

Article original

Un score individuel de reconnaissance d'une série de stimuli intermédiaires entre deux sources : la perception catégorielle des expressions faciales

An individual recognition score of a series of intermediate stimuli between two sources: categorical perception of facial expressions

R. Bruyer^{a,*}, P. Granato^b, J.-P. Van Gansberghe^{c,†}

^a *Unité de neurosciences cognitives, UCL, faculté de psychologie,
université de Louvain-la-Neuve, 10, place du Cardinal-Mercier, 1348 Louvain-la-Neuve, Belgique*

^b *Inserm, centre d'investigation clinique, CHRU de Lille, centre hospitalier de Valenciennes, Valenciennes, France*

^c *Consultant en analyse de données, Bruxelles, Belgique*

Reçu le 31 mai 2005 ; accepté le 20 février 2006

Résumé

La technique informatique de « morphing » permet de créer des images intermédiaires qui sont des hybrides de deux images sources, dans des proportions variables. Au cours des 15 dernières années, elle a fortement contribué aux développements de nos connaissances relatives aux mécanismes de la reconnaissance des expressions faciales. En particulier, il se confirme que la perception des expressions est catégorielle. En effet, si on fait varier de manière continue et linéaire la transition entre une expression et une autre, la variation de perception qui en résulte n'est pas linéaire, mais sigmoïdale. Le même constat s'applique à la reconnaissance des identités faciales et à la perception du genre des visages. Nous montrons ici que les données issues de tests de perception catégorielle peuvent également être utilisées pour mesurer la qualité de la reconnaissance, indépendamment de l'aspect catégoriel. Elle porte sur les réponses d'un sujet que l'on confronte à celles d'un groupe témoin. La méthode proposée permet de décider si une réponse individuelle à une image hybride est correcte ou erronée, par comparaison aux réponses d'une population de référence. Plus précisément, la proportion de sujets du groupe de référence qui a produit la même réponse que le sujet sert d'estimateur de la probabilité que le sujet soit issu d'une population saine. Cette quantification n'était pas possible lorsqu'il s'agissait d'enregistrer le choix, sans définition a priori de la nature correcte ou erronée de la réponse. Notre ajustement nous semble généralisable à toute épreuve destinée à mesurer la perception catégorielle.

© 2006 Elsevier Masson SAS. Tous droits réservés.

Abstract

Morphing is a computerized tool allowing the generation of pictures which are « hybrids » of two sources-pictures, in controlled proportions. Over the last decade, this tool did contribute greatly to the knowledge of mechanisms underlying facial expressions recognition. Especially, it appeared that categorical perception is involved. Indeed, while the physical variation from one source to the other is continuous and linear, the perceptual variation takes the form of a sigmoidal or threshold function. The same applies to facial identity recognition and to recognition of face's gender. The present paper shows that data collected to evidence categorical perception can also be used to assess accuracy of expression recognition, whatever categorical or not, and to compare, say, the performance of an individual subject to that of a control group. In short, the proportion of subjects of the control group making the same choice as the individual subject is used as an estimator of the probability that this

* Auteur correspondant.

Adresse e-mail : raymond.bruyer@psp.ucl.ac.be (R. Bruyer).

subject is issued from a normal population. This procedure furnishes criteria to decide if a given response is correct or not, which is not possible when free choices are recorded. This adaptation can be generalized to every tool designed to evaluate categorical perception.

© 2006 Elsevier Masson SAS. Tous droits réservés.

Mots clés : Perception catégorielle ; Morphing ; Expressions faciales

Keywords: Categorical perception; Morphing; Facial expressions

1. Introduction

Les mécanismes qui sous-tendent la reconnaissance des expressions faciales émotionnelles font l'objet de nombreux travaux depuis une trentaine d'années. Ces travaux portent sur le sujet sain (nouveau-né, enfant, adulte ou âgé) et des patients atteints de pathologies psychiatriques ou neurologiques, à des fins diagnostiques, de recherche clinique ou de recherche fondamentale. Les méthodes utilisées sont des tests comportementaux, le recueil de potentiels évoqués cérébraux et la magnétoencéphalographie, la stimulation magnétique transcrânienne et l'imagerie cérébrale fonctionnelle (tomographie par émission de positrons, résonance magnétique fonctionnelle).

Les développements de l'informatique, en particulier des logiciels de traitement d'images, ont permis des progrès. C'est le cas de la méthode du « morphing ». Elle permet de créer une image qui est un hybride de deux images sources (A et B) en mélangeant, en quelque sorte, les pixels de ces deux images sources : on obtient ainsi un composite artificiel qui présente des « traits communs » avec chacune des deux images sources. Cette méthode a en particulier été appliquée aux identités faciales et aux expressions faciales (Fig. 1). Elle présente l'intérêt de maîtriser les variations des contributions de chacune des deux images sources dans l'hybride qui en résulte.

