Neurosciences

Image phase or amplitude? Rapid scene categorization is an amplitude-based process

Nathalie Guyader [a,b,*], Alan Chauvin [a,b], Carole Peyrin [a], Jeanny Hérault [a], Christian Marendaz [b]

[a] *Laboratoire des images et des signaux (CNRS), INPG, 46, av. Félix-Viallet, 38031 Grenoble cedex, France*
[b] *Laboratoire de psychologie et neurocognition (CNRS), UPMF, BP 47, 38040 Grenoble cedex, France*

Presented by Pierre Buser

Abstract

Models of the visual cortex are based on image decomposition according to the Fourier spectrum (amplitude and phase). On one hand, it is commonly believed that phase information is necessary to identify a scene. On the other hand, it is known that complex cells of the visual cortex, the most numerous ones, code only the amplitude spectrum. This raises the question of knowing if these cells carry sufficient information to allow visual scene categorization. In this work, using the same experiments in computer simulation and in psychophysics, we provide arguments to show that the amplitude spectrum alone is sufficient for categorization task. *To cite this article: N. Guyader et al., C. R. Biologies 327 (2004).*
© 2004 Académie des sciences. Published by Elsevier SAS. All rights reserved.

Résumé

Information de phase ou information d'amplitude ? La catégorisation rapide des images est un processus basé sur le spectre d'amplitude. En traitement d'image, la modélisation du cortex visuel permet une décomposition des images selon leur spectre de Fourier (amplitude et phase). Il est communément admis que l'information de phase est nécessaire à l'identification d'une scène. Or, nous savons que les cellules complexes du cortex visuel (les plus nombreuses) codent uniquement le spectre d'amplitude. Se pose alors la question de savoir si ces cellules véhiculent une information suffisante pour permettre la catégorisation de scènes visuelles. Nous montrons, par une simulation informatique inspirée de la biologie du système visuel et par une expérience de psychophysique, que le spectre d'amplitude seul suffit à catégoriser des scènes. *Pour citer cet article : N. Guyader et al., C. R. Biologies 327 (2004).*
© 2004 Académie des sciences. Published by Elsevier SAS. All rights reserved.

Keywords: image Fourier transform; natural scenes; categorization

Mots-clés : transformée de Fourier ; scènes naturelles ; catégorisation

---

* Corresponding author.
 E-mail address: nguyader@lis.inpg.fr (N. Guyader).

1. Introduction

Experimental studies have shown that complex natural scenes can be categorized within a short time, faster than 150 ms [1], suggesting a simple and efficient coding process. In terms of signal representation, the Fourier components of an image can be expressed as amplitude and phase spectra; this has raised the question of which component, phase or amplitude, is the most diagnostic for scene perception. For two decades, it has been commonly believed that phase information dominates the perception of visual scenes [2–5]. However, it is well established that the primate primary visual cortex is widely dominated by complex cells [6,7], that is cells that respond preferentially to orientation and spatial frequency, but not to spatial phase [8]. On the other hand, we have shown that, from a theoretical viewpoint, scenes can be categorized by using only their energy spectrum (i.e., the squared amplitude spectrum) [9–11]. In this study, by means of simulation and psychophysical experiments, we show that amplitude-based processes are sufficient to allow rapid scene categorization.

2. Stimuli

We constructed four types of stimuli: (*i*) the non-modified or 'integral' image, which can be reconstructed from both its amplitude and phase spectra by Inverse Fourier Transform (IFT); (*ii*) an 'amplitude-component' image that was created by applying an IFT on the image amplitude spectrum modulated by a white-noise phase spectrum; (*iii*) a 'phase-component' image that was created with the IFT of the image phase spectrum associated to a flat-amplitude spectrum; (*iv*) a 'chimera', containing information from two different scenes, that was created by mixing the amplitude spectrum of one scene with the phase spectrum of another scene. In order to investigate whether amplitude- or phase-spectrum information was used by human observers in scene categorization, we conducted two psychophysical experiments using a priming paradigm. We hypothesized that if prime and target scenes had the same amplitude spectra (i.e., consistent), categorization of the target should be faster than when prime and target had different amplitude spectra (inconsistent). In order to increase the exper-

iment sensitivity, we used only two scene categories (beaches and cities), due to their very dissimilar amplitude spectra (Fig. 1a and b).

