The visual perception of the movements of objects and ourselves is clearly of vital importance in interpreting and predicting our world. Catching a ball, crossing the road, driving a car, filling a kettle and innumerable other tasks rely on these abilities – those who have lost them through brain damage give us a fleeting insight into their value. There appears to be a consensus amongst most researchers that there are at least two stages to interpreting the pattern of changing luminance on our retina that forms the input to motion perception. The first stage is a local extraction of the ‘amounts’ of motion in each possible direction at each point in the image. Secondly, these local estimates need to be organized so as to eliminate noise, form surfaces, segment differing areas, and so on. Indeed this type of problem is not confined to visual motion processing and so one hopes that the ideas reviewed here will have implications for many other domains and applications.

Stages of motion processing

During the 1980s, consideration of the ‘first’ stage of motion processing led to a plethora of computational models, while, though varying in their details and emphasis, also have many similarities. Many of these models can be considered as extensions of the so-called ‘Reichardt detector’ where signals from two points on the retina are compared, with a delay introduced to one of the signals. Such was the success of these models (and the difficulty in telling them apart) that, despite the intervening years, these models still hold sway among most researchers. This review is mute with regard to the various merits of these models, and will simply refer to this first stage as the extraction of local velocity through appropriately tuned filters.

The second stage of motion processing requires that the signals arising from such units be organized – our visual world seems to consist of objects, surfaces and the like, rather than simple local motions. To do this we must have some way of combining the signals that ‘belong’ together and separating the ones that do not. In essence this is akin to the Gestalt notion of ‘common fate’. The 1990s have seen a deluge of activity on this front at both the psychophysical and physiological level (for an earlier review see Ref. 4) culminating in several computational models of our perceptual behaviour and of the units that underpin these perceptions. In this review, we shall briefly outline some of the psychophysical and physiological data that gave rise to these models, we shall describe the basic outline of the two types of model, and then attempt to draw some commonalities between them, whilst suggesting how their differences might lead to new approaches and understanding of our perception of dynamic events.
Motion transparency

The major problem faced by the ‘second’ stage of motion perception is to combine the signals that ought to go together whilst separating those that ought not to. One particularly clear version of this problem is the condition termed motion transparency. Here, the motions arising from two surfaces or objects are physically overlapping so that one cannot segment the surfaces from position-based cues alone (Fig. 1).

Unlike most laboratory stimuli, real objects normally move across cluttered environments that could, by virtue of their own movements, our eye/head movements, and noise, also contribute potential motion signals to the analysis of the scene. Hence, the need to remove, or at least segment, the additional signals is paramount. Motion transparency therefore provides a valuable test case for potential models of motion integration.

Box 1. Transparent motion and its aftereffect

One of the interesting aspects of transparent motion is its aftereffect (MAE) (Refs a,b). If an observer looks at a transparent motion stimulus, say patterns of random dots moving in orthogonal directions, two segregated surfaces are perceived. However, when the motion is stopped the MAE is unidirectional. The direction is the inverse of the vector sum of both inducing patterns (Ref. c). The direction is the inverse of the vector sum of both inducing patterns (Ref. c). This observation has challenged classic theories of the MAE (Ref. c) as well as models of motion detection.

However, the MAE of transparent motion is not always an integrative phenomenon, that is, unidirectional. Under certain conditions there are two aftereffect directions visible, moving opposite to the adapting directions. For example, if two transparent planes are segregated by introducing depth from disparity the MAEs are contingent on this disparity and therefore a transparent motion aftereffect is perceived (for a review, see Ref. f). Speed and the type of test pattern are also factors as to whether integration or segregation in the aftereffect occurs. Adaptation to transparent motion, where one pattern is moving at a low speed and the other at a high speed, can result in a transparent MAE. The type of test pattern, however, must have dynamic as well as static components (e.g. noise on a detuned TV superimposed on a checkerboard, respectively) (Ref. g). In this case the physically static parts of the test pattern show the aftereffect of the slow adaptation speed whereas the dynamic parts move opposite to the fast adaptation speed. The implication is that the visual system has access to two different sources (adapted neural substrates for each direction) for the aftereffect and might integrate this information after adaptation, with a strong dependency on the type of test pattern.

