptual whole.

Temporal resolutions for individual features also likely de-

pend on stimulus parameters. For example, temporal resolution for orientation may depend on the luminance contrast of the gratings. Single-unit recording studies have found that neuronal responses become more sluggish as luminance contrast is reduced (Gawne et al., 1996). Temporal resolution for orientation might therefore have been lower had we used gratings of lower luminance contrast. A related point is that the chromaticity of each grating was, by design, approximately flat. That is, chromatic contrast was set as close to zero as possible, within the color resolution of our CRT. Although there are, to our knowledge, no single-unit recording studies showing that neuronal responses become more sluggish as chromatic contrast is reduced, it may be that our estimate of temporal resolution for color would have been higher if our stimuli had included greater chromatic contrast. The main conclusion to be taken from present experiments is thus not the particular estimates of temporal resolution for color, orientation and conjunctions, but rather the finding that, in every subject tested, temporal resolution was measurably lower for conjunctions than for individual features.

Although additional insight into the mechanisms of feature integration will require neuronal recording studies, the present findings do have implications for existing models of feature integration. One model that has been proposed to accomplish integration is that the features of a given object are linked by tagging them with a synchronous 40 Hz oscillation. If this mechanism integrates color and orientation in our paradigm, synchronization must be able to emerge very quickly, as integration occurred within ~32 ms, ~23 ms of which was required for features to reach perceptual threshold.

Next, consider models in which features are bound together by spatial attention (Treisman and Gelade, 1980; Wolfe et al., 1989). There is considerable psychophysical and neuropsychological evidence that attention does play a role in feature binding (Treisman and Schmidt, 1982; Wolfe and Bennett, 1997). Damage to the parietal lobes, which are thought to play a part in the allocation of attention, can result in illusory conjunctions (Friedman-Hill et al., 1995). Also, search for a target defined by a conjunction of features is often less efficient than search for targets defined by a feature singleton (Wolfe et al., 1989), as would be expected if spatial attention must be applied to each item to bind the features of each object together before identification. In our study only one stimulus appeared at any point in time, and that stimulus always appeared at an attended location. Therefore, this study provides measures of the temporal resolutions for features and for conjunctions in the presence of spatial attention. It remains to be seen whether the temporal resolution for features and conjunctions, as measured in the present paradigm, will change when attention is drawn away from the grating sequence. If attention does play a role in integration within our paradigm, withdrawal of spatial attention should result in a pronounced reduction in temporal resolution for conjunctions.

A third model that could potentially account for the observed temporal cost of feature integration is that time may be required for the visual system to accumulate evidence for a given conjunction, based on responses of feature selective neurons. The distribution of response times and accuracy in sensory discrimination tasks are well described by models in which a decision variable is treated as a diffusion process, and sensory detection or discrimination is achieved when this variable crosses a threshold (Ratcliff, 1980; Smith et al., 2004; Palmer et al., 2005). One might imagine that when a given grating appears, the resulting activity of color- and orientation-selective neurons is read out by a downstream decision process, leading to accumulation of evidence for that

conjunction. In the present experiments, the total amount of time that a given conjunction appeared was fixed (at 500 ms), regardless of presentation frequency. Thus, reduced accuracy in reporting a conjunction would have to be attributed to a reduced rate of evidence accumulation. It seems reasonable to assume that at frequencies approaching threshold for individual features, the relevant feature-selective neurons would respond weakly, resulting in a reduced rate of evidence accumulation.

A fourth possibility is that the difference between the time required to process the slowest feature and the time required to process the conjunction reflects the time required to complete a feedforward process of feature integration. It is unknown whether neurons selective for a particular conjunction give rise to the perception of that conjunction, but the temporal constraints imposed by the present results are in line with the temporal characteristics of feedforward neuronal signals. The response latencies of neurons in macaque area V1 are ~40–50 ms. After leaving V1, visual information traveling down the temporal lobe passes through four areas (V2, V4, TEO and TE). In area TE, response latencies are ~80–100 ms (Wallis and Rolls, 1997). Thus, each stage of processing requires on the order of 10 ms before sending a feedforward volley of activity to the next stage of processing. This is consistent with the idea that the 10 ms required for integration is achieved when the feedforward transmission of action potentials passes through a single cortical area. Also consistent with this estimate are direct estimates of the time required for neurons to begin to discriminate one stimulus from another, which range from 5–25 ms after the neuronal response is initiated (Oram and Perrett, 1992; Wallis and Rolls, 1997). The observation that feature integration required ~10 ms supports the view that integration, at least in the case of color and orientation, could be achieved as a natural consequence of the feedforward progression of neuronal responses as they advance through visual areas with progressively more complex receptive fields.