Cette méthode a permis de progresser dans un ancien débat, opposant deux théories. D'une part, la théorie *dimensionnelle*,

(Plutchik, 1980) considère que les expressions sont des combinaisons variables, continues et très nombreuses de quelques dimensions de base — intensité, similarité, valence positive ou négative, etc —. D'autre part, la théorie *catégorielle*, (Ekman, 1992) considère que les expressions constituent quelques catégories émotionnelles de base — peur, colère, joie, tristesse, etc. — qui peuvent se combiner pour rendre compte de la grande variété d'expressions que nous percevons. Plusieurs études, toutes récentes (Bruyer et Granato, 1999 ; Calder et al., 1996 ; Campanella et al., 2002 ; De Gelder et al., 1997 ; Etcoff et Magee, 1992 ; Granato et Bruyer, 2002 ; Granato et al., 1996 ; Kiffel et al., 2005 ; Young et al., 1997), ont eu recours à la méthode de morphing pour construire des expressions hybrides. Il est apparu que la reconnaissance de ces émotions n'est pas une fonction linéaire de la proportion de chaque image source dans l'hybride, mais une fonction d'allure sigmoïde (ou logistique : Fig. 2). En effet, si on présente *dans un ordre aléatoire* les différents hybrides d'une même série AB, les sujets reconnaissent A sans hésitation jusqu'à une certaine proportion de contribution de la source A dans l'image, ils reconnaissent B sans hésitation à partir d'une certaine proportion de contribution de la source B, et il n'y a qu'une petite zone d'incertitude qui définit une frontière aux limites mal définies entre les deux catégories A et B. En d'autres termes, alors que la variation des stimuli est linéaire, celle de la réponse ne l'est pas. La nature catégorielle de cette perception peut ensuite être confirmée avec une autre tâche. On présente

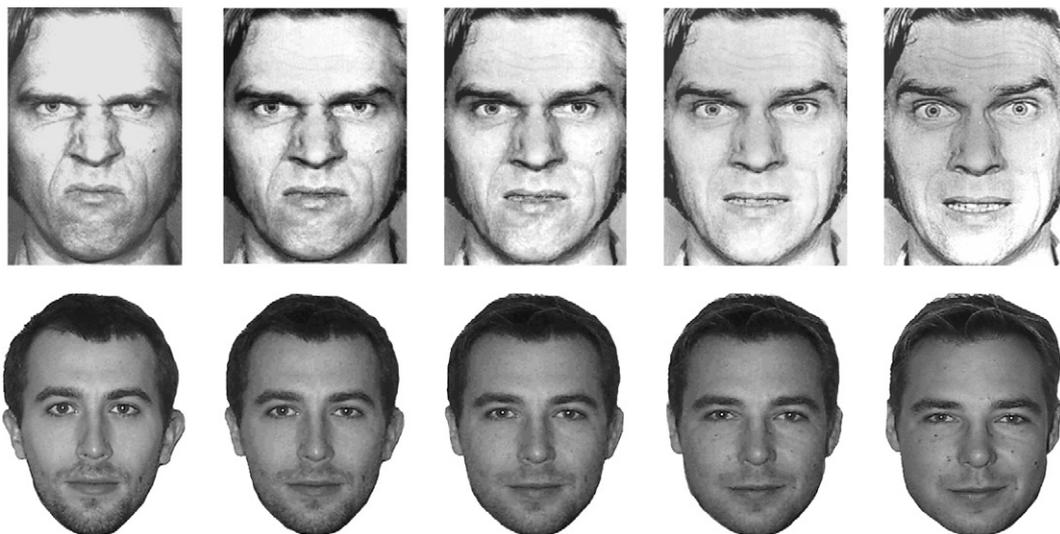


Fig. 1. La technique du morphing : ces exemples illustrent le morphing de deux expressions faciales (rangée du haut) et de deux identités faciales (rangée du bas). Les images sources occupent les extrémités gauche et droite, et les images intermédiaires sont des hybrides où le « poids » de l'image source de droite est manipulé (de gauche à droite : 0, 30, 50, 70 et 100 %). Les images sources des expressions (dégoût et peur) sont reproduites de Ekman et Friesen, 1976, avec la permission des auteurs.

