Neurocomputing 73 (2010) 2522–2531

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Coarse scales are sufficient for efficient categorization of emotional facial expressions: Evidence from neural computation

Martial Mermillod [a,*], Patrick Bonin [b], Laurie Mondillon [c], David Alleysson [d], Nicolas Vermeulen [e]

[a] Laboratoire de Psychologie Sociale et Cognitive, Clermont Université, Université Blaise Pascal, and Centre National de la Recherche Scientifique, UMR 6024, Clermont-Ferrand, France
[b] LEAD-CNRS UMR 5022, Université de Bourgogne, Dijon, France
[c] Laboratoire LIP/PC2S-EA 4145, Université de Savoie, France
[d] Laboratoire de Psychologie et NeuroCognition, CNRS UMR 5105, Universite Pierre Mendes, France
[e] Psychology Department, Université catholique de Louvain (UCL) and National Fund for Scientific Research, Belgium

ARTICLE INFO

Article history:
Received 24 September 2008
Received in revised form
20 May 2010
Accepted 10 June 2010
Communicated by M.S. Bartlett
Available online 25 June 2010

Keywords:
Neural computation
Cognitive science
Cognitive neuroscience
Computer vision
Emotional facial expressions

ABSTRACT

The human perceptual system performs rapid processing within the early visual system: low spatial frequency information is processed rapidly through magnocellular layers, whereas the parvocellular layers process all the spatial frequencies more slowly. The purpose of the present paper is to test the usefulness of low spatial frequency (LSF) information compared to high spatial frequency (HSF) and broad spatial frequency (BSF) visual stimuli in a classification task of emotional facial expressions (EFE) by artificial neural networks. The connectionist modeling results show that an LSF information provided by the frequency domain is sufficient for a distributed neural network to correctly classify EFE, even when all the spatial information relating to these images is discarded. These results suggest that the HSF signal, which is also present in BSF faces, acts as a source of noisy information for classification tasks in an artificial neural system.

1. Introduction

1.1. Cognitive neuroscience data

The purpose of the present article is to determine whether LSF components are sufficient for efficient categorization of an EFE. This hypothesis is based on different neuroimaging and cognitive science experiments showing that the human cognitive system may have a fast way of accessing LSF components relating to threat recognition in the visual environment [23,14,38]. For instance, a neuroimaging study (functional MRI; [38]) suggests the possibility of a preferential link between magnocellular layers in the lateral geniculate nucleus (LGN) and the amygdala. This study [38] revealed the existence of a hemodynamic response at the level of a subcortical pathway involving the superior colliculus, the pulvinar and the amygdala when participants processed LSF pictures of faces depicting a fearful expression compared to LSF faces displaying a neutral expression. These results, therefore, suggest that the transmission of the signal associated with the facial expression of fear might bypass the primary visual cortex by taking a subcortical pathway [10,14], possibly emanating from the magnocellular layers in the LGN, which transports the low spatial frequency information very quickly. This type of fast access to visual information might be of particular value for increasing sensory exposure to potentially dangerous events [37]. LSF and HSF stimuli are processed by two different visual streams at the level of the lateral geniculate nucleus. Whereas the magnocellular neurons primarily provide rapid, but low spatial frequency cues which encode configural features as well as brightness and motion, the parvocellular neurons provide slower, but also higher spatial frequency information (finer visual details) about local shape features, color and texture [19]. Conversely, this fMRI study [38] also showed that faces filtered at high spatial frequencies only slightly activated the amygdala and that the signal activated different structures in the ventral pathway (the occipito-temporal cortex and the face fusiform area). This result is corroborated by ERP studies [29].

Thus, the underlying question we address in the current paper is to determine if the biological structure of the human cognitive system is adapted to the computational properties of the visual environment when providing rapid access to an LSF information in a suitable EFE recognition task. In other words, we assume that the phylogenetic development of the human neural structures dedicated to categorization of an EFE may be such as to provide faster access to coarse scale information conveyed by

* Corresponding author. Tel.: +33 4 73 40 62 54.
E-mail address: Martial.Mermillod@univ-bpclermont.fr (M. Mermillod).

0925-2312/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2010.06.002

magnocellular neurons, because this information is more efficient for categorization of EFE.

1.2. Artificial neural networks and emotional facial expressions

A number of different methods for the modeling of artificial systems capable of performing emotional facial expression recognition tasks have been reported in the literature [3]. Among these different recognition systems, the use of artificial neural networks permits an efficient classification of an EFE [16]. The first step involved in the use of connectionist networks consists of compressing the visual information. Thus, authors have suggested using radial basis function (or RBF) networks consisting of Gaussian receptive fields which compress the information relating to various parts of an image [7,28]. However, different studies have shown that the use of Gabor wavelets permits a better modeling of the receptive fields of the simple cells of the primary visual cortex [5]. This research has shown that the statistical evaluation of the residual error between the difference in the response profiles of V1 simple cells and Gabor filters is not distinguishable from chance [11,12,39,40] therefore suggesting Gabor wavelets for face recognition tasks performed by artificial systems. Using this technique, series of Gabor wavelets are convolved with specific parts of the image, in order to extract the information relating to different wavelengths and different orientations. Similarly, [16] successfully used Gabor wavelets in combination with a Support Vector Machine (SVM) for the static or dynamic recognition of an EFE [2]. Finally, [15,4] have shown that the convolution of Gabor wavelets, which are sensitive to different wavelengths and different orientations using a sliding window applied to the entire image permits reliable categorization, which is comparable to human data when the wavelets are associated with an SVM, discriminant analysis [18] or an artificial neural network [4].

The computational model that we propose in this study is very similar to the model proposed by Dailey et al. [4] but with two differences. At the level of the perceptual encoding of the stimuli, the Gabor wavelets are implemented not in the spatial domain, but in the frequency domain by means of the modulus of the discrete Fourier transform (DFT). Each wavelet is therefore multiplied by the local energy spectrum of the Fourier transform. This local energy spectrum represents the quantity of energy associated with each spatial frequency and each orientation, independently of the spatial location of the wavelet. This is convenient because (i) it avoids a subsequent visual data compression step and (ii) it makes the representation of the image (i.e. the output computed by the Gabor filters) phase invariant (as V1 complex cells [6]). The second difference in our model can be found in the artificial neural network, which simulates the association between the perceptual output of the stimuli and the category label which encodes each EFE category. Dailey et al. [4] used a single layer perceptron in association with the softmax activation function at the level of the output layer, in order to perform non-linear categorization of their images, whereas we used the standard back-propagation algorithm on a multi-layer perceptron [20,31]. This actually represents small differences with an exception, however, that our method actually constitutes a novel approach to visual data compression that deserves to be discussed in the light of a comparison with the methods used previously [2,4,15,16].

