

# Evidence for multi-functional interactions in early visual motion processing

**Martin A. Giese**

*Center for Biological and Computational Learning,  
Massachusetts Institute of Technology,  
Cambridge, MA 02139-4307, USA  
E-mail: [giese@ai.mit.edu](mailto:giese@ai.mit.edu)*

(To appear in *Trends in Neurosciences*)

16. May 1999

**Running title:** Multi-functional interactions in early motion processing

**Keywords:** apparent motion, object recognition, multi-modal integration, P / M pathway, perceptual channel, visual cortex, neural coding

### ***M. Giese: Multi-functional interactions in early motion processing***

---

Scientific work on vision has often been based on the assumption that the visual system can be decomposed into multiple modules that are relatively independent, and which can be studied separately, both experimentally and theoretically. Also, early vision is usually regarded as being independent from other sensory modalities. In psychophysics, early vision has been conceptualized as a multi-channel system with multiple (information) processing streams for extracting different stimulus attributes, like shape, motion, color, and depth<sup>1,2</sup>.

Despite this historical emphasis on modularity, there is increasing evidence from neurophysiology and psychophysics that a strict modularization of visual function is not possible. This article focuses on two psychophysical experiments that provide such evidence. The first experiment by Yantis and Nakama<sup>3</sup> demonstrates that the recognition of stationary, non-moving characters is impaired along paths of perceived motion. This suggests that motion perception influences the perception of stationary shape. A second experiment by Sekuler *et al.*<sup>4</sup> demonstrates an influence of acoustic stimuli on the perceptual organization of visual motion. Such results raise the question if information processing in early vision can really be modularized.

## *Modularization of visual function*

The concept of a modularization of visual function has emerged from different results in neurophysiology and psychophysics. It has also been important in computational approaches to vision.

Early clinical observations of distinct visual impairments that are induced by local lesions in certain regions of the brain led neurologists to hypothesize an association of visual functions with specific local neuroanatomical structures<sup>5,6</sup>. The fundamental work by Hubel and Wiesel<sup>7,8</sup> demonstrated that neurons in the primary visual cortex are characterized by well-defined tuning properties, and that the complexity of these tuning properties increases along the visual pathway. This has been taken as evidence for a theoretical conception that modularizes early vision in terms of a hierarchy of feature detectors, or filters. The complexity of the tuning properties in higher visual areas was accounted for by the assumption that complex feature detectors are formed by combining the output signals from simpler feature detectors on earlier levels<sup>7</sup>. Beyond hierarchical organization, the concept of parallel pathways has been important for the modularization of visual function. On the basis of systematic lesion studies, it was proposed to distinguish between a “where pathway”, in which mainly the spatial characteristics of objects are processed, and a “what pathway”, that is responsible for color vision and object recognition<sup>10</sup>. A similar decomposition, that is based on neuroanatomical, physiological and psychophysical

### ***M. Giese: Multi-functional interactions in early motion processing***

---

results<sup>1,2,6,9</sup>, distinguishes an “M pathway” (including the magnocellular regions of the lateral geniculate nucleus), and a “P pathway” (that passes through the parvocellular layers of the lateral geniculate nucleus). The M pathway was associated with the processing of object motion, whereas the P pathway was assumed to be responsible for the processing of color and object shape. The extent to which the two pathways are functionally and anatomically separated has been the issue of extended scientific discussions. It seems now that the separation is less strict than was originally claimed<sup>9,11,12,13,14</sup>.

A modularization of the visual function has been a successful approach in computational approaches to vision<sup>15</sup>. The perception process was decomposed into separate computational problems, like “motion correspondence”, or “shape from motion”. Such problems could then be formalized mathematically, such that a solution with suitable (computer) algorithms was possible. Later work in computational theory and computer vision has focused on the introduction of interactions between different computational functions in complex artificial vision systems<sup>16,17</sup>.