au sujet des paires d'hybrides issus de la même série et il s'agit de décider si les deux images sont strictement identiques ou non ; en réalité, elles le sont dans la moitié des cas et, lorsqu'elles diffèrent, elles proviennent toutes deux du même côté de la frontière catégorielle (25 % des cas), ou pas (25 % des cas) : par exemple, si la frontière perceptive est située à hauteur de l'image comprenant 60 % de contribution de B (et 40 % de A), les images comprenant 30 et 50 % de B sont situées du même côté de la frontière, tandis que les images comprenant 50 et 70 % de B sont situées de part et d'autre de la frontière. On observe que les erreurs sont plus nombreuses et les réponses plus lentes si les images proviennent du même côté que si elles sont issues chacune d'une « catégorie », alors que la différence physique entre les deux images, en proportion de A (par exemple), est strictement identique dans les deux cas. Ces propriétés classiques de la perception catégorielle (Harnad, 1987) étaient bien connues pour des stimuli « simples » comme la longueur d'onde de la lumière qui en détermine la couleur (Bornstein et al., 1976) et le voisement des phonèmes (Liberman et al., 1957) : on montre donc à présent que la perception de stimuli visuels complexes multidimensionnels obéit aux mêmes principes, qu'il s'agisse de l'expression faciale (ci-dessus), de l'identité faciale (Beale et Keil, 1995 ; Campanella et al., 2003 ; Kiffel et al., 2005 ; Levin et Beale, 2000 ; Stevenage, 1998) ou du genre d'un visage (Bülthoff et Newell, 2004 ; Campanella et al., 2000 ; Campanella et al., 2001).

Ce type d'épreuve peut servir à examiner finement les propriétés de la perception des expressions faciales chez un sujet donné (un patient cérébrolésé atteint d'un déficit sélectif, un patient psychiatrique, etc.). Le but est de comprendre la nature du trouble éventuel et d'examiner si la perception est bien catégorielle chez le patient (De Gelder et al., 1997). Toutefois, on ne souhaite pas toujours procéder à une analyse aussi fine, mais simplement disposer, en routine clinique, d'un test normalisé, réalisable au lit du patient, sensible et fidèle, permettant d'apprécier, d'une manière globale, la qualité de la reconnaissance des expressions émotionnelles chez un sujet particulier. Nous proposons dans cet article que les séries émotionnelles utilisées pour explorer la perception catégorielle peuvent également servir à dériver un score global pour un sujet donné, qui

peut être comparé au score global d'une population « idéale » (c'est-à-dire qui ne commet aucune erreur).

Nous prendrons pour exemple l'outil « MARIE » décrit initialement par Granato et al., (1996) puis utilisé par Granato et Bruyer (2002) ainsi que Bruyer et Granato (1999), et pour lequel nous disposons à présent des résultats de sept groupes de sujets sains (de 20 à 70 ans) qui feront prochainement l'objet d'une publication. Il est important de noter que l'objectif du présent article n'est pas de proposer des normes permettant un quelconque diagnostic : en aval de cette étape normative, nous proposons plutôt d'abord un mode de calcul du score du sujet pour une série donnée d'hybrides compte tenu de la nature particulière d'une telle série.

2. Méthode

2.1. Sujets

Les données ont été recueillies auprès de sept groupes de sujets :

- 21–30 ans ($n = 30$) ;
- 31–40 ans ($n = 30$) ;
- 41–50 ans ($n = 30$) ;
- 51–55 ans ($n = 30$) ;
- 56–60 ans ($n = 30$) ;
- 61–65 ans ($n = 30$) ;
- 66–70 ans ($n = 24$).

Nous avons rétréci les tranches d'âge entre 50 et 70 ans, dans la mesure où les effets cognitifs du vieillissement se manifestent au-delà de 50 ans. Dans chaque groupe, il y avait 50 % de femmes et 50 % d'hommes. Tous les sujets étaient droitiers et avaient une vision normale ou corrigée. Des antécédents de troubles neurologiques ou psychiatriques constituaient un motif d'exclusion. Les groupes étaient comparables en termes de niveau socioculturel (scolarisation).

2.2. Matériel

« MARIE » est l'abréviation de « Méthode d'Analyse et de Recherche de l'Intégration des Émotions ». Le détail est décrit dans les publications qui viennent d'être mentionnées. Il s'agit de présenter, sur un écran d'ordinateur, des séries de 19 images dont deux sont les images sources (A et B), toujours présentées en fin de test, et 17 hybrides de ces images sources, présentés dans un ordre aléatoire défini a priori. La taille des images est 10×18 cm et elles sont présentées à une distance d'environ 40 cm.

Le logiciel contient 63 séries de ce type, qui sont les 21 combinaisons possibles entre la joie, la tristesse, la peur, la colère, la surprise, le dégoût et l'expression neutre, exprimées par trois personnages différents (« la blonde », « la brune », « l'homme »). Les images sources sont empruntées à Ekman et Friesen (1976), avec leur permission.