1.3. Visual data compression

The main difference between our perceptual model of vision and previous models proposed in [2,4,15,16] resides at the level of visual data compression. Performing Gabor filtering in the spatial domain results in a huge perceptual vector (for example, a 40,600 vector size in [4]). Therefore, Gabor filtering in the spatial domain requires an additional step in order to reduce the size of the perceptual layer for subsequent neural network processes. For instance, some authors [4] have chosen to reduce the perceptual space by means of a "gestalt layer" produced by means of a principal component analysis (PCA), and then focusing neural computation on the first 50 eigenvectors. Using this technique, they have obtained efficient results for an EFE categorization. However, this technique raises an important methodological problem given the objectives of the present article. It has been shown elsewhere that the first eigenvectors correlate highly with LSF information [1]. In other words, virtually all the eigenvectors are necessary to retain HSF details, with the result that it is not possible to reduce visual information, while investigating the role of SF channel when using this method.

Another way to reduce the size of visual information for subsequent processing in an artificial neural network is to use feature selection algorithm such as Adaboost [16]. Adaboost is an efficient technique for selecting Gabor filters that are relevant for a subsequent associative task (for example, categorizing an EFE). In other words, Adaboost selects different Gabor filters that are sensitive to the different SF or orientations that are necessary to categorize different EFE. However, we did not use this algorithm to test the second main hypothesis of this paper, namely that an efficient categorization of an EFE can be obtained even after the spatial location of the Gabor filter has been completely removed. This hypothesis is based on biological evidence that is described below.

It is possible to perform Gabor filtering by means of a convolution of a Gabor kernel in the spatial domain (at a specific location or within a sliding window), which is the formal equivalent to a multiplication of this Gabor filter in the Fourier domain. When this multiplication is applied to the Fourier transform of the entire image, this method resembles a type of holistic vision and means that each Gabor filter provides an average energy value for the whole image. The originality of this method is that it does away with spatial locations.

At a methodological level, it has the advantage of retaining exactly the same amount of information, in quantitative and qualitative terms, for each SF channel, thus making it possible to compare the different SF channels in the most balanced way possible. It should also be noted that all the main references in the same field [2,4,15,16] as our current perceptual model, also use the energy spectrum of the Gabor filters, and thus discard the phase information necessary to reconstruct the spatial information of an original image. However, using the magnitude spectrum of Gabor filters within a sliding window is just one step on the way towards eliminating spatial information: it removes phase information which is important for the spatial reconstruction of the image but retains the spatial location of the filters. In this paper, we will show that we can go a step further in removing spatial location for the purposes of the efficient categorization of an EFE. At a theoretical level, we assume that taking an average energy value of the Gabor filters over the whole image might be sufficient for the efficient categorization of an EFE. This assumption is based in part on biological data reported in a single-cell recording study, which showed that neurons in the medial temporal lobe (MTL) respond independently to spatial location [30]. In other words, neurons become less and less sensitive to spatial location during the bottom-up process from the retina to the temporal lobes, with the result being a completely abstract representation at the end of this process (at the level of the MTL). In other words, in this paper, we propose the provocative idea that spatial location might not be necessary for an efficient categorization of complex stimuli such as an EFE.

The first simulation is based on a limited but widely used database of EFE: the pictures of emotional affect (POFA) Database [8]. The aim of this pilot simulation is to determine whether the superiority of LSF signals for the categorization of EFE emerges even with a very small number of training exemplars. Simulation 2 was then designed to confirm the results obtained in Simulation 1 using a broader database, the *Karolinska Directed Emotional Faces* (KDEF) [17] as well as to extend this result to the full spectrum of spatial frequency channels. We have used the KDEF and POFA not only because they are two commonly employed databases, but also because they consist of very carefully controlled pictures selected for important neuroimaging, behavioral and connectionist modeling papers [8,17,24,38] in the field of EFE categorization. To summarize, the aim of this paper was to test the efficiency of different SF channels on stimuli that were carefully controlled for in neuroimaging and behavioral experiments on the basis of commonly used and standardized computational methods (Gabor filters at the perceptual level and connectionist networks at the associative level).

2. Simulation 1. Pictures of emotional affect database.

2.1. Method

2.1.1. Neural network

To perform our simulations, we used an image database of gray-scale images of facial expressions. The size of the images was $N \times N$ (with $N=256$ pixels). First, we applied a Hann window to avoid boundary effects in the subsequent Fourier transform. Boundary effects could result in a bias toward an over-representation of cardinal orientations, and the Hann window is frequently used to suppress this bias. The following formula describes the one-dimensional Hann window of size N ($i=0, ..., N-1$) applied vertically and horizontally to each image by pixelwise multiplication

$$w(i) = 0.5 - 0.5 \times \cos\left(\frac{2\pi i}{N}\right)$$

Next, we applied Gabor receptive fields in the spectral domain. Multiplying a Gabor receptive field in the spectral domain is equivalent to a convolution in the spatial domain. It is therefore possible to perform the filtering in the spectral domain by multiplying the spatial frequency information by the kernel of the Gabor function

$$G(\mathbf{x},\mathbf{y},\mathbf{f_c},\boldsymbol{\theta}) = \frac{1}{2\pi\sigma_r\sigma_t} e^{-\frac{(\mathbf{x}\mathbf{u})^2}{2\sigma_r^2}} e^{-\frac{(\mathbf{x}\mathbf{u}_\perp)^2}{2\sigma_t^2}} e^{j2\pi\mathbf{x}\mathbf{f_c}}$$

with

$$\begin{cases} \mathbf{x} = [x,y]^t, & \mathbf{f_c} = [f_0\cos\theta, -f_0\sin\theta]^t \\ \mathbf{u} = [\cos\theta, \sin\theta]^t, & \mathbf{u}_\perp = [\sin\theta, \cos\theta]^t \end{cases}$$

A Gabor filter is constructed with a Gaussian modulated by a complex exponential: parameters σ_r and σ_t of the Gaussian determine the spatial extent of the filter. The vector f_c with modulus f_0 and direction θ describes this sine wave. In summary, each individual image was transferred into the Fourier domain and filtered by a set of Gabor filters determining energy coefficients by coding the local energy spectra. We applied a bank of 56 Gabor wavelets corresponding to seven different spatial frequency channels (the distance between two consecutive centers was one octave and spatial extent increased by one octave per spatial frequency channel) and eight different orientations (0, $\pi/8$, $2\pi/8$, $3\pi/8$, $4\pi/8$, $5\pi/8$, $6\pi/8$, $7\pi/8$), with respect to biological data [6].

The second component then took the form of a back-propagation neural network, whose aim was to classify the output vectors provided by the Gabor filters [4,21,22]. Our connectionist network was thus used as a computational tool that allowed us to analyze the subtle and distinctive statistical properties of the faces with respect to their emotional categories. The connectionist network was a 3-layer back-propagation neural network. We used the standard hetero-association training algorithm, whose function is to associate each of the different category exemplars with a specific output vector coding for each category. During the feed-forward phase, activation was rescaled by means of a sigmoid transfer function

$$f(a) = \frac{1}{1+e^{-a}}$$

where $f(a)$ is the output activation value and a is the sum of the input activation vector multiplied by the input-to-hidden weight matrix.

