

Ref: Abdi, H., Valentin, D., (1994). Modèles neuronaux, connexionistes et numériques de la reconnaissance des visages. *Psychologie Française*, **39**(4), 357–392.

Modèles Neuronaux, Connexionistes et Numériques pour la Mémoire des Visages

HERVÉ ABDI*[†] et DOMINIQUE VALENTIN*

* *School of Human Development: The University of Texas at Dallas,
MS: GR.4.1., Richardson, TX75083-0688, U.S.A.*

[†] *Université de Bourgogne à Dijon, Boulevard Gabriel, 21000 Dijon, France*

ABSTRACT

Neural network models for face processing are classified according to three dimensions: Code and information representation, neuronal architecture, and type of task performed. Five types of models can be found: (1) linear auto-associators using a low-level code for coding faces (2) back-propagation networks using a low-level code (3) Gabor filters (or Morlet wavelets) operating on pictures of faces (4) models describing faces with a geometrical code (5) cognitive models using an arbitrary or random code. The code used by the models seems to be their most important feature. Currently, the auto-associator models are the most developed ones and the most computationally powerful.

1. INTRODUCTION

La psychologie cognitive des visages a été plus lente que d'autres composantes des sciences cognitives (*e.g.*, langage) à exposer ses modèles dans un cadre connexioniste. Depuis quelques années, toutefois, un nombre croissant de modèles a vu le jour (*cf.* Valentin *et al.* 1994; O'Toole *et al.* sous presse). Ces modèles se composent d'unités plus ou moins assimilables à des neurones regroupés en couches et reliés les uns aux autres par des connexions d'intensité variable. Les stimuli à apprendre sont présentés à la couche d'entrée, la réponse du système est donnée par la couche de sortie. Certains réseaux possèdent des couches cachées intermédiaires.

Un neurone calcule son activation comme la somme pondérée par le poids des connexions des entrées qu'il reçoit des neurones auxquels il est relié. Une fonction de transfert appliquée à son activation la transforme en réponse. Les formes les plus fréquentes des fonctions de transfert sont linéaires, sigmoïdes, ou gaussiennes.

2. LES DIMENSIONS DES MODÈLES: CODAGE, ARCHITECTURE ET TÂCHES

Les modèles se décrivent en utilisant trois dimensions (1) le codage ou représentation de l'information (2) l'architecture neuronale (3) les tâches simulées.

2.1 Codage et Représentation des visages

Le premier problème consiste dans la définition et le choix de la représentation des visages. Trois codages se trouvent dans la littérature. Les deux premiers

décrivent l'apparence visuelle, le troisième utilise des vecteurs arbitraires ou aléatoires. Ces codages sont:

- *Codage numérique des visages.* Les images de visages sont numérisées et représentées par l'intensité lumineuse de chaque pixel. La plupart des modèles implémentent une première étape de compression comme (1) utiliser un auto-associateur linéaire (2) utiliser les valeurs de réponse des cellules de la couche cachée d'un réseau à rétro-propagation (3) représenter les visages par la réponse de filtres.
- *Description géométrique.* Les visages sont décrits par un ensemble de mesures de caractéristiques des visages (*e.g.*, longueur du nez, distance inter-pupillaire).
- *Représentation arbitraire.* Les visages sont représentés par des séries arbitraires de vecteurs binaires ou purement aléatoires.

2.2 Architecture des modèles

Quatre architectures se trouvent dans la littérature (pour plus de détails, voir par exemple, Abdi, 1994a et b, Haykin, 1994; Hertz *et al.*, 1991):

- *Auto-associateurs linéaires ou analyses en composantes principales.* Ils consistent en un ensemble d'unités linéaires complètement inter-connectées.
- *Réseaux à rétro-propagation d'erreur.* Ils comportent une couche d'entrée, une couche de sortie et une couche cachée. La fonction de transfert est non-linéaire.
- *Fonctions de rayon (radial basis functions).* Ces

réseaux récents comportent une couche cachée dont chaque cellule calcule la similitude entre la configuration présentée en entrée et un prototype.

- *Réseaux d'optimisation de fonction de coût.* C'est une famille de réseaux. Le but est de trouver un ensemble de paramètres qui minimisent une fonction de coût. Le modèle générique est le réseau de Hopfield (1982).

2.3 Tâches

Les modèles recensés effectuent ou simulent trois tâches: identification, reconnaissance, et catégorisation. L'*identification* revient à retrouver l'*identité* d'une personne à partir d'une vue du visage. C'est une tâche cognitive complexe, les modèles qui cherchent à la simuler utilisent, en général, un code arbitraire pour les visages. La *reconnaissance* d'un visage revient à décider si le visage a été vu précédemment (avec la même vue ou non). C'est une tâche de mémoire épisodique. *Catégoriser* un visage équivaut à affecter le visage à une ou plusieurs catégories. La première catégorie est celle des visages. Une catégorie populaire est le sexe (des visages). On trouve aussi, l'émotion et la race. C'est une tâche de mémoire sémantique.

3. CINQ GRANDES CLASSES DE MODÈLES

La combinaison des critères précédents définit 45 modèles possibles. Par chance, l'ensemble des modèles publiés est plus restreint. Il se réduit, de fait, à 5 groupes:

- *Auto-associateurs et analyses en composantes principales.* Le premier groupe comprend l'ensemble des modèles utilisant en entrée un code de bas niveau suivi d'un pré-traitement de l'information par un auto-associateur. Les tâches effectuées consistent en catégorisation (sexe et race) et reconnaissance.
- *Modèles à Holons et rétro-propagation de l'erreur.* Ils utilisent un code de bas niveau. Le pré-traitement de l'information est effectué par un auto-encodeur non-linéaire (*e.g.*, rétro-propagation). Les tâches consistent en identification et catégorisation (sexe et émotion). L'équivalent des composantes principales des auto-associateurs a été baptisé *Holons*.
- *Modèles à filtres.* Ils utilisent en entrée des images de visages. L'étape de pré-traitement de l'information est effectuée par filtrage. Ces modèles essaient d'identifier une personne à partir de plusieurs vues (prise avec différents point de vue). C'est un problème d'optimisation qui peut utiliser, entre autres, des réseaux de Hopfield ou des architectures à liens dynamiques.

- *Modèles géométriques.* Ils décrivent les visages en termes de caractéristiques, de distances entre caractéristiques et de rapports entre distances. Les tâches simulées reviennent à catégoriser (par sexe) et à identifier. Les architectures sont équivalentes à un perceptron ou un réseau à fonction de rayon.
- *Modèles cognitifs.* Ils modélisent des phénomènes cognitifs de haut niveau comme les effets de contexte, d'amorçage sémantique, ou de typicalité. Les visages sont représentés par des vecteurs arbitraires. Les architectures comprennent les réseaux à couches cachées et rétro-propagation de l'erreur, les auto-associateurs non-linéaires, les réseaux à compétition inter-active (réseaux IAC), CHARM et MINERVA.

Pour chaque modèle nous suivrons la même présentation: une brève description suivie des principaux résultats et problèmes.

4. Auto-associateurs et ACP

Ces modèles, dérivés des travaux de Kohonen (1977) et Anderson *et al.* (1977) sur les mémoires auto-associatives (*i.e.*, les stimuli présentés sont associés à eux-mêmes), sont composés de neurones linéaires (*i.e.*, dont la réponse est une fonction linéaire de l'activation), et fonctionnent comme des mémoires auto-adressables (Ils peuvent reconstituer une image complète à partir d'une partie présentée en entrée). Chaque neurone est relié à tous les autres neurones de la mémoire par des connexions modifiables par apprentissage.