References

Fig. 1. Motion transparency stimuli. (A) Two sets of randomly positioned dots moving in different directions and at different speeds are seen to move smoothly through one another. This does not appear to be the case. Several studies report that the presence of a second pattern negatively affects the detectability of the first pattern, suggesting the presence of inhibitory connections. Thus, two motions can be represented – although at a reduced level for each one. The primary visual cortex (V1) is the first area where directionally selective cells are found, and these are often identified with the first stage of the motion processing schemes outlined above (for reviews, see Refs 19–21). The second stage appears to involve the middle temporal area (MT, also known as V5), which has been heavily implicated in the processing of motion information (Ref. 22). When presented with a transparent motion stimulus, single cells in area V1 respond to their preferred direction of motion just as if one motion stimulus had been presented alone. MT cells, on the other hand, tend to give a reduced response to their preferred direction when it is presented as one of two surfaces (Ref. 23). Note that these MT cells merely reduce their rate of response – they do not fall silent. The suppression caused by this second transparent motion has a broad direction tuning, which is in line with psychophysical results (Ref. 24). Thus, the population of MT cells can signal two directions of motion, but with reduced strength.

Whether MT can represent two constructed surfaces or merely represents the first stage of surface construction is still an open question (Ref. 25). The aftereffects observed with transparent motion can also help to inform questions about the cortical substrates of motion perception (see Box 1).
Beyond the random-dot pattern

The transparency seen with random-dot patterns is in contrast to the phenomenon seen when two identical sine-wave grating stimuli moving in opposite directions are superimposed. The gratings appear as a single flickering grating. Qian and colleagues showed that the crucial difference between these patterns is that the motions at a very local level are balanced for the grating pattern, whereas they are only balanced at a global level for the random-dot pattern. If one constructs a ‘random’ dot pattern in which for each dot in one direction there is another dot very close by moving in the opposite or orthogonal direction (Fig. 1B), then one merely sees flicker rather than the impression of two surfaces.

One of the problems faced by the second stage is that known as ‘the aperture problem’, which occurs locally whenever there is only one dimension to the image structure. Consider the case of two squares moving in opposite directions (Fig. 2A). The motion occurring in one small aperture, illustrated in a magnified portion just above the red square, could arise from various possible motions (some of which are illustrated by the arrows) all of which produce the same displacement of the line within the aperture. Such a stimulus produces activity in V1 cells, not only in detectors tuned for the ‘correct’ motion, but in a range of detectors whose speed and direction tuning are all consistent with the possible movements of this line. Hence it might be inappropriate
to think of a single local velocity being signalled at each location in space – instead there might be a multiplicity of ‘possible’ motions at each point in the image.

One dimension is that the ‘probabilistic’ representation exists for any one-dimensional stimulus the solution seems easy – call for another dimension. A two-dimensional pattern can be produced by simultaneously perceiving two one-dimensional patterns whose lines are at different orientations and are moving in different directions – often termed ‘plaids’ because of their ‘tartan’ appearance (Fig. 2B). Such patterns often appear to move coherently in a single direction as a single entity (3). In some circumstances, though, the two patterns appear to drift transparently through one another. When do we see coherence and when transparency? When we see a coherent motion in a single direction, what is this direction and how is it computed?

For our example of the square (Fig. 2A) the situation is not so simple. The information about the two dimensions needed to resolve the ‘aperture problem’ is actually displaced in space so that information must somehow propagate across this spatial dimension. This might not seem too difficult to solve until one considers the more complex situation of the overlapping squares. There are now several possible motions that could be combined in any arbitrary combination. Our visual systems appear to have little problem in this complex situation – can models of motion integration also produce correct answers?

We have already noted that cells early in the motion processing pathway suffer from the aperture problem. Concordant with this notion such cells produce two peaks in activity when presented with plaids of varying orientation (Fig. 2C), each peak corresponding to one component of the plaid that is moving in a direction optimal for this cell, thus earning these cells the shorthand name ‘component’ cells (20). Some cells, however, produce a single peak that corresponds to direction of the pattern – hence the shorthand ‘pattern’ cells (Fig. 2D). These pattern cells appear later in the system and around one third of the cells in area MT (V5) are of this kind (20). Such cells are consistent with our perception but merely tell us the ‘problem’ is ‘solved’ by this stage – they do not tell us how the problem is solved.

