PERCEPTUAL PROCESS OF FILLING-IN FOR DYNAMIC TEXTURES

M. Yokota* and Y. Yokota**

* Dept. of Information and culture, Nagoya Bunri University, Inazawa Aichi, Japan
** Faculty of Engineering, Gifu University, Gifu, Japan

yokota@nagoya-bunri.ac.jp

Abstract: Filling-in is a phenomenon by which a small figure (a filling-in target), which is presented in an observer's peripheral vision surrounded by a dynamic texture, is perceived to be invaded by the surrounding pattern and to disappear within a few seconds. Filling-in is regarded as a key to understanding the manner of effective visual information processing in human vision. Some characteristics of filling-in have been reported. Filling-in time has been adopted as a criterion of filling-in facilitation in many reports. From this perspective, we have proposed a model of the filling-in process that represents filling-in facilitation using filling-in target distinguishability from the surroundings (EMBEC 2002). The present study measures filling-in time for dynamic textures, surrounding a filling-in target. Those textures have various spatio-temporal frequencies. Furthermore, we estimated spatio-temporal frequency sensitivity of human vision at the retinal eccentricity where the filling-in target is projected. These results were applied to the filling-in process model. Results suggest that filling-in is facilitated when the LGN M-channel sensitivity is high; it is inhibited when that of the LGN P-channel is great.

Introduction

Perceptual filling-in is a famous illusion whereby a small object, i.e. a filling-in target, presented in peripheral human vision appears to become filled-in by its surrounding texture within a few seconds: it thereby becomes invisible. Visual information captured in the retina is putatively eliminated by such filling-in if it has less priority. Therefore, analyses of perceptual filling-in are important for understanding some aspects of information processing in human vision.

For that reason, much attention has been paid to factors that induce perceptual filling-in. Although the influences of several textural attributes, including color, spatial density, and moving speed, on filling-in facilitation have been reported, the influence of the spatio-temporal power spectrum of dynamic texture on filling-in facilitation has not been investigated sufficiently [1–5]. Filling-in is inferred to occur in the primary visual cortex, in which properties are explained in spatio-temporal frequency characteristics. For that reason, clarifying spatio-temporal frequency characteristics of filling-in is important for investigating filling-in occurrence mechanism. From this point of view, we have proposed a model of the filling-in process to address the phenomenon that occurs when a filling-in target, which is surrounded by spatio-temporal frequency limited dynamic textures, is presented to an observer's peripheral vision [6]. This model expresses filling-in target distinguishability from the surroundings. Filling-in facilitation is evaluated using that measure.

In this study, we measured the time to filling-in for various spatio-temporal frequencies of dynamic textures, and estimated spatio-temporal frequency sensitivity of human vision. These results were applied to the model, suggesting points of discussion for facilitation of filling-in.

Model of filling-in process [6]

We previously proposed a model that represents the filling-in process using a spatio-temporal power spectrum of the dynamic texture [6]. We introduced perceptual power, which represents visibility of the filling-in target, as distinguished from the surrounding dynamic texture. The perceptual power decreases from its initial value while the stimulus is presented, as depicted in Fig. 1.

The initial perceptual power P_{init} is defined as the inner product of the difference of the spatio-temporal power spectrum of the filling-in target and the dynamic texture, expressed as $W(f^t, f^s)$, and the squared spatio-temporal frequency sensitivity for dynamic texture in human vision, expressed as $S^2(f^t, f^s)$. It is written as

$$P_{init} = \iint W(f^t, f^s) S^2(f^t, f^s) df^s df^t, \qquad (1)$$

where f^s and f^t respectively represent spatial and

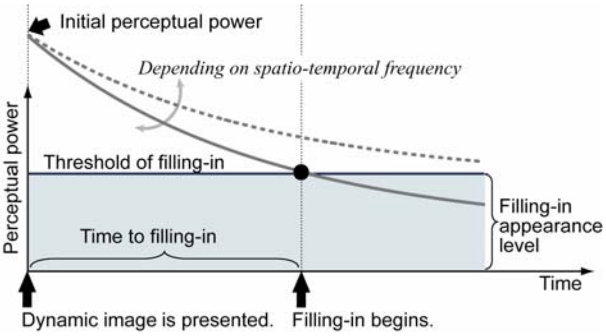

Figure 1: Diagram of proposed model with respect to the perceptual filling-in process.