3. Model

For each experiment, the priming effects were assessed by a biologically inspired mathematical visual-processing model. This model consists of two different stages: in the first stage, image spectra are whitened to simulate the retina function (Fig. 2).

In the second stage, the energy (squared amplitude) spectrum of images is sampled by Gabor filters into eight orientations and six frequency bands (Fig. 3), to simulate the activity of V1 cortical complex cells.

So, each image is represented by a 48-dimension vector; each dimension corresponds to the image energy at the frequency f_m and at the orientation θ_n:

$$E_{f_m,\theta_n} = \sum_{f_x,f_y} \left|I(f_x, f_y)\right|^2 \left|G_{f_m,\theta_n}(f_x, f_y)\right|^2$$

$|I(f_x, f_y)|^2$ is the spectral energy density function of the analysed image; $G_{f_m,\theta_n}(f_x, f_y)$ is the transfer function of a Gabor filter.

Then, the Euclidean distance between the 48-dimensional vectors of prime and target was computed: a low value of this distance means that prime and target have very similar energy spectra.

4. Experiment I

Seventy-two observers participated in the first experiment. Stimuli subtended $9.7° \times 9.7°$ of visual angle. On each trial, a fixation point was displayed in the centre of the screen for 2 s. Then the following sequence was displayed: a prime image (10 ms) a mean grey screen (30 ms), a dynamic mask [12] (160 ms), a target scene (20 ms), and again a mean grey screen (1.5 s). Stimuli were 256×256 pixel images with a 256-grey-level scale. Prime stimuli were either 'integral', 'amplitude-component' or 'phase-component' images. Primes were chosen in order to have typical amplitude spectra (mainly *horizontal* for city scenes and mainly *vertical* for beach scenes, in the spectral domain [9]). Observers were asked to press a button as fast and accurately as possible when the target scene
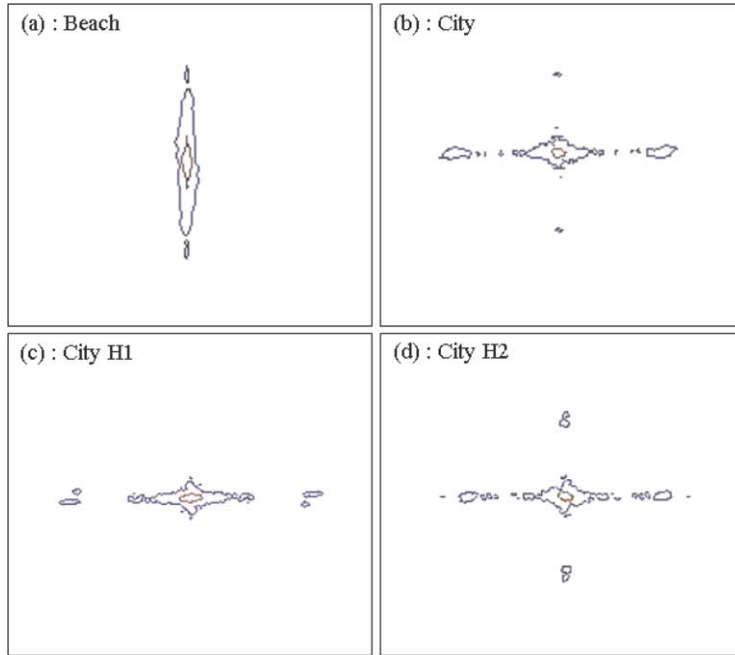
Fig. 1. The mean amplitude spectra, in the spectral domain of (**a**) beach scenes or (**b**) city scenes for Experiment I; (**c**) H1 city target group and (**d**) H2 group for Experiment II (see text).

Fig. 2. The original image (on the left) is locally compressed in shading area by photoreceptors (in the middle), then high-pass filtered by the retina circuits (into the right image). We will note the contrast equalisation in the whole image.
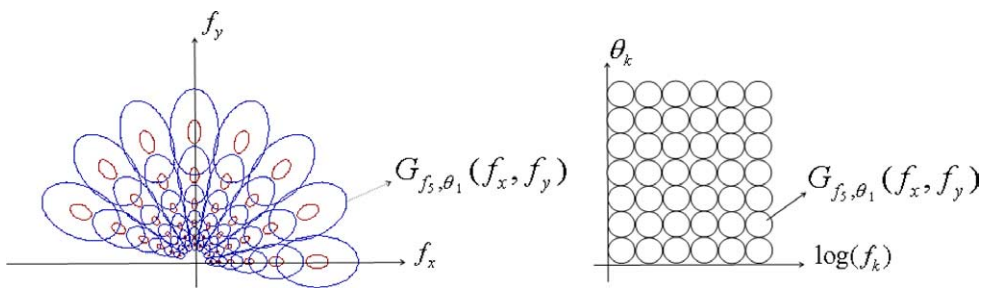