The emphasis on modularization in neurophysiology and computational theories has also influenced psychophysics. Mostpart of the experimental and theoretical research has focused on the study of *individual* visual functions. Psychophysicists have attempted to isolate different perceptual (information processing) channels, e.g. by demonstrating selective adaptation. During experiments on individual functions, potential interactions with other functions were minimized through the choice of adequate stimulus sets, and by

balancing the experimental designs with respect to other, presumably uninteresting functional dimensions. This has led to a valid characterization of visual processing under stimulus conditions where individual functions, or channels, can be regarded as being independent from each other.

### ***Challenging experimental evidence***

Some recent psychophysical experiments have studied the interactions between different visual functions that were previously often assumed to be independent, and which were typically assigned to different anatomical pathways and perceptual channels. One experiment by Yantis and Nakama<sup>3</sup> investigated an interaction between the perception of apparent motion and the recognition of stationary non-moving characters. The stimulus combines an ambiguous apparent motion pattern with an object-recognition task (see Fig. 1C). The motion stimulus, which is sketched in Fig. 1A, consists of four dots that are positioned in the corners of an imaginary square. During the first stimulus frame, the two dots in one pair of diagonally opposite corners of the square (marked by white in the figure) are illuminated, and during the second frame, the other pair of dots (which is marked by black in the figure). This two-frame sequence is periodically repeated. The stimulus induces two alternative, mutually exclusive apparent-motion percepts<sup>18</sup>, which are indicated by the gray arrows in Fig. 1A. Motion is perceived either along the horizontal paths between the dots (h), or along the vertical paths (v). Never horizontal and vertical motion are perceived

### ***M. Giese: Multi-functional interactions in early motion processing***

---

at the same time. This implies that the motion percept is bistable. Detailed quantitative experimental and theoretical work has shown that the dynamics of percept formation can be captured by a bistable dynamical system that shows hysteresis<sup>19,20</sup>. This hysteresis property was exploited by Yantis and Nakama to establish a horizontal or vertical percept at the beginning of the experimental trials. They first occluded one pair of dots, so that the initial motion percept was uniquely horizontal (or vertical) (Fig. 1B). When the occluder is then slowly removed, the initial percept persists because of the hysteresis property of the perceptual dynamics.

After the initial motion percept had been established, a letter was presented on one of the four possible motion paths (Fig. 1C). The subjects had to discriminate between two different letters indicating their response by pressing a key. The average reaction time until the key press was measured. It was significantly longer (772 ms versus 723 ms) when the letter appeared on the path of the perceived motion, compared to when it appeared in the path of the non-perceived motion .

This result demonstrates that the recognition of the character is influenced by the perceived motion-path. There are different possible explanations for this effect. A control experiment, demonstrates that the detection of a small gap in a ring (Landolt figure) is also impaired along the perceived apparent-motion path. This suggests that there might exist an interaction between motion perception and the perception of stationary patterns. Such interactions between the perception of stationarity and apparent motion are also indicated

### ***M. Giese: Multi-functional interactions in early motion processing***

---

by an experiment by Hock *et al.*<sup>21</sup>. This experiment demonstrates bistability and hysteresis for the transition between the percepts of motion and stationarity (with flicker) for a local apparent-motion stimulus that consists only of a single pair of dots. This hysteresis effect can be accounted for by a mutual inhibition between the perception of motion and stationarity, which would also explain the effect observed by Yantis and Nakama. Interactions between the perception of motion and stationarity could result from top-down influences. They may also arise from couplings between motion-sensitive neural structures and structures that are sensitive stationary contrasts<sup>3</sup>. Such couplings may exist already early in the visual pathway, e.g. on the level of elementary motion detection. In any case, the interaction implies that the information processing streams for motion and stationary patterns are not strictly separated.