ISSN: 1727-1983

temporal frequency. Furthermore, $W(f^t, f^s)$ is given as the spatio-the temporal power spectrum of the dynamic texture if the filling target is homogeneous: it has a power spectrum of zero.

Filling-in is considered to result from neural cell adaptation [1]. The number of adapted cells in a neural cell population can be expressed stochastically as an exponential function. Such stochastic models are also used for predicting machine breakdown. For that reason, perceptual power is assumed to decrease exponentially from the initial value during presentation of the dynamic texture. Perceptual power $P(t)$ at time t is expressed as

$$P(t) = P_{init} e^{-\lambda t}, \quad (2)$$

in which λ is the attenuation factor. The filling-in target becomes invisible, i.e., filling-in occurs when the perceptual power is less than a certain threshold value.

Experiment 1: Measurement of time to filling-in

(1) Method

Dynamic texture: A moving image (768 × 768 [pixel] × 32 [frame]) is generated. Its pixel values independently follow a Gaussian probability density distribution with mean μ = 127 and standard deviation $\sigma = 127/3$. Three-dimensional Fourier transformation transforms the moving texture into the spatio-temporal frequency domain. The dynamic texture expressed in the spatio-temporal frequency is band-limited in the ranges of [f^S_{low}, f^S_{up}] and [f^T_{low}, f^T_{up}]. Table 1 shows ranges used in the experiment. The scale of spatial frequency is expressed as cycles per degree of visual angle (cpd). Spatial and temporal frequencies are limited by narrow bands with ½ [octave] and broad bands with 2 [octave] width, which includes four narrow bands. Each frequency band is denoted as a logarithmic median frequency between [f^S_{low}, f^S_{up}] or [f^T_{low}, f^T_{up}], as narrow spatial (Ns), narrow temporal (Nt), broad spatial (Bs), or broad temporal (Bt). To equalize the power of dynamic texture, the band-limited dynamic texture expressed in spatio-temporal frequency is multiplied by

$$\frac{(f^S_{up} - f^S_{low})\sqrt{f^T_{up} - f^T_{low}}}{f^S_s \sqrt{f^T_s}}, \quad (3)$$

where f^S_s and f^T_s respectively denote the spatial and temporal Nyquist frequency. Three-dimensional Fourier transformation transforms the resultant dynamic textures expressed in spatio-temporal frequency into the spatio-temporal domain. Finally, the dynamic textures are quantized into integers of [0, 255].

Filling-in target and fixation point: How to maintain a subject's fixation is a serious problem in filling-in time measurement experiment because filling-in does not occur when a subject's fixation is not perfect; alternatively, time to filling-in is prolonged. We adopt an annular ring as a filling-in target because the eye tends to move toward the filling-in target spontaneously

Table 1: Limited spatial and temporal frequency

	Spatial frequency [cpd]				Temporal frequency [Hz]			
	f^S_{low}	-	f^S_{up}		f^T_{low}	-	f^T_{up}	
Narrow bands	0.41	-	0.56	Ns-0.48	0	-	0	Static
	0.58	-	0.82	Ns-0.69	2	-	2	Nt-2.0
	0.84	-	1.14	Ns-0.98	3	-	3	Nt-3.0
	1.16	-	1.64	Ns-1.38	4	-	5	Nt-4.5
	1.66	-	2.31	Ns-1.96	6	-	8	Nt-6.9
	2.33	-	3.28	Ns-2.76	9	-	11	Nt-10.0
	3.30	-	4.61	Ns-3.90	12	-	16	Nt-14.3
	4.63	-	6.55	Ns-5.51				
Broad	0.41	-	1.64	Bs-0.82	2	-	8	Bt-4.0
	1.66	-	6.55	Bs-3.30	4	-	16	Bt-8.0

when a filling-in target is located a certain position in the peripheral visual field. The inner and outer radii of the annular ring are, respectively, 250 and 300 pixels.