Fig. 3. The set of Gabor filters (-3 dB and -0.2 dB) in the spectral domain (left) and its log-polar representation (right).

was recognized (beach or city) in a go/nogo procedure. Response Time (RT) and accuracy were recorded. The presentation order of the experimental trials was randomised.

4.1. Simulation results

According to the model simulation, smaller prime-target Euclidean distances were observed for consistent (Cs) than for inconsistent (ICs) conditions for both 'integral' and 'amplitude-component' images, but not for 'phase-component' images. As our model is only based on the amplitude spectrum of images, the same distance difference for 'integral' and 'amplitude-component' image primes was found (Fig. 4a).

4.2. Psychophysical results

Across the experimental conditions the error rate was very low ($< 1.2\%$); therefore only RTs (on correct trials) were analysed. Even if mean reaction time is shorter for the 'phase-component' prime, the type of prime did not influence the mean reaction time.

When the prime was an 'integral' image, target categorization time was significantly faster for the Cs condition (379 ms) than for the ICs condition (397 ms) [$F(1, 66) = 15.6$, $p < 0.01$]. A similar priming effect occurred when the prime was an 'amplitude-component' image (Cs, 395 ms; ICs, 410 ms; $F[1, 66] = 11.11$, $p < 0.01$). However, as predicted, no effect of degree of prime-target consistency was observed when the prime consisted of 'phase-component' image (Cs, 373 ms; ICs, 375 ms; $F < 1$). Moreover, the priming effect obtained with 'integral' primes did not differ from the one with the 'amplitude-component' image prime condition ($F < 1$), but differed from that obtained with 'phase-component' primes ($F[1, 66] = 6.02$, $p < 0.02$; Fig. 4a).

As predicted by our model, scene categorization could be based on the global information provided by the image-amplitude spectrum components.

Also, the priming effect was of the same magnitude whether the prime consisted of an 'integral' or an 'amplitude'-component image. However, our data showed that the priming amplitude (RTs difference between ICs and Cs conditions) was higher for beach than for city scenes (15 ms and 10 ms, respectively).

This could be explained by the fact that the comparison of the amplitude spectrum among city targets showed more differences than among beach targets, which could reduce the priming effect. But it was impossible to examine this possibility in Experiment I because, in order to limit perceptual learning, each exemplar was seen once, i.e., was preceded by only one type of prime. So, a second experiment was necessary to clarify this point.

5. Experiment II

The second experiment had two aims: first, to estimate the consistency effect as a function of the degree of homogeneity of amplitude spectrum of city images (so, each exemplar target was preceded by all types of prime); second, to strengthen the experimental test of image amplitude spectrum as a basic mechanism subtending the rapid categorization of images by using a more complex type of prime: a 'chimera'. A 'chimera' has the amplitude component of one category (beach or city) and the phase component of the other category (city or beach). Within the framework of the model, a spectral 'chimera' should show the same priming power as 'integral' or 'amplitude component' image primes, considering that all these primes have the same amplitude spectrum.

Ten observers participated in Experiment II. The same priming paradigm as for the first experiment was used: prime spectra had a non-ambiguous orientation (*vertical* for beaches and *horizontal* for cities) and, in consistent conditions, the prime and the target had amplitude spectra belonging to the same category. Stimuli and design were partially modified in relation to the above-mentioned aims. The prime was either an 'amplitude component' image (as in Experiment I), or a 'chimera'. The orientation of the amplitude spectrum of city images was either homogeneous (horizontal; H1) or heterogeneous (horizontal and vertical; H2) (Fig. 1 c and d).

