Motion Capture Unmasked

Jeffrey B. Mulligan and Helen A. Cunningham
NASA Ames Research Center
MS 239-3
Moffett Field, CA, 94035

[FIRST SLIDE ON (sum of parts)]

The ultimate goal of our work is to understand what determines what percept will be evoked by an arbitrary stimulus. What we mean by the statement that "the whole is rarely the sum of the parts" will hopefully become clear in a moment. First we want to demonstrate an example where the whole IS the sum of the parts. [NEXT SLIDE (random dot equation), ROLL TAPE]

In the first section of the video, we see two random dot patterns drifting in opposite directions. When a new sequence of patterns is generated by adding the luminances of the individual patterns, two transparent sheets are seen, producing the sensation of motion in BOTH directions at each point in the field. [NEXT SLIDE] We can represent the categories of the possible motion percepts using a Venn diagram. In this slide we see ovals enclosing stimuli which are seen to move to the left [POINT] and to the right [POINT]; the intersection region [POINT] represents transparent stimuli where both motions are seen like what we just saw on the tape.

This situation of adding leftward motion to rightward motion and ending up with both leftward and rightward motion is certainly not the general result. [ROLL TAPE] It is well known, for example, that if two GRATINGS of equal spatial frequencies are superimposed (while drifting) that the result will be a counterphase flickering grating, for which NO motion is seen, only flicker. This illustrates that we have to revise our idea of how to categorize motion percepts by adding another category for flicker or noise. [NEXT SLIDE] The bubble for noise is shown to overlap with the bubbles for leftward and rightward motion. The next section of the tape illustrates how the sensation of flicker or "twinkle" can coexist with directional percepts.

Another stimulus we would like to understand is the one investigated by Ramachandran and Cavanagh, which led them to coin the term "motion capture." In this stimulus, temporally uncorrelated random dots were superimposed on a low frequency grating which underwent an abrupt displacement. Although the dots by themselves would not generate a directional percept, when presented with the grating they appeared to be "stuck" onto the grating and moved with it.

When a square wave is displaced by a quarter cycle, the higher Fourier components are displaced by more than a complete cycle, resulting in an ambiguity if they are considered singly. When the components are presented singly, each component will appear to move in the direction requiring the smallest phase shift, which alternates for the odd-harmonic components. Nevertheless, when the square wave is shown, the ambiguity is resolved in favor of the coherent percept.

It has been shown by Adelson, however, that if the fundamental is removed, resulting in a "fluted" square wave, then the visual system does not adopt the interpretation resulting in a coherent percept, but rather follows the direction of the third harmonic. We can image that the direction might reverse yet again if we were to remove the third harmonic. [ROLL TAPE?] I noticed that this issue will be addressed in the following paper, which made me glad to realize that we were not the only people that thought there was something here that deserved further study.

It is obvious that motion capture can be abolished by reducing the contrast of the capturing grating sufficiently, and that the motion reversal of the fluted square wave should be eliminated by adding some contrast of the fundamental, or reducing the contrast of the third harmonic. None of the current theories of motion perception, however, are sufficiently detailed to make a quantitative prediction of just HOW MUCH contrast might be required.

In the hopes of gaining an understanding of the spatial frequency selectivity of the motion process, we adopted a motion cancellation paradigm, in which the motions of gratings of various spatial frequencies were opposed, and the contrasts varied to achieve a directional null. This technique had been used successfully by Cavanagh to determine contrast equivalence between chromatic and achromatic gratings, although he only looked at equal spatial frequencies. The details of our methods are summarized on the NEXT SLIDE. [Sample stimuli here???]

In order to illustrate what our stimuli looked like, I would like to first review a well known image, first created at the Craik lab in Cambridge, which Beau Watson has informed me is referred to in Cambridge as the $64,000 picture. [NEXT SLIDE] The image is schematized on the slide and we see the image on the video tape. The image consists of a frequency sweep grating, the contrast of which is ramped along the grating bars. Each point in the image represents a frequency-contrast pair that might be used in an experiment to measure contrast sensitivity. This figure is neat because it allows observers to "see their own contrast sensitivity function" as the boundary of the region where the pattern can be seen. We can also visualize the spatio-temporal CSF by having the pattern drift or flicker.