The filling-in target is painted a homogeneous pixel value 127. A black cross fixation point is attached at the center of every image with an included dynamic texture. Furthermore, we gradiently blurred the texture around the fixation point within a circle with 200-pixel radius to mitigate the transient release of fixation at the very moment when the dynamic texture appeared. Examples of a frame including a dynamic texture are shown in Fig. 2.

Apparatus and subjects: A dynamic texture is presented with a frame rate of 1000 / 30 [Hz] on a 21-inch CRT display with 1024 × 768 resolution. The dynamic texture, a series of 32 images, appears repeatedly. Subjects sat 35 [cm] from the CRT monitor facing it. The CRT luminance, $L(k)$, was set to be proportional to pixel values, k, using a video card function: $L(k) = 0.36\,k$ [cd/m^2]. The annular filling-in target inputs were 15–18 [deg] on the observer's retina.

Subjects are four 21–37-yr-old males. The subject views the display in his dominant eye; the other eye is covered by an eye-mask.

Task procedure: Initially, a homogeneous gray image attached with only a fixation point was presented to a subject. We instructed the subject to gaze at the fixation point and to press the computer mouse button when the subject felt that fixation was completed; thereby the dynamic texture began to move. Maintaining his gaze, he pressed the computer mouse

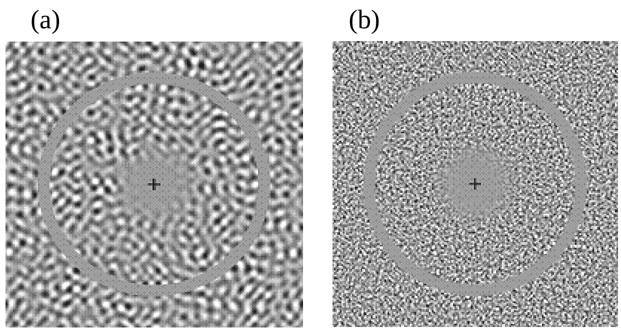

Figure 2: Examples of dynamic images in which spatial frequency is (a) narrow band and (b) broad band

button again when he felt the filling-in target disappear. The time interval from when the computer mouse button was first pressed until it was pressed a second time is regarded as the time to filling-in. Presentation of the dynamic texture was terminated within 20 [s] whether the subject pressed the computer mouse button or not. Dynamic textures were presented in random order. Each subject performed the above experiment three times for each dynamic texture.

(2) Results

The measured time to filling-in includes individual differences; respective subjects' means are 6.2, 5.3, 4.3, and 3.8 [s]. Moreover, data measured under uncompleted fixation must be included in measured data. Filling-in time tends to become longer or filling-in does not occur when fixation is not completed. For these reasons, measured times are divided by respective subjects' mean and multiply the mean time for all subjects, 4.9 [s]. Then we adopt the time averaged from 2nd to 9th lowest as the representative time to filling-in in order to exclude the data under uncomplete fixation condition.

Figure 3 shows the representative filling-in time for spatial and temporal frequency of dynamic texture.

Most filling-in times for spatial or temporal broad bands are longer than that for narrow bands included in the broad band.

Experiment 2: Estimation of spatio-temporal frequency sensitivity of human vision

(1) Method

Generally, sensitivity is estimated as an inverse of the sensory threshold. We measured the luminance sensory threshold for spatio-temporal frequency at the retinal eccentricity where filling-in targets are input.