Another explanation for the results by Yantis and Nakama is an interference between object recognition and motion perception. Direct interactions between these two visual functions are also suggested by other experiments that demonstrate that high-level knowledge about the structure of three-dimensional objects, like human bodies, can influence whether rigid or non-rigid motion is perceived<sup>22</sup>. Such high-level information determines also if coherent apparent motion is perceived at all for image sequences that consist of strongly degraded images of human faces (Mooney faces)<sup>23</sup>. Results like these argue against a strict functional separation of the processing streams for object shape and motion, and they demonstrate that the perceptual organization within individual perceptual channels can be substantially

### ***M. Giese: Multi-functional interactions in early motion processing***

---

influenced by processes in other channels. The physiological basis of such modulations could be direct couplings between anatomical pathways, and top-down influences from higher multi-functional structures<sup>3</sup>.

In another experiment, Sekuler *et al.*<sup>4</sup> showed an influence of auditory stimuli on the perception of visual motion stimuli. An ambiguous motion stimulus was presented that consisted of two disks that moved in opposite directions towards each other, and crossed<sup>24,25</sup>. (The true motion of the discs is indicated by the black arrows in Fig. 1D, left.)

This stimulus can be perceived in two ways, that are indicated by the gray arrows in Fig. 1D (right): Either the discs seem to move continuously and cross (“streaming”), or they seem to bounce, and reverse directions. During a part of the trials, the coincidence of the two discs was paired with an acoustic click. The auditory stimulus led to a substantial increase of the probability that subjects saw the bouncing percept (57 versus 27 % without the click, on average). In a control experiment, in which the collision of the discs was paired with a short interruption of a continuous sinusoidal sound, no increased probability of the bouncing percept was observed. This rules out an explanation in terms of general unspecific attentional factors.

The experimental results by Sekuler *et al.* demonstrate that the perceptual organization of apparent motion is not independent of auditory perception. Indeed, interactions between visual and auditory cues are well-established for spatial localization tasks<sup>26,27</sup>. A classical example is the ventriloquist illusion: The location of an acoustic source can be

### ***M. Giese: Multi-functional interactions in early motion processing***

---

misperceived when an additional visual stimulus, suggesting another location of the source, is presented<sup>28</sup>. (This happens frequently in the cinema). Subcortical structures, like the colliculus superior, seem to be involved in the multi-sensory integration during cue localization<sup>27</sup>. The perceptual organization of motion that is relevant to the experiment by Sekuler *et al.* is typically assumed to take place in structures that are located early in the visual pathway (like area MT in the macaque). This means that the neural activity in such structures must be either be modulated by central multi-modal structures, or that multi-modal representations are present also on early levels of the visual pathway. Evidence for multi-modal neurons in the primary visual cortex of cats<sup>29</sup>, and in area V4 of monkeys<sup>30</sup> lends plausibility also to the second account.

There exists also psychophysical evidence for a reverse influence of visual stimuli on the perceptual organization of auditory stimuli. A classical example is the McGurk effect<sup>31</sup>: If an auditory syllable, e.g. /ba/, is paired with a visual stimulus showing a person that says the syllable /ga/, the subjects have the “intermediate” percept of the syllable /da/. The coupling between visual and auditory perception seems, therefore, to be bidirectional.

### ***Implications and unanswered questions***

This paper has reviewed a number of psychophysical experiments that demonstrate interactions between different functions in vision, and between vision and auditory perception. These interacting functions are usually assigned to different anatomical

### ***M. Giese: Multi-functional interactions in early motion processing***

---

pathways, and to different functional modules, or channels of sensory processing. The existence of interactions between different visual functions is consistent with anatomical studies that have revealed substantial connections between different anatomical pathways<sup>9</sup>. Interactions between early vision and other sensory modalities have not been studied very much. It seems at least possible that such interactions exist, and they would have a meaningful biological function.