5.1. Simulation results

According to our model, we observed a difference of Euclidean distances between Cs and ICs conditions only for H1 targets, the magnitude of this difference
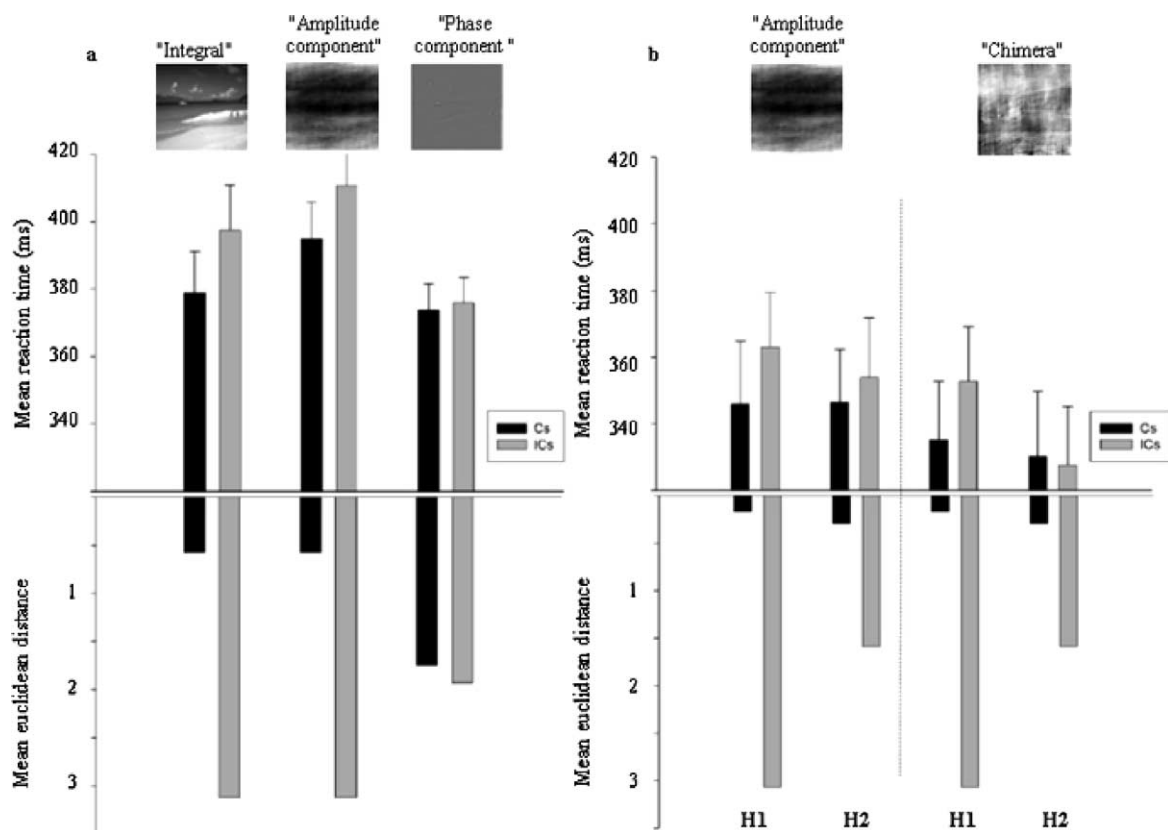
Fig. 4. On each graph (**a** and **b**), experimental results are given above the *x*-axis and model results are given below the *x*-axis. Model results (lower histograms) give the mean Euclidean distance between prime and target for each condition. Experimental results (upper histograms) show the mean reaction time for each condition of prime. (**a**) Experiment I: RTs as a function of factors type of prime ('integral', 'amplitude'- or 'phase-component' prime) and prime-target consistency (Cs, consistent vs. ICs, inconsistent). (**b**) Experiment II: RTs as a function of factors type of prime ('amplitude-component' or 'chimera' prime), target amplitude spectrum heterogeneity (H1 vs. H2) and prime-target consistency (Cs vs. ICs); see text.

being the same whatever the type of prime ('amplitude component' and 'chimera') (Fig. 4b).