Stimuli: The dynamic textures were generated as surrounding homogeneous gray and annular targets, to have band-limited texture for each narrow spatio-temporal frequency in Table 1. Furthermore, we prepared dynamic textures that have various luminance amplitudes, multiplying band-limited dynamic texture for each spatio-temporal frequency by $n = 0, 1, 2, \dots, 24$. Finally, the dynamic textures are quantized into integers of [0, 255]. Experimental conditions were identical to those of Experiment 1.

Procedure: Subjects are four 22–37-yr-old males, some of whom were participants in Experiment 1.

Initially, a homogeneous gray image attached only the fixation point was presented to a subject. We instructed the subject to gaze at the fixation point with his dominant eye. He pressed a computer mouse button when he felt that fixation was completed. Thereby, the dynamic texture was presented for 1 [s]. The subject reported whether the filling-in target was visible or invisible. We measured the visible luminance threshold for each spatio-temporal frequency. The spatio-temporal frequency of the dynamic texture was chosen in random order. The luminance amplitude was chosen in

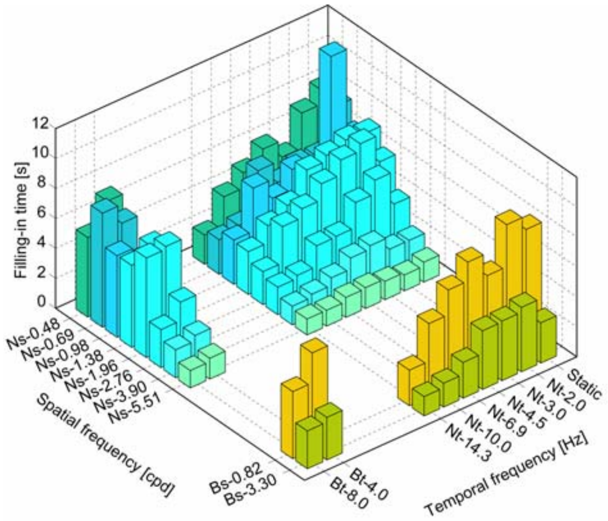

Figure 3: Normalized time to filling-in for spatial and temporal frequency.

dichotomy order. Each subject repeated the tasks three times for each dynamic texture.

(2) Results

The measured luminance sensitivity threshold has no great differences among subjects.

Spatio-temporal frequency sensitivity is estimated as the inverse of the measured luminance threshold (Fig. 4). The estimated spatio-temporal sensitivity is almost identical to the sensitivities in central or peripheral vision reported in other studies [7–9].

Applying the results to the model of filling-in process

Time course of perceptual power: In the following analysis, the highest spatial frequency band, $f^S = 5.51$ [cpd], was eliminated because the filling-in time and sensitivity are too low to neglect some experimental disturbances, e.g. subjects' reaction times in filling-in time measurement.

Results of Experiments 1 and 2 are applied to the proposed model of the filling-in process. Initial perceptual power is calculated with Eq. (1). The time course of perceptual power is estimated using Eq. (2).

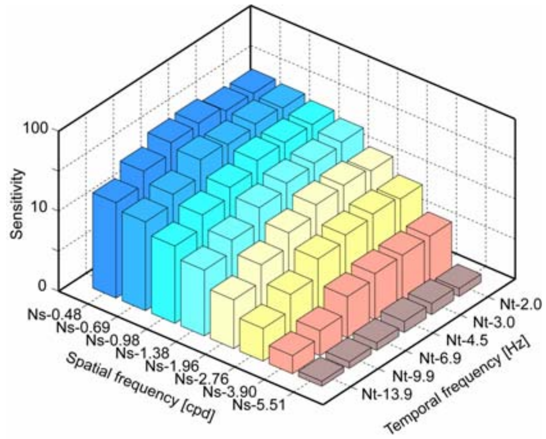

Figure 4: Estimated spatio-temporal frequency sensitivity of human vision.