The question arises if such interactions are relevant for perception under natural conditions. The attempt to build technical vision systems has led to the insight that their robustness can be considerably improved by introducing interactions between different computational functions (e.g. between texture detection, and optic-flow and depth estimation) already at early processing levels (where, for instance, correspondence problems are solved)<sup>16,17</sup>. It seems plausible that the visual system uses the same strategy to achieve robustness<sup>11</sup>. The presence of multi-functionality, even on early levels of the visual pathway, is likely also a basis for avoiding the “binding problem”. This problem arises in the classical conception of separate parallel processing channels because the information in different channels must later be recombined, and assigned in the right way to different objects in the visual field<sup>32</sup>.

The experiments, and theoretical considerations in this article suggest a conception of the visual system which is not based on a strict modularization of function. One possibility is to assume that different functional modules or channels are coupled, such that processes in one channel can modulate the processing in another channel<sup>2</sup>. This interpretation faces,

### ***M. Giese: Multi-functional interactions in early motion processing***

---

however, the conceptual problem to delineate different modules or channels when their functional properties are overlapping. Another conception, that avoids this difficulty, assumes that in each structure of the visual system many different functional dimensions are represented at the same time, but with different degrees of accuracy<sup>11</sup>.

The idea of such multi-functional neural representations has been criticized based on the argument that neural representations with many functional dimensions require an unrealistically high amount of neural hardware, and are characterized by very slow learning<sup>32</sup>. This problem is well-known in the theory of neural networks<sup>33</sup> as “curse of dimensionality”: The number of neurons that is required to encode a multi-dimensional feature space, assuming a fixed approximation error of the representation, increases exponentially with the dimensionality of the feature space. Results in statistical learning theory show, however, that very high-dimensional feature spaces can be encoded by relatively small neural networks with good learning properties when the precision of the encoding is controlled<sup>34</sup>. Such a control of precision seems also to be present in the visual cortex: For example, the precision with which the retinal position is encoded in the primary visual cortex (areas V1 and V2) seems not to be retained in the motion-sensitive area MT. This is indicated by the size of the receptive fields that are much larger in area MT than in the primary visual cortex<sup>35</sup>. Similarly, non-dominant attributes, like color in the M pathway, may be encoded with a strongly reduced accuracy. In this way, they could still take part in the perceptual organization process without leading the discussed dimensionality problem.

## ***Acknowledgements***

The author thanks L. Spillmann, H.S. Hock, T. Poggio, G. Geiger, K.P. Hoffmann, A. Benali, and R. Leonhardt for helpful comments. The authors research was supported by the Deutsche Forschungsgemeinschaft.

## ***References***

- 1 Livingstone, M. and Hubel, D.H.** (1988) *Science* 240, 740-749
- 2 Shapley, R.** (1990) *Annu. Rev. Psychol.* 41, 635-658
- 3 Yantis, S. and Nakama, T.** (1998) *Nat. Neurosci.* 1, 508-512
- 4 Sekuler, R., Sekuler, A.B. and Lau, R.** (1997) *Nature* 385, 308
- 5 Holmes, G.** (1945) *Proc. R. Soc. Lond. Ser. B*, 41, 348-361
- 6 Zeki, S.** (1993) *A Vision of the Brain*, Blackwell
- 7 Hubel, D.H. and Wiesel, T. N.** (1962) *J. Physiol.* 160, 106-154
- 8 Hubel, D.H. and Wiesel, T.N.** (1977) *Proc. R. Soc. Lond. Ser. B*, 198, 1-59
- 9 Felleman, D.J. and van Essen, D.C.** (1991) *Cereb. Cortex* 1, 1-47
- 10 Mishkin, M., Ungerleider, L. G. and Macko, K.A.** (1983) *Trends Neurosci.* 6, 414-417
- 11 DeYoe, E.A. and van Essen, D.C.** (1988) *Trends Neurosci.* 11, 219-226
- 12 Schiller, P.H. and Logothetis, N.K.** (1990) *Trends Neurosci.* 13, 392-398
- 13 Merigan, W.H. and Maunsell, J.H.R.** (1993) *Ann. Rev. Neurosci.* 16, 369-402
- 14 Gegenfurtner, K.R. and Hawken, M.J.** (1996) *Trends Neurosci.* 19, 394-401
- 15 Marr, D.** (1982) *Vision*, Freeman and Comp.
- 16 Poggio, T., Gamble, E.B. and Little, J.J.** (1988) *Science*, 242, 437-440
- 17 Aloimonos, J.Y. and Shulman, D.** (1989) *Integration of Visual Modules*, Academic Press
- 18 von Schiller, P.** (1933) *Psycholog. Forschg.* 17, 179-214
- 19 Hock, H.S., Kelso, J.A.S. and Schöner, G.** (1993) *J. Exp. Psych.: Hum. Perc. Perf.* 19, 63-18
- 20 Giese, M.A.** (1999) *Dynamic Neural Field Theory of Motion Perception*, Kluwer
- 21 Hock, H.S., Kogan, C. and Espinoza, G.** (1997) *Percept. Psychophys.* 59, 1077-1088
- 22 Sinha, P. and Poggio, T.** (1996) *Nature*, 384, 460-463
- 23 Ramachandran, V.S., Armel, C, Forster, C. and Stoddard, R.** (1998) *Nature* 395, 852-853