The input vector activation was then propagated through the network, layer-by-layer, until it reaches the output layer. Then, the supervised learning algorithm computed the sum of squared error (SSE)

$$E = \frac{1}{2}\sum_n\sum_k (t_{pk} - o_{pk})^2$$

In this equation, p indexes the pattern in the training set, k indexes the output nodes, t_{pk} the desired output for the *kth* output node for the *pth* pattern, o_{pk} the observed output for the *kth* output node for the *pth* pattern.

Then, the error signal was computed, using the standard back-propagation algorithm [31] and back-propagated through the network until convergence of the neural network. Fig. 1.

2.1.2. Stimuli

The stimuli used came from the picture of facial affect (POFA) database [8]. We used the same pictures as in a deep electrode ERP study [13]. The identities used are referred to as EM, JJ, PE, WF, C, MF, MO, NR, PF and SW. These identities, presented face on, are expressing 6 basic emotional expressions (joy, disgust, fear, anger, sadness and surprise) to which a neutral expression was added for control purposes (Fig. 2).

Each image was defined on a gray scale and was centered in a 256×256 pixel frame for computational reasons relating to the symmetry of the rosette of Gabor wavelets applied to an image. We then applied a low-pass Gaussian filter (Fig. 3) with a cut-off frequency, which made it possible to retain the spatial frequencies lower than 8 cycles per image for the LSF images and higher than 24 cycles per image for the HSF images [33]. The mean brightness of the stimuli was normalized at 90.3 on a scale of 256 levels of gray for the LSF, HSF and BSF images.

2.1.3. Procedure

The neural network was used for simulations in hetero-associative mode similar to [4], i.e. each output vector generated by the Gabor wavelets for each of the images was associated with a specific code relating to an EFE category (1 0 0 0 0 0 0 for angry faces, 0 1 0 0 0 0 0 for disgust, 0 0 1 0 0 0 0 for fear, 0 0 0 1 0 0 0 for joy, 0 0 0 0 1 0 0 for neutrality, 0 0 0 0 0 1 0 for sadness and 0 0 0 0 0 0 1 for surprise). The energy vectors produced by the Gabor wavelets were standardized between 0 and 1 at the input to the network and independently, in order to avoid inducing any bias in favor of a spatial frequency or specific orientation. The network architecture consisted of 56 input units (the energy response of

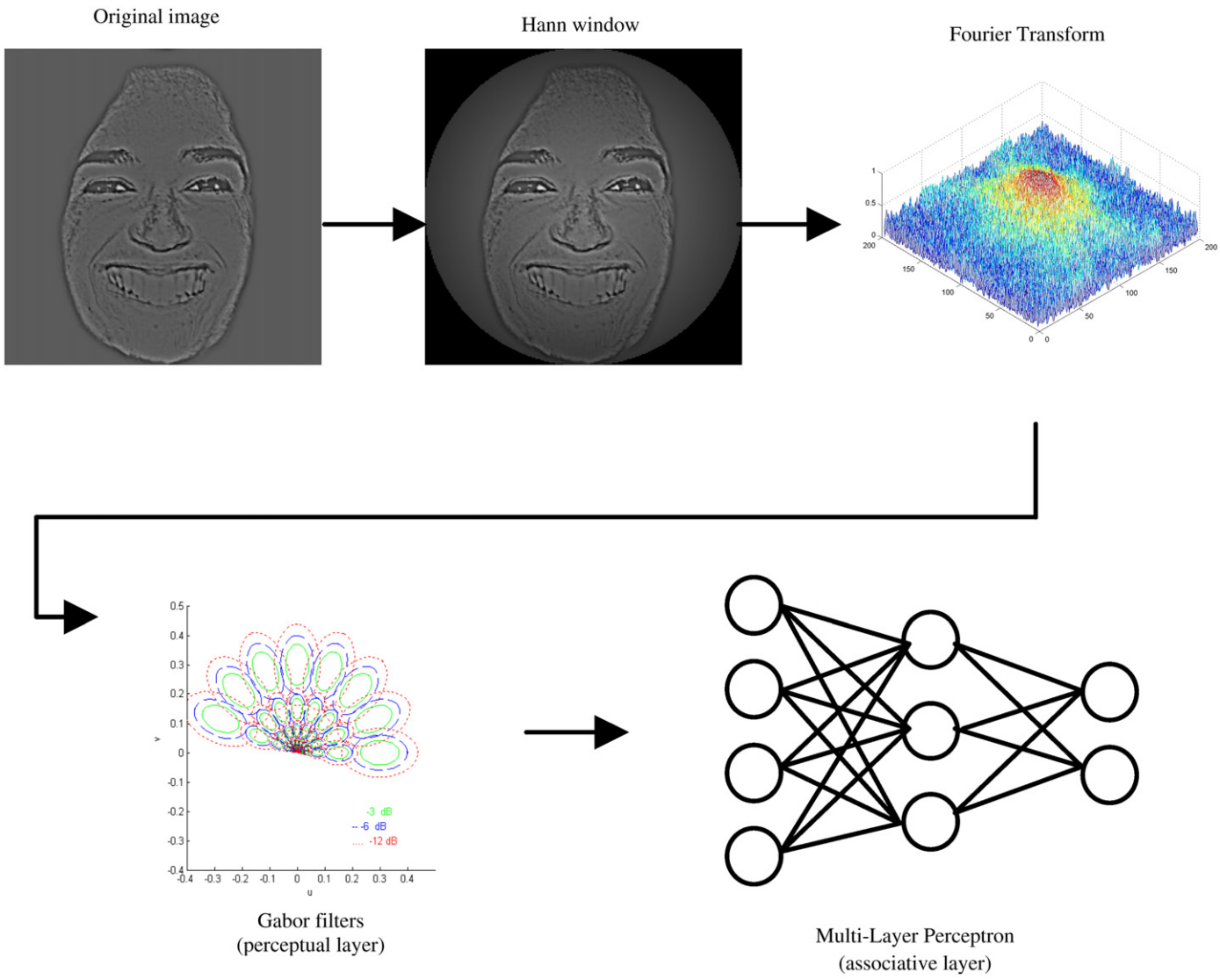


Fig. 1. Experimental procedure.

Fig. 2. Example of stimuli from the picture of facial affect database.

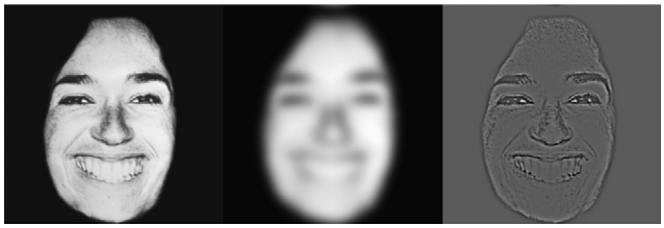


Fig. 3. Example of BSF, LSF (provided by magnocellular layers) and HSF faces (only provided by parvocellular layers).

the Gabor wavelets), 42 hidden units and 7 output units (the category code for each category). The learning rate was fixed at 0.1 and the momentum at 0.9. The learning algorithm, the procedure and the learning parameters were selected to correspond to the parameters that are most widely used in the literature. Modifying these parameters (number of hidden units, learning rate, momentum) changed the speed of learning, but under no circumstances the statistical properties associated with each EFE or each spatial frequency channel. In other words, increasing or reducing the learning speed induced a global advantage for all the categories (and not just for one specific category compared to another).