La première étape (voir figure 1) numérise une image du visage qui est ensuite transformée en vecteur par "dépliage". Chaque élément du vecteur code l'intensité lumineuse d'un pixel qui correspond à un neurone. Pour stocker des visages on modifie les connexions entre neurones proportionnellement à la corrélation entre pixels: c'est l'apprentissage de Hebb. Pour rappeler un stimulus, on présente en entrée une image numérisée et on laisse chacune des cellules calculer sa nouvelle réponse en fonction de l'activation reçue des autres cellules. La qualité du rappel s'évalue par le cosinus (*e.g.*, coefficient de corrélation) entre l'entrée et la réponse. Plus le cosinus est proche de 1 plus il est probable que l'image a été apprise. Certains vecteurs possèdent l'intéressante propriété de donner effectivement un cosinus de 1: ils restent inchangés lorsqu'ils sont rappelés par la matrice, à l'exception d'une multiplication par une constante qui correspond à une amplification par la mémoire. Ils s'appellent les *vecteurs propres* de la matrice de connexions, le coefficient d'amplification s'appelle la

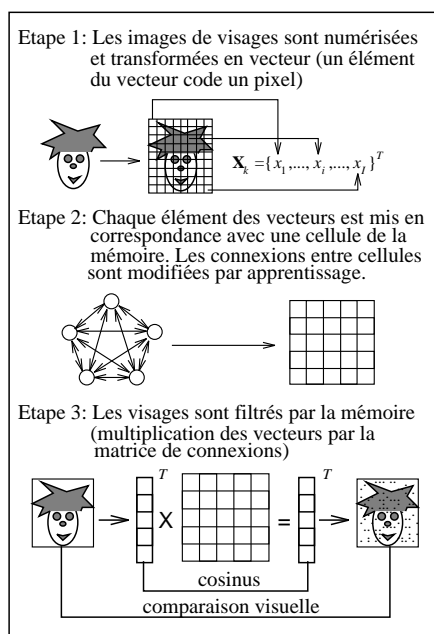


FIG. 1. — Les principales étapes du stockage et la récupération des visages avec un autoassociateur

valeur propre. Psychologiquement, ils correspondent à des *prototypes* (ils déclenchent des fausses alarmes avec certitude maximale).

Les vecteurs propres s'interprètent aussi comme des *macro-caractéristiques*: chaque stimulus appris peut se reconstituer par une somme pondérée (*i.e.*, une combinaison linéaire) des vecteurs propres. Chaque visage stocké peut, donc, se représenter par l'ensemble des coefficients de sa reconstruction. Du point de vue statistique, ceci revient à effectuer l'analyse en composantes principales (ACP) des visages et à en calculer les projections sur les composantes principales (*i.e.*, les vecteurs propres sont les composantes principales). Les vecteurs propres s'ordonnent en fonction de la valeur propre (le premier vecteur propre a la plus grande valeur propre). Pour l'ACP, la valeur propre donne la part de variance expliquée par la composante principale. Les vecteurs propres constituent l'outil essentiel d'analyse des mémoires auto-associatives linéaires.

On peut améliorer la performance avec une procédure d'apprentissage utilisant l'erreur entre la réponse donnée et celle désirée. L'erreur est calculée par chaque cellule qui modifie ensuite la valeur de ses connexions de manière à diminuer l'intensité de l'erreur lors d'une présentation subséquente du stimulus. C'est la règle d'apprentissage de *Widrow-Hoff* ou règle *Delta*. En termes de vecteurs et valeurs propres, la règle d'apprentissage égalise les valeurs propres: la matrice de connexions est dite *sphérisée*.

Lorsque la procédure d'apprentissage converge, les visages stockés sont parfaitement reconstitués. Les visages inconnus sont déformés par la mémoire en fonction inverse de la ressemblance d'ensemble avec les visages appris (voir figure 2).

4.1 Mémoires auto-associatives: principaux résultats

Kohonen (1977) utilise (incidemment) des visages pour montrer, avec une figure classique, qu'une mémoire peut reconstituer un visage à partir d'un fragment donné en indice. O'Toole *et al.* (1988) montrent qu'un auto-associateur peut reconnaître des visages même après filtrage en fréquences spatiales.

Les modèles plus récents utilisent vecteurs et valeurs propres comme outils d'analyse. Sirovich et Kirby (1987) indiquent que les vecteurs propres d'une matrice de connexions de visages ressemblent à des visages. Ces auteurs cherchent simplement un moyen de compresser l'information en représentant des visages par leur projection sur leurs vecteurs propres. Ils montrent que 115 visages peuvent se reconstruire à partir de leurs 40 premiers vecteurs propres avec une qualité moyenne de 97%.

Abdi (1988; *cf.* Abdi *et al.* sous presse) utilisant une généralisation des mémoires auto-associatives équivalente au modèle linéaire général, confirme que les vecteurs propres (généralisés) préservent l'aspect général des visages. Il indique que les deux premiers vecteurs propres séparent spontanément les



FIG. 2. — Réponses d'un auto-associcateur (entraîné avec des visages caucasiens) à des visages présentés en entrée (1-ère ligne) après apprentissage avec la loi de Hebb (2-ème ligne) et la loi Delta (3-ième ligne). Le premier visage a été appris ($r_H^2 = .71$; $r_D^2 = 1.00$). Le second visage n'a pas été appris et provient de la même population que les visages appris ($r_H^2 = .66$; $r_D^2 = .85$). Le troisième visage provient d'une autre population que les visages appris ($r_H^2 = .10$; $r_D^2 = .71$). La quatrième image se compose de points aléatoires ($r_H^2 = .00$; $r_D^2 = .00$). r_H^2 , et r_D^2 sont le carré du coefficient de corrélation entre l'image d'origine et sa reconstruction après apprentissage avec la loi de Hebb ou Delta.

visages masculins et féminins et que le premier vecteur propre joue le rôle de détecteur de visage (mais ne donne pas d'analyse statistique). En outre, il montre que les visages masculins sont moins dispersés que les féminins et donc qu'une décision serait biaisée vers les visages masculins.

Plus récemment, Turk et Pentland (1991) utilisent les vecteurs propres obtenus à partir d'une vue de 16 personnes pour les identifier à partir de nouvelles vues obtenues avec un éclairage différent, avec la tête plus ou moins penchée, ou avec une prise de vue plus ou moins éloignée de la prise de vue originale (*i.e.*, la tête occupait une superficie différente de l'image originale). Ils utilisent les projections des visages (après soustraction du visage moyen de chaque visage) sur les 8 premiers vecteurs propres. Le système se révèle peu sensible au changement d'éclairage (96% de reconnaissance correcte), quelque peu sensible au changement d'orientation (85%) et assez sensible au changement de taille (64%). Ils indiquent, également, que les premiers vecteurs propres peuvent jouer le rôle de détecteurs de visage dans une photographie. Pour ce faire, ils convolvent les premiers vecteurs propres avec l'image et varient la taille des vecteurs propres. Une corrélation importante indique la présence et la position d'un visage de la taille des vecteurs propres dans l'image (les auteurs, néanmoins, ne donnent pas d'indication sur les performances de la technique).