- 24 Metzger, W. (1934) *Psycholog. Forschg.* 19, 1-49
- 25 Michotte, A. (1963) *The Perception of Causality*. Basic Books
- 26 Lehwald, J. and Ehrenstein, W. (1996) *Exp. Brain Res.* 108, 473-485
- 27 Stein, B.E. and Meredith, M.A. (1993) *The Merging of the Senses*, MIT Press
- 28 Driver, J. and Spence, C. (1998) *Proc. R. Soc. Lond. Ser. B*, 353, 1319-1331
- 29 Morrel, F. (1972) *Nature*, 238, 44-46
- 30 Haenny, P.E., Mausell, J.H.R. and Schiller, P.H. (1988) *Exp. Brain Res.* 69, 245-259
- 31 McGurk, H. and MacDonald, J. (1976) *Nature*, 264, 746-748
- 32 von der Malsburg, C. (1995) *Curr. Opp. Neurobiol.* 5, 520-526
- 33 Girosi, F., Jones, M. and Poggio, T. (1995) *Neural Comp.* 7, 219-269
- 34 Vapnik, V.N. (1998) *Statistical Learning Theory*, Wiley
- 35 Gatass, R. and Gross, C.G. (1981) *J. Neurophys.* 46, 621-638

### ***Figure Caption***

**Fig.1: Stimuli used in the psychophysical experiments:** (A) The left panel illustrates the ambiguous apparent-motion stimulus (motion quartet). The dot pair that is marked by white is illuminated during the first stimulus frame. Then the dot pair which are indicated in black is illuminated in the second frame. This sequence is periodically repeated. For adequate spatio-temporal parameters, a percept of apparent motion occurs (indicated by the gray arrows in the panels in the middle and right). The percept is bistable. Subjects see either motion along the horizontal paths (h), or motion along the vertical paths (v). Both motion directions are never perceived at the same time. (B) When one pair of dots is occluded the percept becomes unique (vertical, for example). (C) The stimulus by Yantis and Nakama<sup>3</sup> combines the motion quartet with an object-recognition task. A letter is presented for a short time interval on one of the four possible motion paths (which are indicated by the

***M. Giese: Multi-functional interactions in early motion processing***

---

gray lines). **(D)** The ambiguous apparent motion stimulus that was used by Sekuler *et al.*<sup>4</sup> consists of two disks that are moving in opposite directions. (The physical motion of the discs is indicated by the black arrows, left in the figure.) The stimulus can be perceived as “streaming motion”, where the discs move continuously and cross, or as “bouncing motion”, where the discs seem to bounce, and change motion direction. (The perceived motion paths are indicated by the gray arrows in the right part of the figure.)