Thus these stimuli, and in particular, the complex motions provided by transparency, provide a rich test bed for the models. We can also compare the components of any model against the measured properties of the cells along the motion pathway. The models must aim to answer these three questions that such stimuli pose:

1. How can we represent two motions at the same location in space?
2. How can the locally ambiguous signals be disambiguated?
3. How are signals assigned to appropriate objects/surfaces?

Models of motion integration

Models that try to explain motion transparency come from different traditions. We are well aware that we cannot do justice to all the models that have been suggested in the past (e.g. Refs 30–37). Here we would like to examine two models of motion integration and segmentation. One comes from the filter selection tradition and the other uses a rather new approach, that has received a lot of recent attention (38) based on ‘Bayesian’ inference. The reason that we choose these models is that they try to relate their modelling to the knowledge we have about the behaviour of the neural substrates likely to be involved in the integration and segmentation processes.

1. A filter selection model

The model of Nowlan and Sejnowski (20) starts by extracting motion estimates in the conventional manner outlined above. These signals are then fed to two separate arrays, one calculating the ‘local velocity’ and one the ‘selection weights’ (see Fig. 3). The local velocity calculation is performed via a system of inhibitory connections between units tuned for different velocities, and as such has much in common with the ‘winner-takes-all’ algorithm except that the use of ‘soft maxima’ allows for either strong evidence for one particular motion or weaker evidence for multiple motions. The selection network uses a much coarser scale to compute a global
competition between detectors that have the same velocity preference, with the result that only areas with the best evidence for a given velocity remain active. This selection map is then used as a set of weighting functions that are multiplied with the local velocity estimates to derive the final velocity estimates. The actual algorithm for determining the pattern of interactions in the local velocity map and selection arrays was resolved via a learning procedure in which the network was trained upon a large variety of different stimuli, including plaids and transparent motions, as well as single objects. This selection process learns the value of discontinuities in the motion inputs, such as occurs when the signals around the discontinuity were consistent with a particular object motion. An example of the output of a selection unit to various inputs is illustrated in Fig. 3B. Note that whilst each of these examples contains rightward motion the strongest evidence for rightward motion is for the diamond.

The model accounts for transparency in random-dot patterns by having, at the level of the local velocity units, two sets of active units whose individual activity is less than that produced for a single surface. This is a consequence of the soft maxima (or normalization process) and is present in other models that have this feature and agrees well with both the psychophysical and physiological data. This model eventually produces estimates of motion only for regions where it is most confident; thus, the representation at this level is spatially patchy and can depend upon the area where one motion is represented compared with the other motion. This means that the model has trouble handling locally balanced motions. Normally, this might be thought of as a failing of the model; however, given our own inability to decipher the motion in such stimuli, this is a desirable failing.

The nature of the selection process means that signals at the corners of our square, or at the intersections of our plaid, are chosen. At these features have two dimensions they do not suffer the ambiguity of the aperture problem. However, as mentioned earlier, not all two-component patterns cohere; many appear as the transparent motion of two surfaces. Variations in the luminance of the intersections can alter the perception of the gratings between coherence and transparency.
Review

Snowden and Verstraten – Motion transparency

B Smoothing speed. (See text for further details.) (Adapted from Ref. 44.)

...the lines with arrows depict excitatory connections. Note that any two units are inhibitory if...

...that motion.

...of the ambiguous signals at the centre of the bar is most probable and thenceforth selects...

...motion from the ends of the bar to help calculate which...

...motions at the centre of the bar. The selection model calculates the correct local motion at the ends of the bar but calculate several ‘possible’ local...

...what different from the correct motion owing to the averaging-in of the ‘incorrect’ local...

...the differing models encode the movement of a bar moving up and to the right (light grey arrow). The smoothing-type algorithms start by calculating a single local motion at each location and then averages across these estimates (within some local area). Note that whilst such algorithms provide a global estimate for motion at each location this estimate is some...