Learning phase: each trial started with the random selection of 63 learning exemplars (9 faces out of 10 for each of the 7 EFE categories). The learning phase then consisted of associating each of these 63 learning exemplars with the correct category code over 1000 iterations.

Test phase: after learning each of the EFE expressed by 9 randomly selected actors, the network was tested for its ability to

generalize to the tenth actor. The energy vector of each of the EFE was provided at the input to the network, and we recorded the output calculated at the output from the model. A "winner-take all" procedure was then applied to the observed output, in order to determine the network's response to each EFE. This procedure was applied 10 times for each network, with a new random selection of the learning and test faces on each trial, for each network, in order to calculate a correct response percentage on 10 different trials for each network. The dependent variable was the percentage of correct responses after this procedure had been run 30 times.

2.2. Results

We performed an ANOVA with repeated measures on the EFE category (joy, disgust, fear, anger, sadness, surprise and neutral) X frequency channel (LSF, HSF, BSF), with the percentage of correct responses as the dependent variable. The ANOVA revealed a significant main effect of EFE type ($F(6,174)=22.3$; $MSE=180$; $p<0.001$). Exhaustive multiple comparisons (Tukey's HSD) showed that the categories anger ($M=70.78$; $SE=2.65$), sadness ($M=68.56$; $SE=1.91$) and surprise ($M=69.56$, $SE=2.05$) were significantly better recognized by the networks than the other emotional expressions. At the same time, the ANOVA also revealed a significant effect of frequency channel ($F(2,58)=156.2$; $MSE=182$; $p<0.001$) which took the form of a better categorization of LSF information ($M=75.6$; $SE=2.51$) compared to both the HSF information ($M=52.7$; $SE=1.87$; $F(1,29)=515.56$; $MSE=106.4$; $p<0.001$) and the BSF images ($M=60.2$; $SE=2.51$; $F(1,29)=109.43$; $MSE=225.6$; $p<0.001$).

The EFE X frequency channel interaction was also significant here ($F(12,348)=12.9$; $MSE=217$; $p<0.001$). As we expected, the LSF resulted in better categorization performances (compared to both the HSF and BSF images) on the faces expressing fear ($p<0.001$), joy ($p<0.001$) and surprise ($p<0.001$). The LSF channel therefore permitted a better categorization than the HSF channel (but not the BSF images) for the sad faces ($p<0.001$). In contrast, there was no significant advantage for the LSF compared to the HSF channel for the recognition of disgust, anger or for the neutral faces. Table 1.

2.3. Discussion

The data collected by the neural network enables us to perform a precise computational analysis of the relevance of each frequency channel in the recognition of specific EFE. The first important insight contributed by this analysis is that the low frequency channel leads to better overall EFE recognition and categorization performance. These results are consistent with human behavioral data showing that the low spatial frequencies seem to be the ones that are most useful for the recognition of a specific joyful, angry or neutral EFE, whereas the high spatial frequencies seem to be most useful in determining whether a face is expressive or not [33].

The expressions of fear, joy and surprise were better identified via the LSF than the HSF channel. The LSF images also resulted in better performances than the BSF images for these three expressions. As far as the expression of sadness is concerned, this was recognized better in LSF than in HSF mode, but not better than the BSF images. Finally, no advantage of the LSF over the HSF was observed for the expressions of disgust, anger or neutrality.

Finally, we have to note that the overall computational performance of the neural network is lower than that observed in the main references in the field [2,4,15,16]. This could be due (i) to the fact that spatial information is necessary and, therefore, that applying Gabor filters in the frequency domain results in the loss of diagnostic information that is important for the purposes of EFE categorization or, alternatively, (ii) to the small training patterns used. Indeed, 9 training exemplars may not be sufficient to enable a back-propagation classifier to find the correct boundaries among each category in the high-dimensional space provided by the perceptual layer. It should also be noted that 9 training exemplars might not be sufficient for a human trained on completely novel categories. The purpose of Simulation 2 was to test this hypothesis.

Table 1
Mean correct percentage for each emotion and each spatial frequency channel.

Emotion	Spatial frequency		
	BSF	LSF	HSF
Anger	65.4	77.8	68.2
Disgust	56.4	64.6	61.4
Fear	47.2	72	49
Happiness	44.2	81.6	42.6
Neutral	51.8	56	63.8
Sadness	72	79.6	47.2
Surprise	69.6	86.4	41.4

3. Simulation 2. Karolinska directed emotional faces database

3.1. Method

3.1.1. Neural network

The neural network, parameters and procedure were strictly identical to Simulation 1 except that we used 448 stimuli, 64 stimuli × 7 emotions (from the KDEF database) for the training session (instead of 9 stimuli × 7 emotions in the POFA database, Simulation 1) and 1 remaining item per emotion to test the generalization property of the neural network. Categorization rate was computed across 100 runs with a new test item being selected at random for each run.

3.1.2. Stimuli

The stimuli used came from the Karolinska Directed Emotional Faces (KDEF) database [17]. Among the 70 identities constituting the database, 5 were removed because of bad lighting or low image quality. We then applied a Hann window identical to that used in Simulation 1. Each picture was converted to a gray-scale image centered in a 256 × 256 pixel frame, and we applied band-pass filters which increased by one octave between two consecutive filters: below 8 CpI; 8–16 CpI; 16–32 CpI; 32–64 CpI, above 64 CpI (Fig. 4). The mean brightness was then normalized to 118.71 for all images.

3.2. Results

The results reported in Table 2 indicate a consistent decrease in categorization accuracy of the neural network on change-over from LSF to HSF channels.

As in Simulation 1, we performed an ANOVA on the EFE category (joy, disgust, fear, anger, sadness, surprise and neutral) and X frequency channel (BSF; <8 CpI; 8–16 CpI; 16–32 CpI; 32–64 CpI; >64 CpI). As in Simulation 1, the ANOVA revealed a significant main effect of the type of EFE ($F(6,3564)=6.13$; $MSE=0.428$; $p<0.001$). The post-hoc Tukey test revealed that, for this database, fearful and surprised faces were recognized significantly better than angry and happy faces. The differences between all the other EFE were not significant. Of more interest

Fig. 4. Example of stimuli from the Karolinska Directed Emotional Faces database. From the left to the right: BSF faces, < 8 CpI; 8–16 CpI; 16–32 CpI; 32–64 CpI and > 64 CpI.

Table 2
Mean correct percentage for each emotion and each spatial frequency channel.