Le rôle des vecteurs propres comme macro-caractéristiques a été exploré récemment par plusieurs

auteurs. Ainsi, O'Toole *et al.* (1993) examinent le type d'information fourni par différents ensembles de vecteurs propres. Ils suggèrent une dissociation du type d'information: les premiers vecteurs propres tendent à être riches en fréquences spatiales basses (en partie parce qu'elles contiennent beaucoup d'énergie ce qui se reflète dans une valeur propre élevée). Ils sont statistiquement très stables et s'estiment avec un petit nombre de visages (*cf.* Valentin *et al.* sous presse). Si un ensemble de visages hétérogènes (*i.e.*, visages masculins et féminins ou de plusieurs races) est utilisé, les premiers vecteurs propres permettent d'effectuer des classifications dérivées du code visuel comme le sexe ou la race (O'Toole, Abdi *et al.* 1991). En revanche, les derniers vecteurs propres sont instables, riches en fréquences spatiales hautes et donnent des informations spécifiques aux visages de l'échantillon.

Quelques vecteurs propres sont illustrés figure 3. Traditionnellement, pour une ACP on ne garde ou n'interprète que les premiers vecteurs propres. Mais ce serait une erreur pour les visages car les derniers vecteurs propres codent l'identité. Ainsi, les auto-associcateurs retrouvent spontanément des dissociations suggérées par le modèle de Bruce et Young (1986).

La figure 4 illustre la dissociation entre information sémantique et spécifique. Bien que la reconstruction à partir des 30 premiers vecteurs propres explique 90% de la variance de l'image, elle ne permet pas de reconnaître la personne, mais permet de

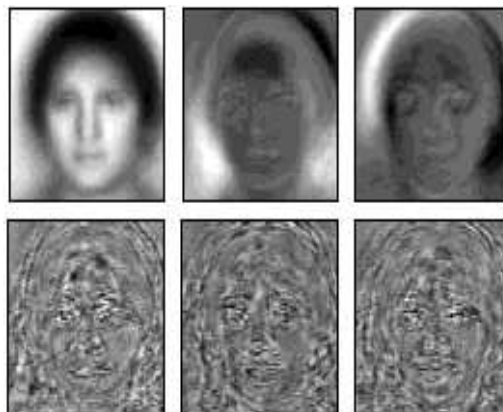


FIG. 3. — Les trois premiers et les trois derniers vecteurs propres d'une matrice de connexions d'un auto-associeur



FIG. 4. — Reconstruction d'un visage (à gauche) avec les 30 premiers vecteurs propres (au centre) et avec les vecteurs propres 31 à 160

catégoriser l'image comme un visage et d'en deviner le sexe. Par contraste, la reconstruction du visage à partir des vecteurs propres de 31 à 160 n'explique que 10 % de la variance de l'image mais permet l'identification.

De fait, les vecteurs propres peuvent s'interpréter comme des filtres (*i.e.*, de Wiener) à travers lesquels les nouveaux visages sont perçus. Les vecteurs propres reflètent les caractéristiques statistiques de la population dont ils sont extraits et dont ils permettent de reconstruire les éléments de manière optimale. S'ils sont utilisés pour représenter les visages d'une autre race, toutefois, ils ne sont plus optimaux, et la mémoire tendra à déformer les visages de l'autre race et à en rendre plus difficile la discrimination (*cf.* figure 2): c'est "l'effet autre race" (O'Toole, Deffenbacher *et al.* 1991).

Une des tâches favorites des réseaux consiste à catégoriser des visages par sexe, probablement parce que les sujets humains montrent un niveau de performance élevé pour cette tâche d'une claire importance pour la survie de l'espèce. Par exemple, Bruce *et al.* (1993) indiquent que leurs sujets devinent correctement le sexe de 96% des visages qui leur sont présentés (bien que les cheveux soient cachés par un bonnet). Récemment, Abdi *et al.* (soumis) évaluent la performance d'un auto-associeur pour 160 visages. Ils testent 4 modèles: (1) les visages sont

représentés par leur projection sur les 160 vecteurs propres et la catégorisation s'effectue par un perceptron; (2) les visages sont représentés par leur vecteurs propres et la catégorisation s'effectue par un réseau à fonction de rayon; (3) les visages sont représentés par leurs 33975 pixels et la catégorisation s'effectue par un perceptron; (4) les visages sont représentés par leurs 33975 pixels et la catégorisation s'effectue par un réseau à fonction de rayon. La performance s'évalue avec une procédure de "Jackknife" (*i.e.*, 159 visages sont utilisés pour prédire le 160-ième), qui permet d'estimer la capacité du réseau à généraliser à de nouveaux visages en provenance de la même population. Les auteurs comparent cette performance à un apprentissage non-supervisé ("nuées-dynamiques") obtenant un taux de succès de 77% (87% pour les visages masculins *vs.* 66% pour les féminins). Le réseau utilisé pour la décision n'a pas d'influence, mais le pré-traitement est décisif. Les modèles 1, 2, 3 et 4 obtiennent respectivement 92%, 91%, 79% et 80% de catégorisations correctes pour des nouveaux visages.

Les auto-associeurs linéaires sont les modèles les plus utilisés et les plus explorés. Ils apparaissent, de ce fait, comme les modèles les plus puissants. Ils doivent leur performance à leur puissance numérique: les algorithmes (*i.e.*, extraction de vecteur propres) qui permettent de les simuler sont bien étudiés et

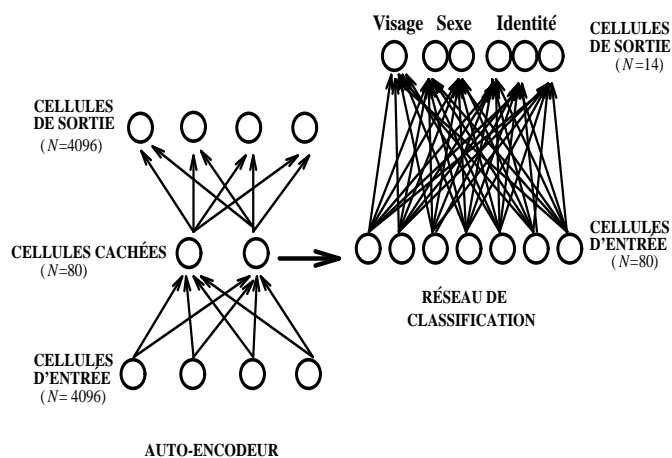


FIG. 5. — Architecture du modèle de Cottrell et Fleming (1990)

optimisés. Par conséquent, ils peuvent prendre en compte de larges ensembles de visages décrits par un grand nombre de pixels. Les effets obtenus découlent directement de la préservation de la structure statistique des visages. Ces modèles rendent compte de manière satisfaisante d'effets cognitifs concernant les visages non-familiers, mais semblent avoir des problèmes avec la reconnaissance de visages présentés avec de nouvelles orientations (*e.g.*, transfert de face à profil). Des résultats récents mais encore préliminaires (Valentin & Abdi, 1994) indiquent que ces problèmes ne sont peut-être pas si importants. En effet, ces auteurs rapportent qu'un auto-associateur ayant appris un nombre suffisant de vues de plusieurs visages arrive à les reconnaître présentés sous un nouvel angle et montre une performance optimale pour les vues de 3/4 face (comme les sujets humains). Le problème de présenter les visages tous à la même échelle reste entier, toutefois. La solution de Turk et Pentland (1991) est trop lourde du point de vue calculatoire et difficile à justifier comme mécanisme cognitif.