5.2. *Psychophysical results*

As the error rate was very low ($< 0.3\%$), data analysis was performed only on correct RTs (Fig. 4b). With 'amplitude-component' primes and H1 targets, RTs were significantly faster for Cs (345 ms) than for ICs condition (363 ms) ($F[1, 8] = 6.14$, $p < 0.05$). In contrast, no consistency effect occurred with H2 targets (Cs, 346 ms; ICs, 353 ms; $F[1, 8] = 2.9$, ns). The same pattern of effects was obtained when the prime was a 'chimera': The effect of consistency was significant for H1 targets (Cs, 335 ms; ICs, 352 ms;

$F[1, 8] = 7.88$, $p < 0.05$), but not for H2 targets (Cs, 330 ms; ICs, 327 ms; $F[1, 8] < 1$). Moreover, for H1 targets, the effect of consistency with 'amplitude-component' primes did not differ from the effect obtained with 'chimera' primes ($F < 1$; Fig. 4b).

To sum up as Experiment I, Experiment II shows that the consistency effect depends only on the degree of match between the amplitude spectra of the prime and target scenes.

6. Discussion

In the paper, we show with psychophysical results on categorization task that the priming effect

(difference between consistent and inconsistent conditions) is due to the amplitude component of images. On a theoretical point of view, a simplified model based only on the amplitude spectrum of images achieves more than 80% accuracy on image categorization [13]. Here, we compare for a same experiment of categorization, the reaction time of subjects with a distance, computed by an amplitude-spectrum-based model, between prime and target. These two results present similar implications when prime and target are consistent for 'integral' and 'amplitude-component' primes.

This conclusion fits with the observation that phase information is ignored by complex cells in low level vision [8] and that an artificial neural network is able to categorize natural scenes on the only basis of their amplitude spectra [11]. Visual scene categorization can be viewed as a solution to a general information-processing problem: a trade-off between the need for detailed information and the need for efficiency and generalization. To this end, the amplitude spectrum of a scene has a more general signification than its phase spectrum: the amplitude spectrum is independent of the spatial positions of the scene's components, whereas the phase spectrum is closely tied to these positions. In other words, a given amplitude spectrum is more representative of an image category than a given phase spectrum, which is representative of only the corresponding given image exemplar. Moreover, as the categorization process is mainly based on low spatial frequency components, it could be linked to the magnocellular pathway, which could explain its observed rapidity [14].

References

[1] S.J. Thorpe, D. Fize, C. Marlot, Speed of processing in the human visual system, Nature 381 (1996) 520–522.

[2] A.V. Oppenheim, J.S. Lim, The importance of phase in signals, IEEE Proc. 69 (1981) 514–529.

[3] L.N. Piotrowski, F.W. Campbell, A demonstration of the visual importance and flexibility of spatial-frequency amplitude and phase, Perception 11 (1982) 337–346.

[4] M.J. Morgan, J. Ross, A. Hayes, The relative importance of local phase and local amplitude in patchwise image reconstruction, Biol. Cybern. 65 (1991) 113–119.

[5] S.E. Palmer, Vision Science: Photons to Phenomenology, MIT Press, 1999.

[6] D.H. Hubel, T.N. Wiesel, Receptive fields and functional architecture of monkey striate cortex, J. Physiol. 195 (1968) 215–243.

[7] G.F. Poggio, Spatial properties of neurons in striate cortex of unanesthetized Macaque monkey, Invest. Ophthalmol. 11 (1972) 368–377.

[8] L. Shams, C. von der Malsburg, The role of complex cells in object recognition, Vision Res. 42 (2002) 2547–2554.

[9] A. Guérin-Dugué, A. Oliva, Classification of scene photographs from local orientations features, Pattern Recogn. Lett. 21 (2000) 1135–1140.

[10] A. Schaaf, J.H. Hateren, Modeling the power spectra of natural:statistics and information, Vision Res. 36 (1996) 2759–2770.

[11] J. Hérault, A. Oliva, A. Guérin-Dugué, Scene categorisation by Curvilinear Component Analysis of low frequency spectra, in: Proc. 5th European Symposium on Artificial Neural Network, D Facto Publications, Brussels, 1997, pp. 91–96.

[12] S.J. Thorpe, N.M. Bacon, G.A. Rousselet, M.J. Macé, M. Fabre-Thorpe, Rapid categorisation of natural scenes: feedforward vs. feedback contribution evaluated by backward masking, Perception (Suppl.) 31 (2002) 150.

[13] N. Guyader, J. Hérault, Représentation espace–fréquence pour la catégorisation d'images, GRETSI 01, Toulouse, France, 2001.

[14] J. Bullier, Integrated model of visual processing, Brain Res. Rev. 36 (2001) 96–107.