Emotion	Spatial frequency					
	BSF	< 8 CpI	8–16 CpI	16–32 CpI	32–64 CpI	> 64 CpI
Anger	93	96	96	94	84	57
Disgust	92	94	93	91	86	84
Fear	98	96	96	96	88	88
Happiness	96	95	94	94	79	70
Neutral	90	96	95	99	79	84
Sadness	92	94	97	98	91	63
Surprise	91	100	95	92	89	90
Grand Average	93.14	95.86	95.14	94.86	85.14	76.57

with regard to our hypothesis is the fact that we found a significant effect of spatial frequency channels ($F(5,594)=53.9$; $MSE=0.103$; $p<0.001$), which reveals that the LSF information is categorized better than the HSF information. Moreover there was also a significant interaction effect between SF channels and EFE ($F(30,3564)=7.14$; $MSE=0.07$; $p<0.001$), which suggests that different EFE might have different diagnostic SF channels. However, as shown in Fig. 5, all EFE produced a decreased recognition rate above 32 CpI.

In order to test this effect, we performed exhaustive pair-wise comparisons between all SF channels on the basis of a post-hoc Tukey test. Table 3.

Pair-wise comparisons reveal that the two highest SF channels produced categorization rates significantly below each of the other SF channels. Moreover there was no statistical difference between the lower three SF channels, and there was no statistical difference between these and the BSF faces.

3.3. Discussion

The results provided by Simulation 2 clearly support the finding reported in Simulation 1, namely that LSF channels allow more efficient categorization of EFE by the neural network. This finding was particularly clear for SF above 32 CpI. Fig. 5 reveals that the different diagnostic cues relevant for each individual EFE might be based on different SF channels. Whereas, the diagnostic cues seem to appear around 8 CpI for surprise, anger and disgust for example, it seems that the diagnostic cues for sadness or neutral expressions require higher SF channels around 16–32 CpI. However, the results mainly revealed a considerable deterioration in performance for virtually all EFE above 32 CpI.

Table 3
Exhaustive pair-wise comparisons between all SF channels.

	BSF	< 8 CpI	8–16 CpI	16–32 CpI	32–64 CpI	> 64 CpI
BSF		0.420351	0.918217	0.998911	0.000020	0.000020
< 8 CpI	0.420351		0.953424	0.212609	0.000020	0.000020
8–16 CpI	0.918217	0.953424		0.736495	0.000020	0.000020
16–32 CpI	0.998911	0.212609	0.736495		0.000020	0.000020
32–64 CpI	0.000020	0.000020	0.000020	0.000020		0.000057
> 64 CpI	0.000020	0.000020	0.000020	0.000020	0.000057	

It should also be noted that, with this larger training set, our neural network (which uses a simpler method which eliminates spatial information) performed categorization as efficiently as the best references in the field, which use Gabor filters the spatial domain [2,4,15,16]. However, we should mention that these different simulations did not use the same database and further simulations will be required to address the question of the usefulness of spatial information on the basis of a strictly identical database.

4. Conclusion

The current computational results suggest that LSF information could be particularly effective for the recognition of specific facial expressions. Our connectionist simulations revealed a clear superiority of LSF over HSF information in both simulations. At a computational level, these results are consistent with [16]. These authors used Adaboost, SVM, and a linear discriminant analysis to select Gabor filters, in the spatial domain, that result in a higher categorization rate for each EFE. Using this method, they showed that with a SF ranging from 3 up to 48 cycles per face, the "diagnostic features" may appear at around 17 cycles per face. Our method is complementary to this "bottom-up" approach by showing that, even with the same amount of information at each scale and with a holistic vision which incorporates each spatial location, LSF information permits better categorization performance than HSF information for subsequent associative processes in neural networks.

Another main finding, and an important difference compared to [16], is the fact that this result was obtained irrespective of the spatial location of the Gabor filters. This result, combined with the fact that both methods discard important spatial information (related to the phase of the signal), while using only the magnitude of the output provided by the Gabor filters, raises an important underlying theoretical question: is spatial location necessary to perform reliable categorization and identification of emotional expressions? As far as EFE are concerned, the answer provided by Simulation 2 is clear: spatial location is not necessary, at least at the level of LSF filters, for the efficient categorization of the images. It is important to note here that we are going a step further than other computer vision algorithms in discarding both phase information and the spatial location of the Gabor filters in our approach to this question.

This raises important potential implications concerning the possibility of generalizing this finding to other categories of objects. It is possible that, at least in the case of foveal vision (i.e. a focus on the object of interest, inducing the removal of the background to this object), a global representation of the magnitude spectrum seems to be sufficient to efficiently categorize EFE. With regard to generalization to other types of objects, this finding could have important implications at a behavioral and computational level. This means that spatial location could be