## References

Brown LD, Cai TT, DasGupta A (2001) Interval estimation for a binomial proportion. Stat Science 16:101–133.

Friedman-Hill SR, Robertson LC, Treisman A (1995) Parietal contributions to visual feature binding: evidence from a patient with bilateral lesions. Science 265:853–855.

Gawne TJ, Kjaer TW, Richmond BJ (1996) Latency: another potential code for feature binding in striate cortex. J Neurophysiol 76:1356–1360.

Ghose GM, Maunsell J (1999) Specialized representation in visual cortex: a role for binding? Neuron 24:79–80.

Gross CG, Bender DB, Rocha-Miranda CE (1969) Visual receptive fields of neurons in inferotemporal cortex of the monkey. Science 166:1303–1306.

Holcombe AO, Cavanagh P (2001) Early binding of feature pairs for visual perception. Nat Neurosci 4:127–128.

Hu X, Batchelder WH (1994) The statistical analysis of general processing tree models with the EM algorithm. Psychometrika 59:21–47.

Hubel DH, Wiesel TN (1965) Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. J Neurophysiol 28:229–289.

Hubel DH, Wiesel TN (1968) Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. J Physiol (Lond) 195:215–243.

Maunsell JH, Newsome WT (1987) Visual processing in monkey extrastriate cortex. Annu Rev Neurosci 10:363–401.

Newcombe RG (1998) Improved confidence intervals for the difference between binomial proportions based on pair data. Stat Med 17:2635–2650.

Oram MW, Perrett DI (1992) Time course of neural responses discriminating different views of the face and head. J Neurophysiol 68:70–84.

Palmer J, Huk AC, Shadlen MN (2005) The effect of stimulus strength on the speed and accuracy of a perceptual decision. J Vis 5:376–404.

Ratcliff R (1980) A note on modeling accumulation of information when the rate of accumulation changes over time. J Math Psychol 21:178–184.

Riefer DM, Batchelder WH (1988) Multinomial modeling and the measurement of cognitive processes. Psychol Rev 95:318–339.

Riesenhuber M, Poggio T (1999) Are cortical models really bound by the "binding problem"? Neuron 24:87–93.

Rousselet GA, Fabre-Thorpe M, Thorpe SJ (2002) Parallel processing in high-level categorization of natural images. Nat Neurosci 5:629–630.

Shadlen MN, Movshon JA (1999) Synchrony unbound: a critical evaluation of the temporal binding hypothesis. Neuron 24:67–77.

Sincich LC, Horton JC (2005) The circuitry of V1 and V2: integration of color, form, and motion. Annu Rev Neurosci 28:303–326.

Smith PL, Ratcliff R, Wolfgang BJ (2004) Attention orienting and the time course of perceptual decisions: response time distributions with masked and unmasked displays. Vision Res 44:1297–1320.

Treisman AM, Gelade G (1980) A feature-integration theory of attention. Cognit Psychol 12:97–136.

Treisman AM, Schmidt H (1982) Illusory conjunctions in the perception of objects. Cognit Psychol 14:107–141.

Ungerleider LG, Mishkin M (1982) Two cortical visual systems. In: Analysis of visual behavior (Ingle DJ, Goodale MA, Mansfield RJW, eds), pp 288–321. Cambridge, MA: MIT.

Wallis G, Rolls ET (1997) Invariant face and object recognition in the visual system. Prog Neurobiol 51:167–194.

Wichmann FA, Hill NJ (2001a) The psychometric function: I. Fitting, sampling, and goodness of fit. Percept Psychophys 63:1293–1313.

Wichmann FA, Hill NJ (2001b) The psychometric function: II. Bootstrap-based confidence intervals and sampling. Percept Psychophys 63:1314–1329.

Wolfe JM, Bennett SC (1997) Preattentive object files: shapeless bundles of basic features. Vision Res 37:25–43.

Wolfe JM, Cave KR, Franzel SL (1989) Guided search: an alternative to the feature integration model for visual search. J Exp Psychol Hum Percept Perform 15:419–433.

Zeki S, Shipp S (1989) Modular connections between areas V2 and V4 of macaque monkey visual cortex. Eur J Neurosci 1:494–506.