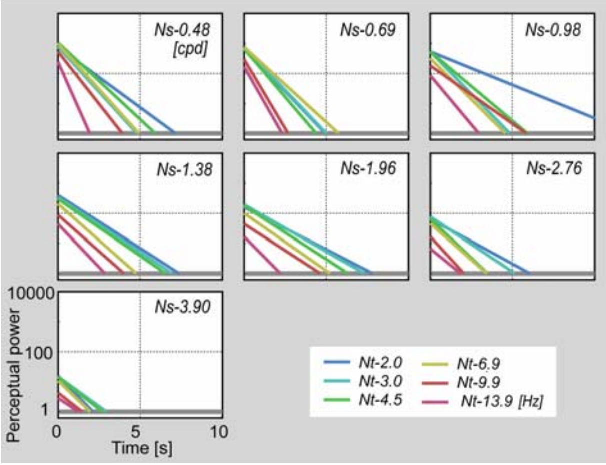

Figure 5: Estimated time course of perceptual power for narrow spatio-temporal frequency bands. Time courses for equal spatial frequency are shown in the same figure. Those for equal temporal frequency are drawn with the same colored lines.

Figure 5 shows estimated time courses of perceptual power for narrow spatio-temporal frequency bands of the dynamic textures.

Viewing Fig. 5, filling-in occurs quickly either when the initial perceptual power is low or when perceptual power is greatly attenuated. The filling-in time is set by both the initial perceptual power and slope of lines (i.e. degree of facilitation in the model). In addition, the slopes of lines in Fig. 5 apparently depend on the spatio-temporal frequency of the dynamic texture.

Thin black lines in Fig. 6 additionally show estimated time courses of perceptual power for broad spatio-temporal frequency bands. Each black line is depicted with that for narrow bands included in the broad band. The initial perceptual power for the broad band is higher than that for narrow ones, so the filling-in time for the broad band is expected to be long. However, most of black lines attenuate greater than or equal to including narrow ones (marked '*'). Perceptual power for the broad band is initially high; it then tends to attenuate greatly.

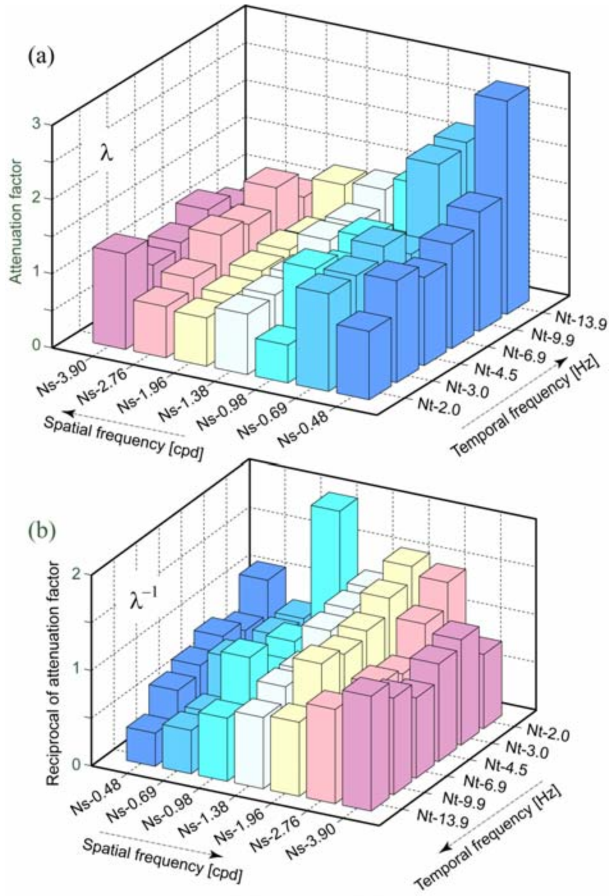

Figure 7: (a) Estimated attenuation factor for spatio-temporal frequency. It means a degree of filling-in facilitation. (b) Reciprocal of attenuation factor. It represents the degree of filling-in inhibition.