...one of the bar moving up and to the right (light grey arrow). The smoothing-type algorithms start by calculating a single local motion at each location and then averages across these estimates (within some local area). Note that whilst such algorithms provide a global estimate for motion at each location this estimate is some...

...the selection model and Bayesian model both calculate the correct local motion at the ends of the bar but calculate several ‘possible’ local...

...the selection model and Bayesian model both calculate the correct local motion at the ends of the bar but calculate several ‘possible’ local...

...inhibitory links from detectors tuned for all...

...a Bayesian inference principle.

...the evidence for upward motion depends upon both the incoming signal and upon the prior belief about the likelihood of that signal. Where does this belief about the likelihood of a signal come from? Koechlin et al. suggest that this can be gathered from nearby units – if they are also sig...

...a Bayesian inference model

...the model of Koechlin et al. uses the idea of gating signals in order to remove spurious signals and disambiguate ambiguous ones. The key assumption of this model is that in systems of lateral excitatory and inhibitory weights implements a Bayesian inference principle.

...et al. suggests that this can be gathered from nearby units – if they are also sig...

...this can be gathered from nearby units – if they are also sig...

...the ideas suggested above (see Fig. 5A). The model also includes a distinction between coherence units driven by ‘on-centre’ fibre units (termed positive) and those driven by ‘off-centre’ units (termed negative). This is to reflect the possible influence of non-motion cues in the processes of segmen...

...of transparency as part of the training diet.

...the Bayesian approach? This model encompasses the ideas from smoothing-type algorithms and from selection processes. Let us illustrate with the exampl...

...the Bayesian approach? This model encompasses the ideas from smoothing-type algorithms and from selection processes. Let us illustrate with the exampl...

...the ideas suggested above (see Fig. 5A). The model also includes a distinction between coherence units driven by ‘on-centre’ fibre units (termed positive) and those driven by ‘off-centre’ units (terned negative). This is to reflect the possible influence of non-motion cues in the processes of segmen...

...is that the selection process chooses the intersections for high-contrast intersections (i.e. the low and high luminance ones) but chooses portions of the gratings away from the intersections for the low contrast ones. Two opposing factors are at work. The inter...

...trend in Cognitive Sciences – Vol. 3, No. 10, October 1999

...as a technique of value to understanding human perception – it seems hard to imagine that our own perception of motion underwent such a learning process.

...the evidence for upward motion depends upon both the incoming signal and upon the prior belief about the likelihood of that signal. Where does this belief about the likelihood of a signal come from? Koechlin et al. suggest that this can be gathered from nearby units – if they are also sig...

...this can be gathered from nearby units – if they are also sig...

...the ideas suggested above (see Fig. 5A). The model also includes a distinction between coherence units driven by ‘on-centre’ fibre units (termed positive) and those driven by ‘off-centre’ units (terned negative). This is to reflect the possible influence of non-motion cues in the processes of segmen...

...is that the selection process chooses the intersections for high-contrast intersections (i.e. the low and high luminance ones) but chooses portions of the gratings away from the intersections for the low contrast ones. Two opposing factors are at work. The inter...

...the ideas suggested above (see Fig. 5A). The model also includes a distinction between coherence units driven by ‘on-centre’ fibre units (termed positive) and those driven by ‘off-centre’ units (terned negative). This is to reflect the possible influence of non-motion cues in the processes of segmen...

...the ideas suggested above (see Fig. 5A). The model also includes a distinction between coherence units driven by ‘on-centre’ fibre units (termed positive) and those driven by ‘off-centre’ units (terned negative). This is to reflect the possible influence of non-motion cues in the processes of segmen...

...the ideas suggested above (see Fig. 5A). The model also includes a distinction between coherence units driven by ‘on-centre’ fibre units (termed positive) and those driven by ‘off-centre’ units (terned negative). This is to reflect the possible influence of non-motion cues in the processes of segmen...

...the ideas suggested above (see Fig. 5A). The model also includes a distinction between coherence units driven by ‘on-centre’ fibre units (termed positive) and those driven by ‘off-centre’ units (terned negative). This is to reflect the possible influence of non-motion cues in the processes of segmen...