5. MODÈLES À HOLON ET RÉTRO-PROPAGATION

Ces modèles codent les visages en entrée par des pixels. En général, la rétro-propagation d'erreur est utilisée pour créer une première étape de compression de l'information. L'idée de base vient de Cottrell *et al.* (1987). Leur réseau comporte 64 cellules en entrée et en sortie, et 16 pour la couche cachée. Le réseau, appelé *auto-encodeur*, associe une image en entrée à elle-même. Comme le nombre de cellules de la couche cachée est plus petit que celui des autres couches, elle doit compresser l'information. Comme l'apprentissage par rétro-propagation minimise la somme des carrés d'erreur entre le signal de sortie et le signal désiré, il optimise le même critère qu'un auto-associateur linéaire et on peut s'attendre

à ce que le comportement des deux réseaux soit assez proche. Cottrell *et al.* notent, en effet, que les 16 cellules de la couche cachée retrouvent le même espace sphérisé que les 13 premiers vecteurs propres de l'auto-associateur linéaire. La réponse des cellules de la couche cachée est utilisée ensuite comme entrée pour d'autres réseaux qui implémentent une tâche de décision.

5.1 Rétro-propagation: principaux résultats

Cottrell et Fleming (1990; Fleming & Cottrell, 1990) utilisent un auto-encodeur suivi d'un réseau classifieur pour identifier et catégoriser des visages. Le réseau comporte 4096 cellules en entrée et en sortie et 80 cellules cachées dont la réponse est ensuite utilisée comme entrée d'un hétéro-associateur avec des cellules de sortie correspondant à un détecteur de visages, à un détecteur de sexe (du visage) et à une série de cellules "grand-mère" pour l'identité de chaque visage appris (voir figure 5). Ils entraînent leur réseau avec un ensemble de 64 images de 11 visages, et 13 images de "non-visages".

La performance est testée avec des vues nouvelles et anciennes des visages appris, des vues de nouveaux visages, des vues dégradées de visages connus et inconnus et des images de "non-visages". Le modèle reconnaît et catégorise parfaitement les vues apprises. Il généralise à 97% sa performance pour identifier les nouvelles vues des visages appris. Les nouveaux visages sont parfaitement catégorisés en tant que visage. La prédiction du sexe est moins bonne (63% de succès, toutes les erreurs portant sur les visages féminins). Le modèle est robuste aussi quant à la dégradation de l'image et au changement d'illumination.

Lorsque le type d'information élaboré par les cellules cachées est exploré en activant une seule d'entre elles, le réseau donne des images proches des images de vecteurs propres des auto-associateurs. Cot-

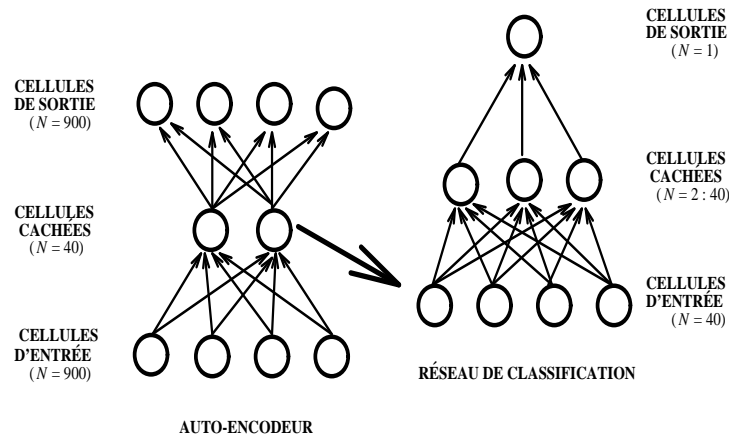


FIG. 6. — Architecture du modèle de Golomb *et al.* (1991)

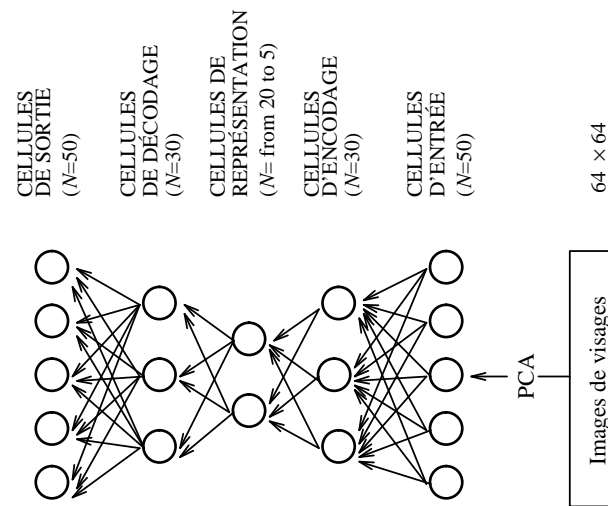


FIG. 7. — Architecture du modèle de DeMers et Cottrell (1993)

trell et Fleming leur donnent le joli nom de *Holons*.

La catégorisation par sexe est aussi exploré par Golomb *et al.* (1991) avec SEX-NET (un réseau similaire à celui de Cottrell et Fleming, *cf.* figure 6). Dans un premier temps, ils compressent des images composées de 900 pixels en les représentant par leurs valeurs pour 40 cellules cachées d'un auto-encodeur. Ces valeurs sont ensuite envoyées en entrée d'un réseau à rétro-propagation avec 10 cellules cachées et une cellule de sortie qui doit deviner le sexe du visage. Ils entraînent le réseau avec 80 visages (pour moitié féminins) puis testent la performance avec 8 ensembles de 10 nouveaux visages. Le réseau catégorise les nouveaux visages avec un taux de succès de 92%.

Cottrell et Metcalfe (1991) utilisent une stratégie proche de Cottrell et Fleming pour la catégorisation d'émotions en plus du sexe et de l'identité de 20 personnes photographiées avec 8 expressions différentes. Le sexe, l'identité des vues appris, les

émotions positives sont correctement devinées, mais le réseau n'arrive pas à distinguer les émotions négatives les unes des autres.

Les réseaux précédents n'ont qu'une couche cachée qui fonctionne comme une ACP, et ne semblent sensibles qu'aux liaisons linéaires. Des architectures plus sophistiquées devraient pouvoir détecter des relations non-linéaires. DeMers et Cottrell (1993) développent cette idée et construisent un réseau dédié à la "compression dimensionnelle non-linéaire" dont le détail est présenté figure 7. Ils utilisent les visages de Cottrell et Metcalfe. Les visages sont tout d'abord représentés par leur projection sur leurs 50 premiers vecteurs propres. Le réseau de compression proprement dit comporte une couche d'entrée et de sortie de 50 cellules chacune et trois couches cachées. La première couche cachée de 30 cellules est une couche d'encodage, la deuxième, de 20 à 5 unités suivant les simulations, s'interprète comme une couche de représentation, la troisième couche de 30 cellules est

une couche de décodage. Dans un premier temps, le réseau apprend à auto-associer les visages par rétro-propagation. Puis la seconde étape réduit le nombre de cellules de la couche de représentation en éliminant les cellules avec la variance d'activation la plus faible, puis le réseau est ré-entraîné. Le réseau arrive à reconstituer de manière satisfaisante des visages avec une couche de représentation ne comportant que 5 cellules. Bien que ces cellules soient difficiles à caractériser, elles semblent capturer une part importante d'information. Les auteurs confirment ce point en utilisant 6 des 8 vues de chaque visage pour entraîner le réseau. Puis, ils utilisent la réponse des 5 cellules de représentation comme entrée d'un hétéro-associateur devant deviner le sexe et l'identité. Lorsque les 2 vues restantes de chaque visage sont utilisées, le réseau devine correctement l'identité et catégorise par sexe avec 95% de succès. Les auteurs ne testent toutefois pas la performance de leur réseau avec de nouveaux visages.