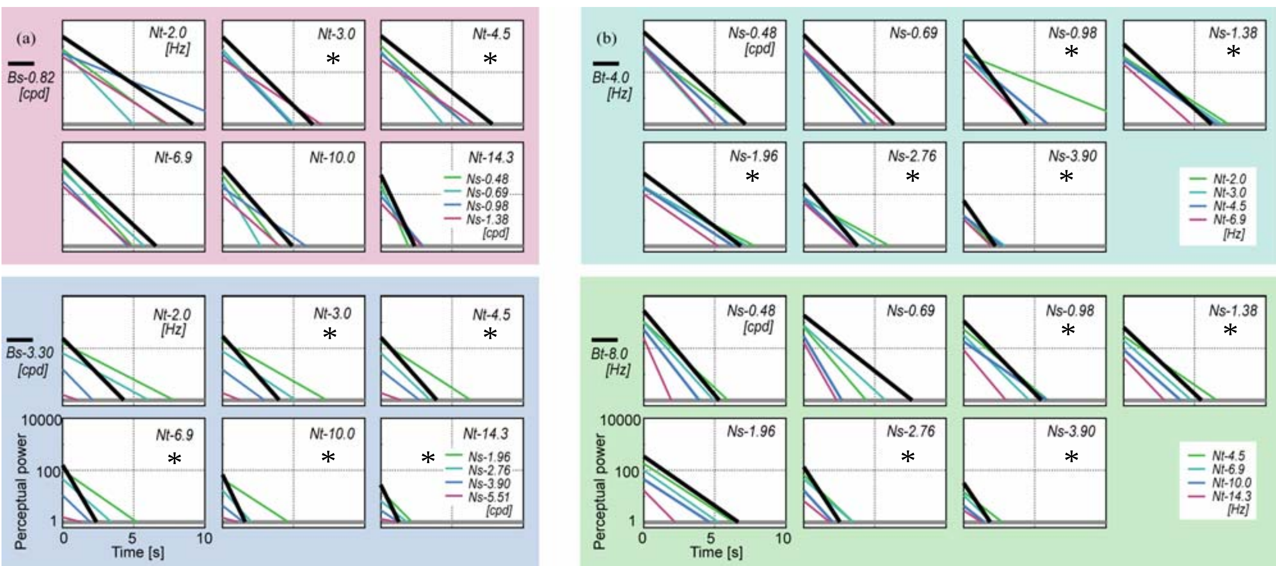

Figure 6: Estimated time course of perceptual power for broad spatio-temporal frequency bands, depicted with that for narrow bands which is included the broad band. The black line in each figure is for broad band, colored lines in the same figure represent the narrow bands.

Attenuation factor: The slope of the time courses, which indicates the degree of attenuation, corresponds to λ in Eq. (2), so we will designate λ as an attenuation factor. The estimated attenuation factor is given in Fig. 7(a) as a function of the spatial and temporal frequency of the dynamic texture. The spatio-temporal frequency of the dynamic texture seems to affect facilitation of filling-in. For illustration, the reciprocal of the attenuation factor, which shows the degree of filling-in time inhibition, is shown in Fig. 7(b).

Discussion

Filling-in that is induced in peripheral vision is reported to be realized in visual area II (V2) or visual area III (V3) [11]. Neural signals in the lateral geniculate nucleus (LGN) project into these visual areas. In the LGN, the signal pathway is divided mainly into Magno (M)-channels and Parvo (P)-channels. The spatio-temporal frequency sensitivity of an M-channel has its peak if the spatial frequency is low and temporal frequency is high. In contrast, the P-channel has high sensitivity if the spatial frequency is high and the temporal frequency is low. Respective spatio-temporal sensitivity characteristics of M-channels and P-channels reported by Merigan and Maunsell are shown in Fig. 8(a) and Fig. 8(b) [10].