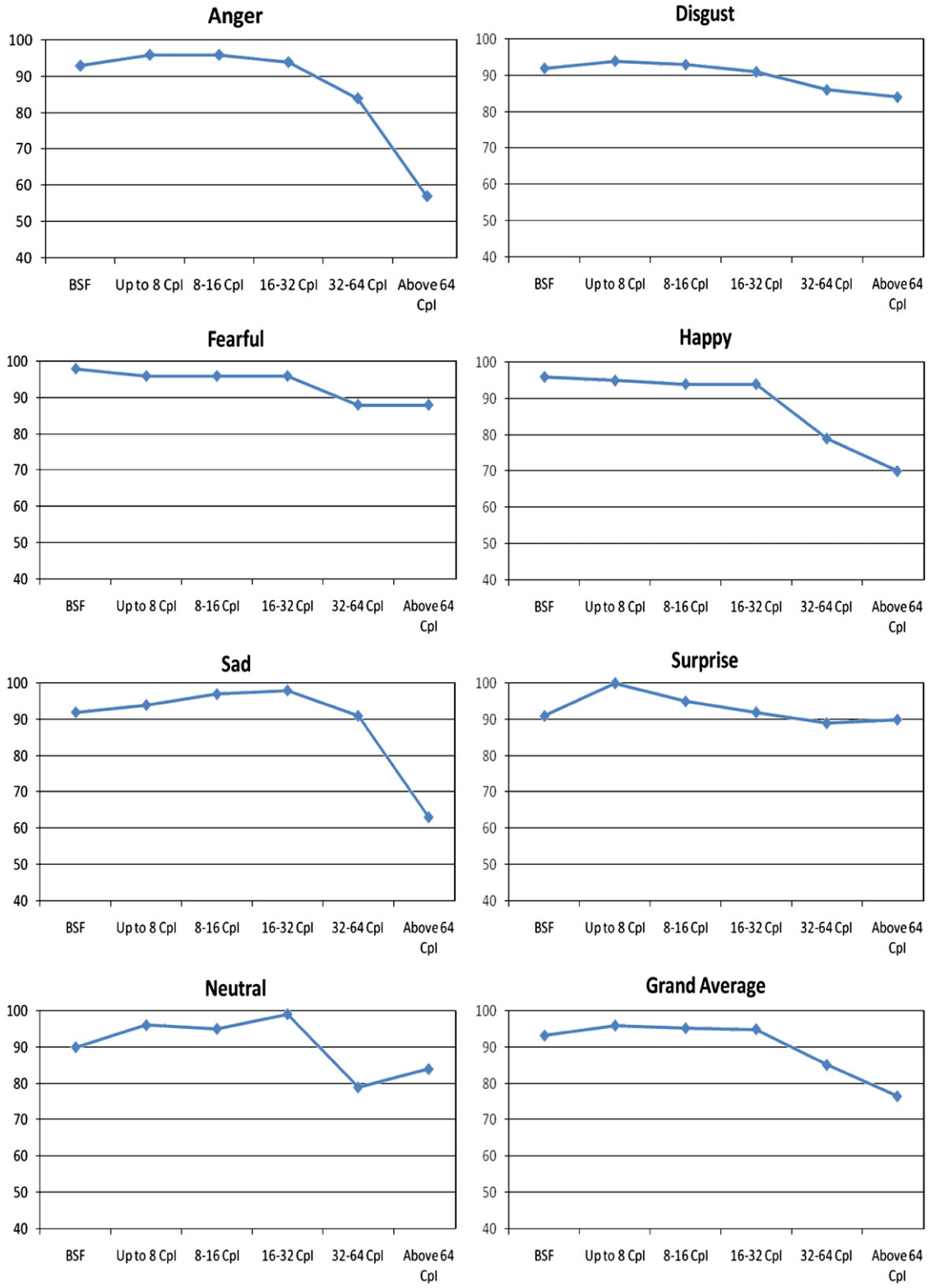


Fig. 5. Recognition rate for each EFE depending on different SF channels.

important simply in that it focuses foveal vision on the object of interest and that, once the focus has been established, a very simple and fast access to the magnitude of the frequency spectrum of the object is sufficient to categorize it reliably. Of course, this hypothesis goes far beyond the scope of the current paper and will require a careful examination of neural network performance across a wide range of objects and natural scene stimuli.

Concerning the superiority of LSF information in the recognition of an EFE: at a qualitative level, recent behavioral

results [35] suggest that the diagnostic scales relevant for each EFE might be similar in both human subjects and our current neural network simulation. In line with other results previously reported in the literature at both the computational [1,9,25] and behavioral levels [23,34,36,38], the results are probably due to the computational properties of the LSF or HSF signals. As suggested by [1], an LSF signal is probably less variable and therefore more useful for categorization purposes than a HSF signal. As far as the superiority of LSF channels over HSF channels for categorization purposes is concerned, our results are also consistent with the behavioral results obtained by [33], as well as other studies conducted in the field of developmental psychology [36] and neural network modeling [9] which have shown that HSF information may in fact constitute a source of perceptual interference in a categorization task. In fact, unlike the HSF signal, the LSF signal is very stable in terms of its perceptual variability, thus explaining the high correlation which exists between the initial eigenvectors of a PCA and the LSF signal [1]. Given this view, the HSF information would not be at all effective in an EFE categorization task even though this information could be of crucial importance for other cognitive tasks, for example finding the identity of a specific individual. It should be noted that the performance impairment due to the HSF signal also induced a considerable impairment in performance on the BSF images in Simulation 1 and a slight impairment in Simulation 2. This represents a major difference compared to behavioral results obtained in humans. In effect, human participants achieve better performances when confronted with the BSF images: they do not therefore seem to be sensitive to the perceptual interference generated by the HSF signal in the case of BSF images. This is an important avenue for future research in the fields of psychology and connectionist modeling. Indeed, it is possible that the coarse-to-fine integration process observed in humans [26,27,32] is associated with a mechanism, which optimizes the integration of the HSF information within the primary structure supplied by the LSF information. Whereas the connectionist network makes identical use of each frequency channel in parallel, humans perform a rapid, initial analysis of the LSF information which complements and refines the signal provided by the HSF information. This process might take advantage of the computational properties of both LSF and HSF information.

Concerning the importance of spatial information, further simulations will have to be performed to test whether this superiority of LSF information persists in the case of spatially located Gabor filters. It is possible that HSF information might potentially be superior if faces are carefully aligned across trials and spatial information is retained for neural computation. However, this new line of research raises several questions: what level of translation would impair the performance of a categorization system which uses HSF channels? Are such HSF channels more sensitive to other geometric transformations (such as rotation or depth for example) than LSF channels? These important questions will have to be carefully addressed in future neural network studies. As a result, new simulations designed to test the effect of a superiority of LSF channels with a sliding window will have to overcome two major problems: it will be necessary (i) to reduce the extent of the perceptual layer for subsequent neural network processes (using PCA or feature selection), while at both the quantitative and qualitative levels, retaining the same amount of information for each SF channel and (ii) to determine an alignment condition and other geometrical transformations, which would make it possible to use HSF information for categorization purpose. However, these different methodological and theoretical considerations do not detract from the finding reported in the current paper, namely that LSF information concerning the global amplitude spectrum is *sufficient* for the efficient categorization of EFE.

In addition, this theoretical point raises other questions at a biological level. For example, we can assume that the human perceptual system has developed to adapt to these computational properties of LSF and HSF channels. This means that the structure of the perceptual system in terms of spatial and temporal frequency properties (at the level of the retina, as well as the magno-, parvo- and koniocellular layers and also within V1 and subsequent cortical areas) could be related to the computational properties of the environment. Again, a coarse-to-fine bias [26,27,32] might constitute the best way of avoiding the problem of the potential sensitivity of HSF channels to geometric transformation (translation, rotation or depth). This question of whether the anatomical and functional properties of the human primary visual system have adapted to the computational properties of the physical environment has to be carefully addressed in future studies in the fields of cognitive neuroscience, psychology and neural computation.

Acknowledgments

This work was supported by the French CNRS (UMR 6024) and a grant from the French National Research Agency (ANR Grant BLAN06–2_145908, ANR Grant BLAN08-1_353820).

Appendix

Simulation 1

Tables 4–6.

Simulation 2

Tables 7–12.

Table 4
Confusion matrix for BSF images in Simulation 1.

BSF	Anger	Disgust	Fearful	Happy	Neutral	Sad	Surprise
Anger	65.4	1.4	7	9	11.6	0.2	5.4
Disgust	0.2	56.4	10.4	15.8	17	0	0.2
Fearful	7	11.2	47.2	3.4	15	1	15.2
Happy	17.2	18.2	1.4	44.2	5.8	0	13.2
Neutral	1.4	8.4	0	12.2	51.8	1.8	24.4
Sad	0.6	11.8	1	1	1.8	72	11.8
Surprise	9.6	1.2	5.2	5	3.2	6.2	69.6

Table 5
Confusion matrix for HSF images in Simulation 1.

HSF	Anger	Disgust	Fearful	Happy	Neutral	Sad	Surprise
Anger	68.2	9.8	5	6.8	5	0	5.2
Disgust	0.2	61.4	0.4	9.6	6.8	9.6	12
Fearful	7.8	0	49	8.4	4	0	30.8
Happy	4.6	12	26.4	42.6	1.6	11.2	1.6
Neutral	1.6	10.6	10.4	2.4	63.8	1.8	9.4
Sad	0.6	9.4	5.6	14.8	7.4	47.2	15
Surprise	4	1.8	26.4	3.6	5.6	17.2	41.4

Table 6
Confusion matrix for LSF images in Simulation 1.