Dans l'ensemble, les modèles à rétro-propagation montrent un niveau de performance comparable ou inférieur aux auto-associateurs. Ils prennent en compte les dimensions statistiques des visages. La représentation qu'ils développent est proche des vecteurs propres de l'ACP. Leur principal défaut découle de l'utilisation de la rétro-propagation comme technique d'apprentissage: les temps de calcul sont excessivement longs et la convergence finale n'est pas assurée.

6. MODÈLES À FILTRES

Ces modèles procèdent en deux temps. La première étape pré-traite l'image par filtrage. La seconde étape équivaut à la recherche par une mémoire associative non-linéaire de la configuration la plus proche de l'image perçue. Une étape additionnelle permet de décider, en fonction de la proximité entre la configuration initiale et celle retrouvée par le système si le visage présenté doit être "reconnu" et identifié ou non.

Le groupe de von den Malsburg (*cf.* Buhman *et al.*, 1991; Lades *et al.*; 1993) utilise une approche dérivée des ondelettes (*cf.* Meyer, 1992) pour coder des visages. Ils placent des récepteurs neuronaux sur une grille rectangulaire de $8 \times 10 = 80$ sommets. Chacun des neurones agit comme un filtre sur l'image. Précisément, à chaque sommet se trouve un ensemble de 40 neurones qui chacun implément un filtrage par une fonction de Gabor (ou *ondelette de Morlet*: produit d'une Gaussienne et d'une exponentielle imaginaire) correspondant à 8 différentes orientations et 5 fréquences spatiales. Les neurones codent l'intensité de la réponse du filtre (*i.e.*, ils ignorent la phase du signal, et donc perdent la possi-

bilité de reconstituer l'image d'origine à partir des réponses du filtre). L'ensemble des valeurs des 40 neurones d'un sommet est appelé un *jet de Gabor*. Un visage est donc représenté par 80 jets de Gabor, ou par un vecteur à $80 \times 40 = 3200$ éléments. Le pré-traitement par des fonctions de Gabor est motivé par leur pertinence pour décrire les processus de bas niveau de la perception visuelle (*e.g.*, les cellules de *V1*).

Pour stocker des visages, les auteurs utilisent une mémoire auto-associative particulière: l'architecture à liens dynamiques. Pour l'essentiel, elle permet de stocker des vecteurs puis de retrouver le vecteur le plus proche de l'entrée. La recherche s'effectue par une procédure d'optimisation équivalente à la recherche du gabarit dont la déformation doit être minimale pour obtenir l'image présentée en entrée. C'est à dire que l'on cherche parmi les images stockées celle dont la somme des distortions (mesurées comme le carré d'une distance euclidienne) des 80 jets de Gabors est minimale. Suivant l'importance de la déformation, une étape subséquente décidera si l'objet perçu a été effectivement identifié.

Pour évaluer la performance, Lades *et al.* (1993) utilisent un ensemble de vues de 87 personnes, prises avec différentes expressions et avec une variation d'orientation de plus ou moins 15 degrés autour d'une vue de face. Ils reportent une performance d'identification correcte de 100% lorsque la vue spécifique du visage a été apprise, et de 85% pour les visages présentés avec une rotation de 15 degrés.

6.1 Evaluation du modèle

Les modèles à filtres sont clairement prometteurs, en partie pour la vraisemblance neuronale des filtres de Gabor. En outre, le filtrage préserve la structure statistique des visages, et donc, ces modèles devraient pouvoir atteindre un niveau de performance équivalent à celui des auto-associateurs ou des modèles à Holons. Ils souffrent des mêmes limites: nécessité de présenter les visages approximativement à la même échelle et orientation.

Le modèle de récupération de l'information (*i.e.*, l'architecture à liens dynamiques) est plus difficile à évaluer, particulièrement la qualité de sa puissance numérique comparée à des méthodes mieux connues (*e.g.*, décomposition en vecteurs propres).

7. MODÈLES GÉOMÉTRIQUES

Les modèles statistiques les plus anciens (*cf.* Samal & Iyengar, 1992, pour une revue) représentent les visages en termes de distances, angles et surfaces dérivés de caractéristiques élémentaires comme le nez ou le menton. L'identification revient à rechercher dans une base de données un visage dont la descrip-

tion correspond au vecteur d'entrée. Cette étape peut s'effectuer, en théorie, par un réseau de neurones.

Les modèles récents s'intéressent essentiellement aux tâches de catégorisation. Ainsi, Brunelli et Poggio (1992) décrivent des visages par leur valeur pour 16 caractéristiques (*e.g.*, largeur du menton, distance inter-pupillaire) extraites automatiquement de 168 images obtenues en prenant 4 vues différentes de 42 visages (dont 21 masculins). Le vecteur décrivant les visages est ensuite utilisé en entrée de 2 réseaux à fonction de rayon (*radial basis function*): l'un devant donner la réponse maximale pour des visages féminins, l'autre pour des visages masculins. Les réseaux comportent 16 unités d'entrée, une unité cachée et de sortie. Après la phase d'apprentissage, un visage est classifié en accord avec le réseau donnant la plus forte activation.

La performance est évaluée en présentant: des vues utilisées pour la phase d'apprentissage, de nouvelles vues de visages connus et des vues de nouveaux visages. Les auteurs évaluent la généralisation avec un "Jackknife". Ils rapportent un taux de catégorisations correctes de 90% pour les vues apprises; 86% pour les nouvelles vues de visages connus; et 79% pour les vues de visages inconnus. A titre de comparaison, la performance de sujets humains pour les mêmes visages (qui étaient pour eux, bien sûr, tous inconnus) atteint 90%.

Le principal problème reste la *sélection* des caractéristiques. Burton *et al.* (1993) illustrent ce problème dans une série de 5 expériences où ils examinent l'efficacité de différents ensembles de mesures pour catégoriser des visages par sexe. Ils utilisent une analyse discriminante ascendante, mais leur résultats sont comparables à un perceptron suivi d'un Jackknife. Dans un premier temps, ils dérivent 19 variables des 73 mesures effectuées manuellement (*e.g.*, taille du menton, distance des yeux aux sourcils) sur 179 visages (dont 88 féminins). L'analyse ne garde que 12 variables et atteint un taux de classifications correctes de 86% et 85% pour les visages masculins et féminins. En combinant des mesures tri-dimensionnelles et différentes transformations des mesures d'origine, les auteurs arrivent à une performance maximale de 93%, qu'ils comparent à la performance de 96% de leur sujets.

Yuille (1991) présente une extension originale présentée dans le cadre de *gabarits déformables*. L'idée de base revient à faire correspondre un gabarit élastique aux cinq caractéristiques de base d'un visage (yeux, nez, bouche, cheveux et joues). Pour chacune d'elles, l'emplacement exact et le type de gabarit (*e.g.*, une bouche souriante ou boudeuse) s'ob-

tient en cherchant celui qu'il faut déformer le moins possible pour obtenir celui représenté sur l'image. La seconde étape place les gabarits les uns par rapport aux autres en cherchant la combinaison à distance minimale de celle observée (Yuille parle de ressorts reliant les gabarits et de recherche de "tension minimale"). La minimisation est présentée dans le cadre numérique classique mais s'interprète naturellement dans le cadre connexionniste (*e.g.*, architecture à liens dynamiques *supra*). Les prédictions cognitives de ce modèle restent, pour l'essentiel, à explorer, mais devraient se révéler intéressantes (*cf.* Bartlett & Searcy, 1993)

En conclusion, Le problème essentiel reste celui de l'identification, du choix et de la pertinence psychologique des caractéristiques.