...the ideas suggested above (see Fig. 5A). The model also includes a distinction between coherence units driven by ‘on-centre’ fibre units (termed positive) and those driven by ‘off-centre’ units (terned negative). This is to reflect the possible influence of non-motion cues in the processes of segmen...

...the ideas suggested above (see Fig. 5A). The model also includes a distinction between coherence units driven by ‘on-centre’ fibre units (termed positive) and those driven by ‘off-centre’ units (terned negative). This is to reflect the possible influence of non-motion cues in the processes of segmen...
The segregation of local motion estimates into particular surfaces could, in theory, take head of cues derived from other sources. One could gather all the estimates that show a particular depth (derived from, e.g., binocular disparity) and aid motion transparency by segmenting these estimates from those that do not share this characteristic. Hilliard and Brainard have provided a quantitative evaluation of this idea (Ref. 6). Sensitivity to motion was greater for a single surface than for a transparent one. If binocular disparity was introduced, so that the two transparent surfaces were at different depths, then this difference vanished. This suggests that the direction of motion is computed separately for each of the surfaces when they are separated in depth. Snowden and Roseman (Ref. 6) further show that performance on a motion coherence task is not disturbed by extra noise dots that have a very different binocular disparity, providing the colour bar has a transparent width to begin with (see also Refs 4 and 5). These results suggest a powerful role for binocular disparity in segmenting motion estimates. Indeed these visual behaviours coincide with the properties of cells in area MT where studies of motion transparency have revealed that the suppression caused by transparent motion in the cell’s non-preferred direction is reduced if this motion is presented at a different depth (Ref. 6).

Can other non-motion cues provide similar segmentation of the motion signal? One obvious candidate is colour. If ‘signal’ dots are defined in a different colour (or polarity) from the ‘noise’ dots, performance is dramatically enhanced compared with when they are the same colour (Ref. 7). However, Edwards and Badcock presented a seemingly contradictory result whereby addition of extra noise dots to a coherence task elevated thresholds even if these noise dots were of a different colour from the ones that defined the task (Ref. 8). Snowden and Edmunds suggest that the differently coloured elements might provide a highly salient ‘pop-out’ area that attracts attention to process this area (Ref. 6). In other motion coherence tasks, however, colour segregation appears. Cells of area MT can show some changes in response when the signal dots are made salient by the use of colour (Ref. 9).

However, the change is more complicated than considered independently. The Bayesian model calculates the overall plausibility of motion in visual noise (Ref. 10). The change in performance is modest compared with the behavioural effect and thresholds in the colour-select condition are still affected by stimulus parameters known normally to affect motion perception. Croner and Albright also suggest the possibility that attentional mechanisms might underlie these changes in the cell’s response (Ref. 11).

Thus, it appears that while non-motion cues may aid in creating motion transparency, not all cases will do. Defining which cues are helpful is an important next step, but perhaps not as important as defining why some cues are helpful and others not.

References


red arrow). The signals in the centre of the bar suffer the aper- tion problem. Smoothing algorithms take all the inputs and average them (or select one based on a winner-takes-all strategy), whereas the selection algorithms merely select the signals from the ends of the bar where the motion is unam- biguous. The Bayesian model corrects all the motions along the bar (including the ambiguous ones) and then uses the unambiguous ones to infer which of the ambiguous ones are ‘correct’. Indeed the Bayesian model can be made to work either like a winner-takes-all strategy or a selection strategy, depending upon the contrast polarity of the plaids with which they are presented. Perhaps more worrying is the finding within the simulation that the filters units in the direction of the pattern motion appear to respond more vigorously than those in the component direction(s). If these filter units are to be analogous to V1 cells then this behaviour is inconsistent with the known neurophysiology.

The Bayesian model certainly successfully answers our first two questions. It can represent two surfaces, and it can dis-ambi-guate ambiguous ones. Where it might have an advan- tage over the filter selection model is that it also allows for grouping behaviour of signals through its excitatory lateral connections. The signal can smooth along regions where there is ambiguous motion information, but will not cross to an area where the local velocity signals are not supportive of this ‘hypothesis’. The model also allows for the temporal integration of motion signals, because the response to locally ambiguous signals depends upon the stimulus history. As time passes, more and more locations will have had a coherently moving dot at that location, which enables a bias in favour of this direction to propagate.
Review

Snowden and Verstraten: Motion transparency

Outstanding questions

- Can motion transparency be used to identify specific brain areas in humans?
- As might be predicted from single-cell recording in non-human primates, recent evidence has suggested that dynamic noise (containing random local motions) strongly activates area V1 with lesser activation in area MT, whereas coherent motion is a more effective stimulus for area MT (Ref. 47).