Next, we compare filling-in facilitation (Fig. 7(a) and Fig. 7(b)) with M-channel and P-channel sensitivities in LGN (Fig. 8). The estimated attenuation factor (Fig. 7(a)) simply decreases as the spatial frequency increases; also, it increases as the temporal frequency increases if the spatial frequency is low. These properties agree with the sensitivity of M-channels in the spatio-temporal frequency range of our experiments (Fig. 8). On the other hand, the attenuation factor is low, around 2 [cpd], concerning the spatial frequency. Therefore, the reciprocal of the attenuation factor (Fig. 7(b)) has its peak there. Additionally, the reciprocal of the attenuation factor decreases as the temporal frequency increases there. These properties are similar to those of the P-channel sensitivity (Fig. 8). That agreement suggests that filling-in is facilitated if the M-channel sensitivity is high, but it is inhibited if the P-channel sensitivity is high.

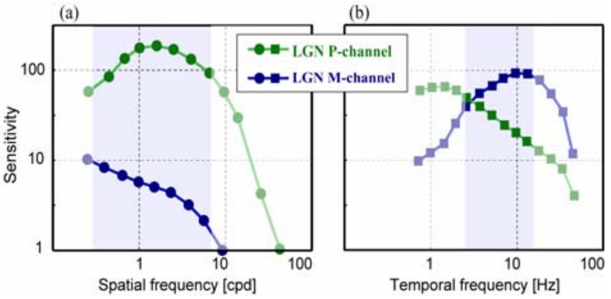

Figure 8: Spatio-temporal frequency sensitivity of the LGN P-channel and M-channel. (a) Spatial frequency sensitivity. (b) Temporal frequency sensitivity. The pale blue ranges were adopted in our experiment. (Modified from Merigan and Maunsell, 1993[10])

With respect to M-channels and P-channels of LGN, Spillmann and Kurtenbach presumed that filling-in is attributable to the fact that P-cells responding to the target are inhibited by M-cells responding to the surrounding dynamic texture [2]. We cannot clarify whether or not the M-channel suppresses the P-channel. Nevertheless, under the assumptions in our model of the filling-in process, our results provide evidence that the M-channel facilitates filling-in and that the P-channel inhibits it. Now we might be able to imagine that neural activation of M-channels responding to surrounding dynamic texture invades into the cells responding to the target. Filling-in will occur if the invading M-channel activation dominates or suppresses P-channel activation.

We will attempt to explain some visual mechanisms based on the above suggestions.

First, we will account for the mechanism by which perceptual filling-in is easily induced artificially in peripheral vision, but rarely in central vision. The P-channel dominates central vision, and the M-channel dominates peripheral vision. The above suggests that filling-in is inhibited in the center, but facilitated in peripheral vision.

We will next consider that filling-in induced in peripheral vision is reported as realized in V2 or V3 [11], although filling-in occurring in the retinal blind spot is realized in V1 [12–14], especially in the 6th layer in V1. Weerd reported that the neural signal that is interpreted as occurrence of filling-in is not found in V1 [11]. Based on that suggestion, "in which layer of V1 the neural signal is measured" is a salient issue because filling-in is facilitated in V1 layers that are projected from LGN M-channel, but it is inhibited in layers projected from the P-channel. Additional detailed physiological experiments are necessary to clarify this matter.

In addition, we adopt an annular filling-in target; the texture around the fixation point is blurred with graduated gray to stabilize subjects' fixation. It is interesting that a part of annular filling-in target locates on a retinal blind spot. However subjects felt that filling-in occurs homogeneously, and perceived no specific appearance at the blind spot.

After experiments, subjects reported that the filling-in target was perceived to be filled from the inner edge to the outer edge. Furthermore, blurred-gray around the fixation point was perceived to expand into its surrounding dynamic texture in certain conditions; this phenomenon is a kind of perceptual filling-in. These subjects' comments reflect that filling-in expands from the central to the peripheral visual field. This fact is very interesting from the viewpoint of understanding the filling-in mechanism. Additional experiments are required in the near future to explain the filling-in expansion phenomenon.