8 MODÈLES COGNITIFS

Ces modèles simulent des effets cognitifs de haut niveau: effets de contexte, amorçage sémantique, typicalité, et mémorisation de composites. Le code d'entrée est une suite de valeurs binaires arbitraires ou un ensemble de nombres aléatoires. On peut envisager que le code utilisé par ces modèles provient de modules de pré-traitement mais l'effet de la structure statistique *per se* des visages est ignoré. Comme chacun des modèles diffère pour l'architecture utilisée, ils sont présentés séparément

8.1 Face-net

Schreiber *et al.* (1991) modélisent des effets de contexte: la reconnaissance est facilitée lorsque le test a lieu dans un contexte semblable à celui d'apprentissage. Comme ils sont intéressés par l'*interaction* entre contexte et visage, ils codent la *situation* d'apprentissage: un visage dans un contexte. Le visage et le contexte sont représentés chacun par 5 blocs de 5 éléments binaires (+1 ou -1), avec la contrainte qu'un seul élément par bloc prend la valeur +1.

La figure 8 décrit l'architecture de FACENET. Le réseau comporte 4 couches avec une configuration de connexions très spécifique. La couche d'entrée comporte 50 cellules (25 pour les visages, 25 pour le contexte). La *première* couche cachée se compose de 80 cellules à fonction de transfert sigmoïde (entre +1 et -1) organisées en 25 "cellules-visages" complètement connectées à l'entrée-visage, 25 "cellules-contexte" complètement connectées à l'entrée-contexte, et 30 cellules d'association complètement connectées à l'entrée-visage et à l'entrée-contexte. La *seconde* couche cachée se compose de 20 cellules sigmoïdes recevant des connexions en provenance des cellules-visage et des cellules d'association de la première couche cachée. La couche de sortie se compose de 80 cellules sigmoïdes organisées en trois groupes:

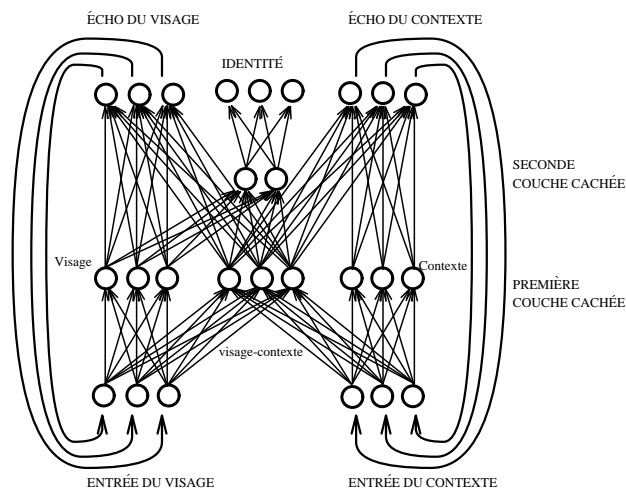


FIG. 8. — Architecture de FACENET (Schreiber *et al.* 1991)

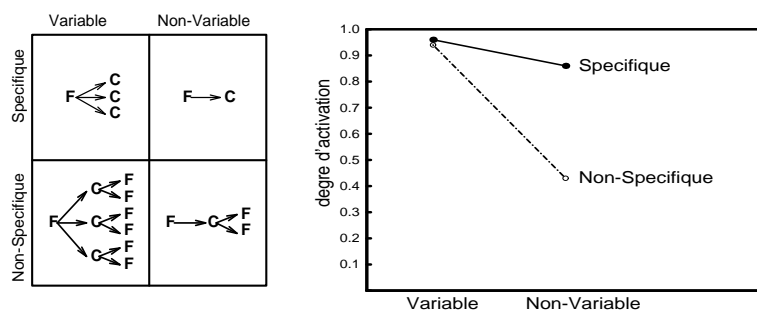


FIG. 9. — Structure des stimuli de FACENET pour la *spécificité* du contexte et la *variabilité* des visages (a) et résultats des simulations (b)

25 cellules pour l'écho-visage (connectées aux cellules-visages et d'association), 25 cellules pour l'écho-contexte (connectées aux cellules-contexte et d'association) et 30 cellules "grand-mère" pour l'identité (connectées uniquement aux cellules de la deuxième couche cachée). Durant la phase de test, la réponse des cellules écho est renvoyée à la couche d'entrée de même catégorie. La phase d'apprentissage utilise la rétro-propagation. Les parties visage et contexte agissent comme un auto-encodeur, chaque cellule d'identité correspond à un visage et agit comme un hétéro-associateur non-linéaire.

Les auteurs créent 4 conditions d'apprentissage par croisement des variables: *spécificité* du contexte et *variabilité* des visages. Un contexte est spécifique s'il est associé à un seul visage. Un visage est variable s'il est associé à plusieurs contextes (*cf.* figure 9a). les 139 stimuli, obtenus par association de 30 visages et 30 contextes sont appris. L'apprentissage terminé, les visages sont présentés avec le même contexte que celui d'apprentissage ou avec un contexte différent.

L'intensité de l'activation des cellules d'identité s'interprète comme un indice de familiarité.

Pour les visages testés dans le même contexte qu'à l'apprentissage, l'activation est la même quelle que soit la condition. Mais, lorsque les contextes diffèrent, on observe un effet de la variabilité, de la spécificité et une interaction (*cf.* figure 9b). Les auteurs notent que les visages appris avec un seul contexte non-spécifique sont plus sensibles au changement de contexte que les visages appris avec un seul contexte spécifique. Ils interprètent ce résultat contre-intuitif par un "fan-effect" qu'ils confirment par d'autres simulations (ils ne rapportent pas de données empiriques avec des sujets humains).

FACENET permet de simuler des effets de contexte. Néanmoins, son architecture complexe rend l'analyse difficile (*e.g.*, en termes de fonction de Lyapunov). C'est d'autant plus un problème que cette architecture spécifique est nécessaire à son fonctionnement. Il est possible, toutefois, qu'elle soit rendue nécessaire par l'aspect arbitraire du codage des vis-

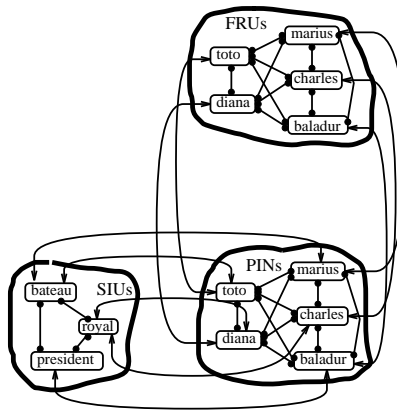


FIG. 10. — Architecture du modèle de Burton *et al.* (1990)

ages et du contexte.

8.2 FRU, SIU et PIN

Burton *et al.* (1990), Burton et Bruce (1992) et Burton *et al.* (1991) utilisent les réseaux IAC (*interactive activation and competition*) de McClelland et Rumelhart (1981) pour réaliser une version connexionniste du modèle de Bruce et Young (1986) et simuler l'identification de visages. Leur modèle comporte 4 couches de cellules (voir figure 10). La première couche d'entrée (non représentée) décrit les visages. Le modèle *stricto sensu* comporte 3 groupes d'unités:

- les FRUs (*face recognition units*) reconnaissent les visages. Elles sont activées par la présentation d'un visage connu *via* les cellules d'entrée. Elles sont "vue-indépendantes".
- les PINs (*person identity nodes*) activées par les FRUs et les SIUs, représentent les unités de reconnaissance des personnes.
- et les SIUs (*semantic information units*) activées par les PINs correspondent à l'information sémantique relative aux personnes.