We have reviewed this image because we have created a similar figure to illustrate our experiment. Just as in the $64,000 picture, each point in the image represents a different stimulus. It should be emphasized that the images we will see now were NOT stimuli in the experiment, but are merely shown to let you have an idea of the range of phenomena.

To create our figure, we simply took two drifting $64,000 pictures, rotated one 180 degrees and added it to the other. Thus the two frequency sweeps proceed in opposite directions accross the screen, and the contrasts are traded off as we move from top to bottom. As is indicated in the slide, the display can be segregated into a number of regions where distinct categories of percept are observed. At the upper and lower edges, where one of the two contrasts is small, just the motion of the higher-contrast component is seen. In the center, where the contrasts and spatial frequencies are similar, flicker is seen. And when the contrasts are similar but the spatial frequencies are sufficiently different, the gratings segregate and transparency is seen.

Before performing the nulling experiments, we first measured thresholds for the frequencies in the study. Thresholds for our three subjects are shown in the NEXT SLIDE. In the nulling experiements, the sum of the contrasts of the two gratings (which we called the "total contrast") was held constant; the total contrast values were determined for each subject on an individual basis. For each grating pair, we measured the null contrast ratio at total contrasts of twice the sum of the thresholds, and twenty times.

The NEXT SLIDE shows the results of the nulling experiments. We have plotted the results in a space where the two axes represent the linear contrast of each of the two components divided by the threshold; that is, all contrasts are expressed in multiples of threshold. The diagonal line with a slope of one represents stimuli where the two gratings are at equal threshold multiples. For the low contrast condition, all the data points are clustered around this line [POINT]. In the high contrast conditions, however, this is not the case. When the components are separated in frequency by two octaves, all subjects showed a preference for the direction carried by the lower of the two frequencies, indicated by the fact that the points fall above the diagonal. The strongest effects were observed for the pairing of 6 cycles per degree and 1.5 cycles per degree, which have been plotted as circles [POINT]; here the contrast ratios were around 5. Similar effects were observed for the pairing of 1.5 cycles per degree and 0.375 cycles per degree, plotted with squares [POINT].

The results were quite different when the gratings were separated by four octaves, however; these points are indicated by triangles [POINT]. In this case the subjects consistently reported that the stimulus appeared transparent. In the face of this inherently ambiguous situation, the subjects were instructed to respond on the basis of which direction appeared "stronger." The points for two of the subjects (the investigators) fall below the diagonal, indicating a bias favoring the high frequency component. Our naive subject, on the other hand, falls above the diagonal. We believe that this is a result of the fact that the appearance of the stimulus changed during the presentation; during the first couple hundred milliseconds, the rapidly moving low frequency stimulus dominated, but was quickly suppressed by the high frequency stimulus. We believe that subject SS may have been consistently basing his responses on his first impression, unlike the other two subjects.

These results suggest that the key question is not "which frequencies dominate which other frequencies," but rather "what determines whether a stimulus is seen as a single motion or as transparent." The two examples of motion capture, Ramachandran and Cavanagh's grating plus dots, and Adelson's fluted square wave, are both situations where a transparent percept is consistent with the data, and yet is not seen.

The phenomenon of transparency illustrates the inadequacy of a single channel motion model. A direction-selective unit, exemplified by the Reichardt detector shown in the NEXT SLIDE, yields an expected response of zero for both superimposed motions and flicker. This can be remedied by postulating a parallel channel model [NEXT SLIDE]. Multiresolution models have been proposed for the motion system before, but to our knowledge these have been motivated by observations of the physiology, and have not been suggested out of necessity to explain psychophysical observations. *(On thin ice here; is the MAE frequency specific???)* We can exploit such a parallel channel model to explain the transparency percept by postulating that transparency is seen when two channels signal different velocities; if the components of the stimulus are combined within a single channel, then the net result will be seen as flicker.

With this framework in mind, we now ask how might the different channels interact to determine the null points in the motion cancellation experiments. We will offer two hypotheses: the first we call the "veto model;" it is inspired by the term "motion capture." In the veto model, there is asymmetric inhibition between the low-frequency channel and higher channels, resulting in the suppression of signals from higher frequency channels. An alternative model [NEXT SLIDE] is that the judgement is made simply on the basis of a weighted sum of the outputs of all the channels.