ISSN: 1727-1983 © 2005 IFMBE

Conclusions

This study analyzed facilitation of perceptual filling-in for spatio-temporal frequencies of dynamic textures surrounding a filling-in target.

First, the time to filling-in was measured for various spatio-temporal frequencies of dynamic textures. Then, the spatio-temporal frequency of human peripheral vision was estimated.

We applied these results to the proposed model of the filling-in process. The results revealed that perceptual power decreases greatly when the dynamic texture has a high-sensitivity spatio-temporal frequency of the M-channel in LGN. In contrast, it decreases slowly where the P-channel sensitivity is high. We infer that filling-in is facilitated in the M-channel, but inhibited in the P-channel.

Acknowledgement

This research was partially supported by a Grant-in-Aid for Scientific Research (C) 2005, KAKENHI #15500332 from JSPS in Japan.

References

[1] RARACHANDRAN V. S. and GREGORY R. L. (1991): 'Perceptual filling in of artificially induced scotomas in human vision', *Nature*, **350**, pp.699–702,

[2] SPILLMANN L. and KURTENBACH A. (1992): 'Dynamic noise backgrounds facilitate target fading,' *Vision Res.*, **32(10)**, pp.1941–1946

[3] WEERD D., DESIMONE R. and UNGERLEIDER L. G. (1998): 'Perceptual filling-in: a parametric study,' *Vision Res.*, **38(18)**, pp.2721–2734

[4] WELCHMAN A. E. and HARRIS J. M. (2001): 'Filling-in the details on perceptual fading,' *Vision Res.*, **41(16),** pp.2107–2117

[5] SAKAGUCHI Y. (2001): 'Target/surround asymmetry in perceptual filling-in,' *Vision Res.*, **41(16)**, pp.2065–2077

[6] YOKOTA M. and YOKOTA Y. (2002): 'Dependence of perceptual filling-in on spatio-temporal frequency', Proc. of 2nd European Medical & Biological Engineering Conference, pt.2, Vienna Austria, 2002, p.1232–1233

[7] CAMPBELL and ROBSON J. (1968): 'Application of Fourier analysis to the visibility of gratings,' *J. Physiology*, **197**, pp.551–566

[8] ROVANO J., VIRSU V. and NASANEN R. (1978): 'Cortical magnification factor predicts the photopic contrast sensitivity of peripheral vision,' *Nature*, **271**, pp.54–56

[9] VIRSU V. and ROVAMO J. (1979): 'Visual resolution, contrast sensitivity, and the cortical magnification factor,' *Exp. Brain Res.*, **37**, pp.475–494

[10] MERIGAN W. and MAUNSELL J. (1993): 'How parallel are the primate visual pathways?', *Ann. Rev. Neurosci.*, **16**, pp.369–402

[11] WEERD P. D., GATTASS R., DESIMONE R., and UNGERLEIDER L. G. (1995): 'Responses of cells in monkey visual cortex during perceptual filling if an artificial scotoma', *Nature*, **377**, pp.731–734

[12] FIORANI M., ROSA M. and GATTASS R. (1992): 'Dynamic surrounds of receptive fields in primate striate cortex: a physiological basis for perceptual completion?', *Proc. Natl. Acad. Sci. USA*, **89**, pp.8547–8551

[13] KOMATSU H., KINOSHITA M. and MURAKAMI I. (2000): 'Neural responses in the retinotopic representation of the blind spot in the macaque V1 to stimuli for perceptual filling-in,' *J. Neuroscience*, **20(24)**, pp.9310–9319

[14] KOMATSU H., KINOSHITA M. and MURAKAMI I. (2002): 'Neural responses in the primary visual cortex of the monkey during perceptual filling-in at the blind spot', *Neurosci. Res.*, **44**, pp.231–236

ISSN: 1727-1983