Les unités à l'intérieur de chaque groupe sont reliées par des connexions inhibitrices. Les connexions bi-directionnelles entre groupes sont excitatrices. Une cellule devient active lorsque son niveau d'excitation dépasse un seuil. Les cellules à l'intérieur d'un groupe sont en compétition pour répondre à l'activation en provenance des cellules des autres groupes (*"the winner take all"*).

Les auteurs rapportent plusieurs simulations où ils mesurent le nombre de cycles d'apprentissage nécessaire pour reconnaître un visage. La reconnaissance s'effectue par les PINs: un visage est jugé fam-

ilier quand l'activation du PIN atteint un certain seuil (décidé *a priori* et constant pour tous les PINs). Les auteurs simulent des effets d'amorçage, de récupération différentielle des noms propres et de l'information sémantique, et de reconnaissance implicite. L'amorçage sémantique est simulé de la manière suivante: quand un des PINs est actif, (*toto* par exemple), son activation se propage à un SIU (*bateau* ici), qui, à son tour, active un autre PIN avec lequel il est sémantiquement relié (*marius*). La difficulté de retrouver les noms propres découle du fait que les SIUs des noms propres ne sont reliés qu'à un seul PIN (les noms propres sont uniques) alors que les autres SIUs sont reliés à plusieurs PINs et sont donc plus fortement activés. La reconnaissance implicite est simulée en affaiblissant la force des connexions entre l'ensemble des FRUs et des PINs. Ainsi, la présentation d'un visage active son FRU qui propage cette activation au PIN correspondant. A cause de la faiblesse de la connexion entre les deux cellules, l'activation du PIN est en dessous du seuil et ne permet donc pas la reconnaissance explicite du visage. En revanche, la connexion intacte entre PINs et SIUs permet encore l'amorçage.

Dans l'ensemble, les résultats ressemblent qualitativement aux phénomènes simulés. Plusieurs pistes demandent toutefois à être développées. (1) Le problème de la création des unités par le système n'est pas résolu (*i.e.*, où, quand, comment?). La représentation des visages, des noms, et de l'information sémantiques est localisée alors qu'il y a de bonnes raisons de la supposer distribuée.

8.3 Typicalité

Valentine et Ferrara (1991) simulent des effets

de typicalité: les visages typiques sont reconnus plus lentement que les visages distinctifs mais sont catégorisés comme visages plus rapidement.

Ils représentent différents types d'information par des vecteurs binaires. L'apparence du visage est codée par un vecteur de 16 composantes aléatoires; l'appartenance (ou non) à la catégorie visages par 8, 5 ou 2 éléments; l'identité du visage par 8 ou 5 éléments. Pour créer des gradients de typicalité à l'intérieur d'une catégorie, elle est représentée par un prototype. Puis, les gradients de typicalité s'obtiennent par variation du nombre d'éléments qui diffèrent du prototype.

Ils testent le rappel de la partie correspondant à l'identité et de celle correspondant à l'appartenance à la catégorie avec un auto-associateur et un réseau à rétro-propagation en fonction de la typicalité des visages utilisés en test. Pour l'auto-associateur, l'information à tester est apprise associée à l'information visuelle qui est seule présentée pour le test. La qualité du rappel est mesurée par le cosinus entre l'entrée et la sortie calculé en fonction du nombre d'itérations avec la règle Delta. Le réseau à rétro-propagation est un hétéro-associateur pour lequel l'apparence du visage constitue le stimulus. La qualité du rappel est mesurée par le carré moyen d'erreur exprimé en fonction du nombre d'époques d'apprentissage.

Les auteurs indiquent qu'ils retrouvent qualitativement les effets de typicalité avec les mesures et les architectures qu'ils utilisent.

8.4 Ja-net: visages composites et moyens

Metcalf (1994) simule des effets expérimentaux utilisant des visages pour explorer les problèmes d'agrégation ("*binding*") de caractéristiques d'un stimulus. Elle montre que des sujets, ayant vu pendant la phase d'apprentissage des visages, réagissent différemment pour la reconnaissance s'ils sont testés avec des visages réels, composites (*i.e.*, obtenus en mêlant des caractéristiques de 2 visages) ou moyens (obtenus en moyennant 2 visages pixels à pixels).

Elle reporte une série de simulations dans lesquels elle représente des visages par des vecteurs aléatoires, qu'elle combine ou moyenne. Elle simule la tâche d'apprentissage en utilisant son modèle CHARM et MINERVA (Hinztman, 1986). Dans les 2 cas, elle indique que les résultats qualitatifs obtenus avec des vecteurs aléatoires se rapprochent des résultats moyens obtenus par ses sujets.

8.5 Evaluation

En général, les auteurs retrouvent qualitativement les effets obtenus avec les sujets humains. Toutes ces simulations indiquent des pistes prometteuses. Nonobstant leur intérêt, le problème de leur perti-

nence pour la mémoire de *visages* reste entier. En effet, il est possible que les effets observés découlent du codage, de l'architecture, ou d'une subtile interaction entre les deux.

9. CONCLUSION

Malgré sa jeunesse, la psychologie cognitive des visages a engendré, déjà, une variété étonnante de modèles dont certains semblent plus proches des préoccupations de l'informatique que de la psychologie. Toutefois, le problème de la modélisation des visages *familiers* reste à explorer et devrait constituer la prochaine étape. Pour les modèles évoqués ici, le premier groupe (auto-associateur) apparaît comme le plus fructueux (non seulement parce qu'il est le favori des auteurs) du fait qu'il préserve la structure statistique des visages allié à une grande puissance calculatoire. Le développement d'architectures neuronales sophistiquées et de la puissance des moyens de calculs devrait rendre ce dernier point moins crucial dans le futur.

RÉFÉRENCES BIBLIOGRAPHIQUES

- Abdi, H. (1988). A generalized approach for connectionist auto-associative memories. In J. Demongeot (Ed.) *Artificial intelligence and cognitive sciences*. Manchester: Manchester University Press.
- Abdi, H. (1994a). *Les réseaux de neurones*. Grenoble: Presses Universitaires de Grenoble.
- Abdi, H. (1994b). A neural network primer. *Journal of Biological Systems*, 2, 247-281.
- Abdi, H., Valentin, D., Edelman, B.J., & O'Toole, A.J. (sous presse, 1995). More about the difference between men and women: Evidence from linear neural networks and the principal component approach. *Perception*, 24, 539-562.
- Abdi, H., Valentin, D., & O'Toole, A.J. (sous presse, 1996). A generalized autoassociator model for face processing and sex categorization. In D. Levine (Ed.) *Optimality in biological and artificial networks*. Hillsdale: Erlbaum.
- Anderson, J.A., Silverstein, J.W., Ritz, S.A., & Jones, R.S. (1977). Distinctive features, categorical perception, and probability learning: Some applications of a neural model. *Psychological Review*, 84, 413-451.
- Bartlett, J.C. & Searcy, J. (1993). Inversion and configuration of faces. *Cognitive Psychology*, 25, 281-316.
- Bruce, V. & Young, A. (1986). Understanding face recognition. *British Journal of Psychology*, 77, 363-383.
- Bruce, V., Burton, A. M., Dench, N., Hanna, E., Healey, P., Mason, O., Coombes, A., Fright, R., & Linney, A. (1993). Sex discrimination: How do we tell the difference between male and female faces?, *Perception*, 22, 131-152.