LSF	Anger	Disgust	Fearful	Happy	Neutral	Sad	Surprise
Anger	77.8	0	0.4	1.4	13.8	0	6.6
Disgust	9.2	64.6	9	0.2	0.8	0	16.2
Fearful	0	13.8	72	12.2	0.2	0	1.8
Happy	9.6	0.2	1.4	81.6	5.8	0	1.4
Neutral	14.4	7	0.4	13	56	5	4.2
Sad	1.2	0.4	3.2	0	15.4	79.6	0.2
Surprise	12.2	1	0	0	0	0.4	86.4

Table 7
Confusion matrix for BSF images in Simulation 2.

BSF	Anger	Disgust	Fearful	Happy	Neutral	Sad	Surprise
Anger	93	1	1	2	2	1	0
Disgust	1	92	1	1	1	4	0
Fearful	1	0	98	0	0	1	0
Happy	0	0	1	96	3	0	0
Neutral	0	1	2	5	90	1	1
Sad	0	2	5	0	0	92	1
Surprise	0	2	1	3	3	0	91

Table 8
Confusion matrix for <8CpI images in Simulation 2.

<8 CpI	Anger	Disgust	Fearful	Happy	Neutral	Sad	Surprise
Anger	96	0	0	1	0	0	3
Disgust	1	94	2	0	0	0	3
Fearful	0	1	96	1	0	2	0
Happy	1	0	0	95	1	2	1
Neutral	0	0	0	4	96	0	0
Sad	0	0	0	4	1	94	1
Surprise	0	0	0	0	0	0	100

Table 9
Confusion matrix for 8–16 CpI images in Simulation 2.

8–16CpI	Anger	Disgust	Fearful	Happy	Neutral	Sad	Surprise
Anger	96	1	0	1	1	0	1
Disgust	0	93	0	1	1	3	2
Fearful	0	0	96	1	0	1	2
Happy	2	0	0	94	1	2	1
Neutral	0	2	0	1	95	2	0
Sad	1	1	1	0	0	97	0
Surprise	0	3	2	0	0	0	95

Table 10
Confusion matrix for 16–32 CpI images in Simulation 2.

16–32 CpI	Anger	Disgust	Fearful	Happy	Neutral	Sad	Surprise
Anger	94	1	1	4	0	0	0
Disgust	0	91	4	0	1	0	4
Fearful	0	2	96	0	1	0	1
Happy	1	1	1	94	1	0	2
Neutral	0	0	1	0	99	0	0
Sad	0	0	0	2	0	98	0
Surprise	0	3	2	1	1	1	92

Table 11
Confusion matrix for 32–64 CpI images in Simulation 2.

32–64 CpI	Anger	Disgust	Fearful	Happy	Neutral	Sad	Surprise
Anger	84	4	3	1	5	2	1
Disgust	4	86	0	2	4	0	4
Fearful	1	1	88	2	0	5	3
Happy	7	3	2	79	4	3	2
Neutral	9	4	1	5	79	1	1
Sad	2	0	1	6	0	91	0
Surprise	3	2	2	2	1	1	89

Table 12
Confusion matrix for >64 CpI images in Simulation 2.

>64 CpI	Anger	Disgust	Fearful	Happy	Neutral	Sad	Surprise
Anger	57	5	2	15	3	15	3
Disgust	4	84	1	10	0	0	1
Fearful	3	1	88	5	1	0	2
Happy	11	8	2	70	2	4	3
Neutral	3	3	4	2	84	3	1
Sad	14	3	0	9	8	63	3
Surprise	3	0	3	1	1	2	90

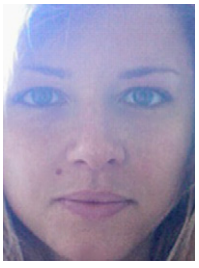
References

[1] H. Abdi, D. Valentin, B.E. Edelman, A.J. O'Toole, More about the difference between men and women: evidence from linear neural networks and the principal component approach, Perception 24 (5) (1995) 539–562.
[2] M.S. Bartlett, G.C. Littlewort, M.G. Frank, C. Lainscsek, I. Fasel, J.R. Movellan, Automatic recognition of facial actions in spontaneous expressions, Journal of Multimedia 1 (6) (2006) 22–35.
[3] J. Cohn, T. Kanade, Use of automated facial image analysis for measurement of emotion expression, in: J.A. Coan, J.B. Allen (Eds.), The Handbook of Emotion Elicitation and Assessment, Oxford University Press, New York, NY, 2006.
[4] M.N. Dailey, G.W. Cottrell, C. Padgett, A. Ralph, EMPATH: a neural network that categorizes facial expressions, Journal of Cognitive Neuroscience 14 (8) (2002) 1158–1173.
[5] J.G. Daugman, Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters, Journal of the Optical Society of America A 2 (7) (1985) 1160–1169.
[6] R.L. De Valois, K.K. De Valois, in: Spatial Vision, Oxford University Press, New York, 1988.
[7] S. Duvdevani-Bar, S. Edelman, Visual recognition and categorization on the basis of similarities to multiple class prototypes, International Journal of Computer Vision 33 (3) (1999) 201–228.
[8] P. Ekman, W. Friesen, in: Pictures of Facial Affect, Consulting Psychologists Press, Palo Alto, CA, 1976.
[9] R.M. French, M. Mermillod, P.C. Quinn, A. Chauvin, D. Mareschal. The importance of starting blurry: simulating improved basic-level category learning in infants due to weak visual acuity. in: Proceedings of the 24th Annual Conference of the Cognitive Science Society, Mahwah, NJ: Lawrence Erlbaum Associates, (2002) 322–327.
[10] B. de Gelder, J. Vroomen, G. Pourtois, L. Weiskrantz, Non-conscious recognition of affect in the absence of striate cortex, NeuroReport 10 (18) (1999) 3759–3763.
[11] J.P. Jones, L.A. Palmer, The two-dimensional spatial structure of simple receptive fields in cat striate cortex, Journal of Neurophysiology 58 (6) (1987) 1187–1211.
[12] J.P. Jones, A. Stepnoski, L.A. Palmer, The two-dimensional spectral structure of simple receptive fields in cat striate cortex, Journal of Neurophysiology 58 (6) (1987) 1212–1232.
[13] P. Krolak-Salmon, M.A. Hénaff, A. Vighetto, O. Bertrand, F. Mauguière, Early amygdala reaction to fear spreading in occipital, temporal, and frontal cortex: a depth electrode ERP study in human, Neuron 42 (2004) 665–676.
[14] J. Ledoux, The emotional brain: the mysterious underpinnings of emotional life, New York: Simon & Shuster (19962004).
[15] G. Littlewort, M. Bartlett, I. Fasel, J. Susskind, J. Movellan, Dynamics of facial expression extracted automatically from video, Image and Vision Computing 24 (6) (2006) 615–625.