Conclusions

The difficulties posed by transparent motions, and the apparent case in which our visual systems deal with this, have prompted new computational approaches that allow units to 'gaze' the incoming signals, either via their value with respect to other signals (filter model) or with respect to their consistency with nearby units (Bayesian model). Thus, spurious noisy signals can be removed and areas with poor signals enhanced while maintaining the other priority of segmenting differing motions. These models also allow for 'flexibility' in the way they deal with inputs. Incoming data could be strong and clean, or could be weak and noisy. It might therefore be possible to use differing strategies to segment such images. Each of these forms an own visual system seems to support the notion that we do have such flexibility.

The approaches described concern the perception of motion. It is clear that similar problems exist in all domains of sensory processing and, indeed, in biological systems in general. Any system faced with noisy and/or incomplete information must decide whether new information is useful and how it interacts with existing information. The approaches emphasized here may prove to have implications well beyond their original setting.

Acknowledgments

F.V. is supported by a Research Fellowship of the Royal Netherlands Academy of Arts and Sciences (KNAW). F.A.J.V. is supported by a Research Fellowship of the Royal Netherlands Academy of Arts and Sciences (KNAW).

References


Conclusions

The difficulties posed by transparent motions, and the apparent case in which our visual systems deal with this, have prompted new computational approaches that allow units to ‘gaze’ the incoming signals, either via their value with respect to other signals (filter model) or with respect to their consistency with nearby units (Bayesian model). Thus, spurious noisy signals can be removed and areas with poor signals enhanced while maintaining the other priority of segmenting differing motions. These models also allow for ‘flexibility’ in the way they deal with inputs. Incoming data could be strong and clean, or could be weak and noisy. It might therefore be possible to use differing strategies to segment such images. Each of these forms an own visual system seems to support the notion that we do have such flexibility.

The approaches described concern the perception of motion. It is clear that similar problems exist in all domains of sensory processing and, indeed, in biological systems in general. Any system faced with noisy and/or incomplete information must decide whether new information is useful and how it interacts with existing information. The approaches emphasized here may prove to have implications well beyond their original setting.
The neuroethology of primate vocal communication: substrates for the evolution of speech

Asif A. Ghazanfar and Marc D. Hauser

In this article, we review behavioral and neurobiological studies of the perception and use of species-specific vocalizations by non-human primates. At the behavioral level, primate vocal perception shares many features with speech perception by humans. These features include a left-hemisphere bias towards conspecific vocalizations, the use of temporal features for identifying different calls, and the use of calls to refer to objects and events in the environment. The putative neural bases for some of these behaviors have been revealed by recent studies of the primate auditory and prefrontal cortices. These studies also suggest homologies with the human language circuitry. Thus, a synthesis of cognitive, ethological and neurobiological approaches to primate vocal behavior is likely to yield the richest understanding of the neural bases of speech perception, and might also shed light on the evolutionary precursors to language.

The species-specific vocalizations of non-human primates are crucial for their social interactions, reproductive success and survival, and some have argued that speech has played a similar role in human history. Investigating the perception and social use of vocalizations in extant non-human primates might be the most direct route to understanding the substrates underlying the evolution of speech and language. It follows, therefore, that investigating the neural processes underlying the vocal behavior of primates might yield important insights into the neurobiology of speech.

Neuroethological research has already added much to our understanding of how natural selection shapes brain-design for complex behaviors such as echolocation in bats, song learning in birds, and mate-choice in frogs. Likewise, in the visual behavior of primates, faces are highly relevant stimuli in their day-to-day social interactions and specialized regions of the temporal lobes appear to be dedicated to face processing. Based on the consistency with which behavioral adaptations are mediated by specialized neural systems in the animal kingdom, we hypothesize that the design of the...