- Brunelli, R. & Poggio, T. (1992). HyperBF Networks for gender classification. *Proceedings of the Image Understanding Workshop*, DARPA, San Diego.
- Buhmann, J., Lange, J., v.d. Malsburg, C., Vorbrüggen, C., & Würtz, R.P., (1991). Object recognition with Gabor functions in the dynamic link architecture. In B. Kosko (Ed.) *Neural Networks for Signal Processing*. Englewood Cliffs, NJ: Prentice Hall.
- Burton, A. M. & Bruce, V. (1992). I recognize your face but I can't remember your name: A simple explanation? *British Journal of Psychology*, *83*, 45–60.
- Burton, A.M., Bruce, V., & Johnston, R.A. (1990). Understanding face recognition with an interactive activation model. *British Journal of Psychology*, *81*, 361–380.
- Burton, A.M., Bruce, V., & Dench, N. (1993). What's the difference between men and women? Evidence from facial measurement. *Perception*, *22*, 153–176.
- Burton, A.M., Young, A.W., Bruce, V., Johnston, R.A., & Ellis, A.W. (1991). Understanding covert recognition. *Cognition*, *39*, 129–166.
- Cottrell, G.W., Munro, P., & Zipser, D. (1987). Learning internal representations of gray scale images. In *Proceeding of the Ninth Annual Cognitive Science Society Conference* Seattle, WA.
- Cottrell, G.W. & Fleming, M.K. (1990). Face recognition using unsupervised feature extraction. In *Proceedings of the International Neural Network Conference*. Dordrecht: Kluwer.
- Cottrell, G.W. & Metcalfe, J. (1991). EMPATH: Face, sex, and emotion recognition using holons. In R.P. Lippman, J. Moody, & D.S. Touretzky (Eds.), *Advances in neural information processing systems 3*. San Mateo, CA: Kaufmann.
- DeMers, M. & Cottrell, G.W. (1993). Non-linear dimensionality reduction, In S.J. Hanson, J.D. Cowan, & C.L. Giles (Eds.), *Advances in neural information processing systems 5*. San Mateo, CA: Kaufmann.
- Fleming, M.K. & Cottrell, G.W. (1990). Categorization of faces using unsupervised feature extraction. In *Proceedings of the International Joint Conference on Neural Networks, Vol. II*. San Diego, CA.
- Golomb, B.A., Lawrence, D.T., & Sejnowski, T.J. (1991). Sexnet: a neural network identifies sex from human face. In R.P. Lippman, J. Moody, & D.S. Touretzky (Eds.), *Advances in neural information processing system 3*, San Mateo, CA: Kaufman.
- Haykin, S. (1994) *Neural Networks*. New York: Macmillan.
- Hertz, J., Krogh, A., & Palmer, R.G. (1991). *Introduction to the theory of neural computing*. Reading: Addison-Wesley.
- Hopfield, J.J. (1982). Neural networks and physical system with emergent collective computational abilities. *Proceedings of the National Academy of Science, USA*, *79*, 6871–6874.
- Kohonen, T. (1977). *Associative memory: A system theoretic approach*. Berlin: Springer-Verlag.
- Lades, M., Vorbrüggen, J.C., Buhmann, J. Lange, J., v.d. Malsburg, C., Wurtz, R.P., & Konen, W. (1993). Distortion invariant object recognition in the dynamic link architecture, *IEEE Transactions on Computers* *3*, 300–311.
- O'Toole, A.J., Abdi, H., Deffenbacher, K.A., & Bartlett, J.C. (1991). Classifying faces by race and sex using an autoassociative memory trained for recognition. In K.J. Hammond & D. Gentner (Eds.), *Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.
- O'Toole, A.J., Abdi, H., Deffenbacher, K.A., & Valentin, D. (1993). A low-dimensional representation of faces in the higher dimensions of the space. *Journal of the Optical Society of America A*, *10*, 405–411.
- O'Toole, A.J., Abdi, & H., Valentin, D. (sous presse, 1995). Face recognition with a neural network. In M.A. Arbib (Ed.), *Handbook of Brain Theory and Neural Networks*. Cambridge MA: M.I.T. Press.
- O'Toole, A.J., Deffenbacher, K.A., Abdi, H., & Bartlett, J.C. (1991). Simulating the other race effect as a problem in perceptual learning. *Connection Science*, *3*, 163–178.
- O'Toole, A.J., Millward, R.B., & Anderson J.A. (1988). A physical system approach to recognition memory for spatially transformed faces. *Neural Networks*, *1*, 179–199.
- Metcalfe, J. (1994). *A computational modeling analysis of binding in human memory*. Paper presented at the meeting of the Society of Mathematical Psychology, Seattle, WA.
- Meyer, Y. (1992). *Ondelettes: Algorithmes et applications*. Paris: Armand Colin.
- McClelland, J.L & Rumelhart, D.E. (1981). An interactive activation model of the effect of context in perception. Part 1. An account of basic findings. *Psychological Review*, *88*, 375–406.
- Samal, A. & Iyengar, P.A. (1992). Automatic recognition and analysis of human faces and facial expressions: A survey. *Pattern Recognition*, *25*, 65–77.
- Schreiber, A., Rousset, S., & Tiberghien, G. (1991). Facenet: A connectionist model of face identification in context. *European Journal of Cognitive Psychology*, *3*, 177–198.
- Sirovich, L. & Kirby, M. (1987). Low-dimensional procedure for the characterization of human faces. *Journal of the Optical Society of America A*, *4*, 519–524.
- Turk, M. & Pentland, A. (1991). Eigenfaces for recognition. *Journal of Cognitive Neuro Science*, *3*, 71–86.
- Valentin, D., Abdi, H., O'Toole, A.J., & Cottrell, G.W.

- (1994). Connectionist models of face processing: A survey. *Pattern Recognition*, 27, 1208–1230.
- Valentin D., Abdi, H., O'Toole, A.J. (1994). Categorization and identification of human face images by neural networks: A review of linear auto-associator and principal component approaches. *Journal of Biological Systems*, 2, 413–429.
- Valentin, D., Abdi, H., & O'Toole, A.J. (sous presse, 1996). Principal component and neural network analyses of face images. In C. Dowling, F.S. Roberts, & P. Theuns (Eds.), *Progress in mathematical psychology*. New York: Elsevier.
- Valentin, D. & Abdi, H. (1994). *How come when you turn your head I still know who you are?* Paper presented at the meeting of the Society of Mathematical Psychology, Seattle, WA.
- Valentine, T. & Ferrara, A. (1991). Typicality in categorization, recognition and identification. *British Journal of Psychology*, 82, 87–102.
- Yuille, A.L. (1991). Deformable templates for face recognition, *Journal of Cognitive Neurosciences*, 3, 59–70.

RÉSUMÉ

Les modèles connexionistes de la mémoire des visages sont décrits suivant trois dimensions: codage et représentation de l'information; type d'architecture neuronale; et type de tâche simulée. Cinq grands groupes de modèles se dégagent: (1) les auto-associateurs linéaires opérant sur des images de visages (2) les modèles à rétro-propagation opérant sur des images de visages (3) les modèles à filtres de Gabor ou ondelettes de Morlet (4) les modèles à codage géométriques (5) les modèles cognitifs à codage arbitraire ou aléatoire.