[16] S. Lucey, A.B. Ashraf, J. Cohn, Investigating spontaneous facial action recognition through AAM representations of the face, in: K. Kurihara (Ed.), Face Recognition Book, Pro Literatur Verlag, Mammendorf2007, pp. 395–406.
[17] D. Lundqvist, A. Flykt, A. ÖhmanKarolinska directed emotional faces, Stockholm: Karolinska Institute and Hospital, Section of Psychology, (1998).
[18] M.J. Lyons, J. Budynek, S. Akamatsu, Automatic classification of single facial images, IEEE Transactions on Pattern Analysis and Machine Intelligence 21 (12) (1999) 1357–1362.
[19] M. Livingstone, D. Hubel, Segregation of form, color, movement, and depth: anatomy, physiology, and perception, Science 240 (4853) (1988) 740–749.
[20] J.L. McClelland, T.T. Rogers, The parallel distributed processing approach to semantic cognition, Nature Reviews Neuroscience 4 (2003) 310–322.
[21] M. Mermillod, N. Guyader, A. Chauvin, The coarse-to-fine hypothesis revisited: evidence from neuro-computational modeling, Brain & Cognition 57 (2) (2005) 151–157.
[22] M. Mermillod, N. Guyader, A. Chauvin, Improving generalisation skills in a neural network on the basis of neurophysiological data, Brain & Cognition 58 (2) (2005) 246–248.
[23] M. Mermillod, S. Droit-Volet, D. Devaux, A. Schaefer N., Vermeulen Are coarse scales sufficient for fast detection of visual threat? Psychological Science, (in press).
[24] M. Mermillod, N. Vermeulen, D. Lundqvist, P.M. Niedenthal, Neural computation as a tool to differentiate perceptual from emotional processes: the case of anger superiority effect, Cognition 110 (3) (2009) 346–357.
[25] M. Mermillod, P. Vuilleumier, C. Peyrin, D. Alleysson, C. Marendaz, The importance of low spatial frequency information for recognizing fearful facial expressions, Connection Science 21 (1) (2009) 75–83.
[26] D.M. Parker, J.R. Lishman, J. Hughes, Evidence for the view that tempor-ospatial integration in vision is temporally anisotropic, Perception 26 (9) (1997) 1169–1180.
[27] C. Peyrin, M. Mermillod, S. Chokron, C. Marendaz, Effect of temporal constraints on hemispheric asymmetries during spatial frequency processing, Brain & Cognition 62 (3) (2006) 214–220.
[28] T. Poggio, S. Edelman, A network that learns to recognize three-dimensional objects, Nature 343 (6255) (1991) 263–266.
[29] G. Pourtois, E.S. Dan, D. Grandjean, D. Sander, P. Vuilleumier, Enhanced extrastriate visual response to bandpass spatial frequency filtered fearful faces: time course and topographic evoked-potentials mapping, Human Brain Mapping 26 (2005) 65–79.
[30] R.Q. Quiroga, L. Reddy, G. Kreiman, C. Koch, I. Fried, Invariant visual representation by single neurons in the human brain, Nature 435 (7045) (2005) 1102–1107.
[31] D.E. Rumelhart, G.E. Hinton, R.J. Williams, Learning internal representations by error propagation, in: D.E. Rumelhart, J.L. McClelland (Eds.), Parallel Distributed Processing, vol. 1, MIT Press, Cambridge, MA, 1986.
[32] P.G. Schyns, A. Oliva, From blobs to boundary edges: evidence for time and spatial-scale-dependent scene recognition, Psychological Science 5 (4) (1994) 195–200.
[33] P.G. Schyns, A. Oliva, Dr. Angry and Mr. Smile: when categorization flexibly modifies the perception of faces in rapid visual presentations, Cognition 69 (3) (1999) 243–265.
[34] M.L. Smith, G. Cottrell, F. Gosselin, P.G. Schyns, Transmitting and decoding facial expressions of emotions, Psychological Science 16 (2005) 184–189.
[35] F.W. Smith, P.G. Schyns. Smile through your fear and sadness transmitting and identifying facial expression signals over a range of viewing distances. Psychological Science, (in press).
[36] G. Turkewitz, P.A. Kenny, Limitations on input as a basis for neural organization and perceptual development: a preliminary theoretical statement, Developmental Psychobiology 15 (4) (1982) 357–368.
[37] N. Vermeulen, J. Godefroid, M. Mermillod, Emotional modulation of attention: fear increases but disgust reduces the attentional blink, Plos One 4 (11) (2009) e7924.
[38] P. Vuilleumier, J.L. Armony, J. Driver, R.J. Dolan, Distinct spatial frequency sensitivities for processing faces and emotional expressions, Nature Neuroscience 6 (6) (2003) 624–631.
[39] L. Wiskott, Phantom faces for face analysis, Pattern Recognition 30 (6) (1997) 837–846.
[40] L. Wiskott, J.-M. Fellous, N. Krüger, C. von der Malsburg, Face recognition by elastic bunch graph matching, in: L.C. Jain, U. Halici, I. Hayashi, S.B. Lee, S. Tsutsui (Eds.), Intelligent Biometric Techniques in Fingerprint and Face Recognition, The CRC Press International Series on Computanional Intelligence, New York,1999, pp. 355–396.

Martial Mermillod was born in 1973. He received his Master's degree at Grenoble Universités, France, in 2000 and Ph.D. from the University of Liège, Belgium, in 2004. In 2004, he was awarded a post-doctoral grant by the Fyssen Foundation at Grenoble Universités. Since 2005, he has been an Associate Professor at the Université Blaise Pascal, Clermont-Ferrand, France. His current research investigates neural networks, cognitive science and cognitive neuroscience.

Patrick Bonin was born in 1966 in Le Creusot. He obtained both his Master's degree and Ph.D. at Université de Bourgogne (Dijon, France) in 1989 and 1995, respectively. In 1996, he acquired the title of an Associate Professor. In 2003, he became Full Professor at the Université Blaise Pascal, Clermont-Ferrand, France. Since September 2009, he has been Full Professor at the Université de Bourgogne. His current research interest is in cognitive psychology and psycholinguistics.



Laurie Mondillon was born in 1979. She received her Master's degree and Ph.D. from Université Blaise Pascal (Clermont-Ferrand, France) in 2003 and 2006, respectively. Since 2007, she has been an Associate Professor at the Université de Savoie, Chambery, France. Her current research focuses on social cognitive science, cognitive neuroscience and embodied knowledge, especially in the field of emotions.

David Alleysson gained his Masters degree and Ph.D. degree from Grenoble Universités, Grenoble, France, in 1994 and 1999, respectively. Between 2000 and 2003, he worked in a post-doctoral position at the Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland. Since 2003, he has been conducting research at Grenoble Universités. He currently specializes in computer science, computer vision and color perception.

Nicolas Vermeulen was born in Brussels, Belgium. He obtained both his Master's degree and Ph.D. from the Université catholique de Louvain (UCL), Belgium, in 2001 and 2005, respectively. Since 2005, he has received a grant from the Belgian Fund for Scientific Research (FRS-FNRS). In 2005, he spent one year working in the Emotion Laboratory headed by Dr. Paula M. Niedenthal at the Université Blaise Pascal, Clermont-Ferrand, France. He is currently working in the fields of cognitive science, attentional processes, embodied knowledge and emotions.