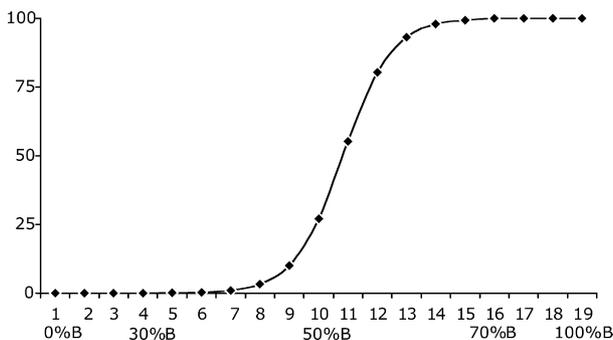


Fig. 2. La perception catégorielle : cette courbe représente le pourcentage de choix « B » des sujets en fonction de la contribution du visage source B dans l'image. On constate qu'alors que la contribution de B est linéaire, le choix des sujets est sigmoïdal plutôt que linéaire.

Les 19 images de chaque série varient par la contribution en pixels de l'image source B, à savoir : 0, 10, 20, 30, 35, 38, 41, 44, 47, 50, 53, 56, 59, 62, 65, 70, 80, 90 et 100 %. Il s'agit donc d'une distribution symétrique dont le pas est plus petit autour de la région a priori critique, à savoir 50 %. Nous voulions de la sorte pouvoir examiner plus finement les réponses des sujets autour de la frontière catégorielle, même si ce n'est pas véritablement l'objet de la présente étude.

Les données de la présente étude ont été recueillies pour 27 de ces 63 séries. Ces séries sont, pour chacun des trois personnages, dérivées des paires colère-peur, colère-tristesse, joie-tristesse, neutralité-colère, neutralité-dégoût, neutralité-joie, neutralité-peur, neutralité-surprise et neutralité-tristesse. On observera que trois séries sont bipolaires et relient une expression à une autre, tandis que les six autres sont unipolaires et relient la neutralité à une expression. Ces séries évaluent donc la perception de l'intensité d'une expression (Hsu et Young, 2004). On notera cependant que des auteurs estiment que la neutralité est elle-même une véritable expression (Shah et Lewis, 2003).

2.3. Procédure

Le sujet doit identifier chaque image en choisissant entre les deux alternatives possibles (A ou B ; par exemple : joie ou tristesse) auxquelles correspondent les boutons gauche et droit de la souris. Les deux labels du choix apparaissent à gauche et à droite du stimulus.

2.4. Mesures

Chaque sujet est exposé une seule fois à chacune des $27 \times 19 = 513$ images. La variable dépendante est donc, par série, la proportion p de sujets qui ont choisi la réponse « B » (par exemple « peur » pour la série « colère-peur »). Le choix étant forcé, il en découle que la proportion de réponses « A » vaut $1 - p$.

Trois principes seront sous-jacents au calcul du score global par série. Le premier consiste à *définir l'erreur et la réponse correcte*. En effet, rien n'autorise de décider a priori que la réponse « A » (ou « B ») est correcte pour toutes les images où la contribution de A (ou B) est supérieure à 50 %, et que toute autre réponse est une erreur, car c'est la perception par le sujet qui est mesurée. Les notions d'erreurs et de réponses correctes sont donc à comprendre comme des réponses minoritaires et majoritaires. Nous définissons en effet comme correctes les réponses individuelles qui correspondent à celles de la majorité (plus de 50 %) des réponses du groupe : par exemple, pour une image donnée, si 65 % des réponses du groupe expriment le choix A, c'est « A » qui sera considéré comme une réponse correcte et « B » comme une erreur. Le second principe consiste à *grader l'ampleur* de l'erreur ou de la réponse correcte. En effet, intuitivement, on conçoit facilement que répondre « A » est plus aisé — et répondre « B » plus difficile — si l'image est composée à 90 % de A et 10 % de B que

si elle est composée respectivement de 60 % de A et 40 % de B. Enfin, le troisième principe consiste à *définir le score global* du sujet, exprimé en pourcentage du maximum possible.

3. Résultats

On trouvera en annexe la distribution des choix par groupe, paire d'expressions et personnage, pour chacune des 19 images de chaque série. Pour la suite de l'exposé nous prendrons, en guise d'illustration, les résultats du groupe 41–50 ans auquel on a présenté la série « colère-tristesse » de la blonde.

3.1. Interprétation initiale de la réponse

Les données brutes recueillies par l'ordinateur portent sur la réponse émise par le sujet : « A » ou « B ». Le Tableau 1 présente le nombre de sujets ayant choisi B. La première étape consiste donc à traduire cette réponse en « correcte » ou « erronée » (dans le sens défini précédemment).

Comme indiqué plus haut, cette catégorisation se fait suivant le choix majoritaire. La réponse est correcte si elle correspond au choix de plus de 50 % des sujets du groupe correspondant à l'âge du sujet, et erronée dans le cas inverse. Ainsi, si un sujet de 48 ans répond « A » à l'image de la série colère-tristesse chez la blonde où la contribution de B est de 62 %, cette réponse sera qualifiée d'erreur puisqu'elle correspond au choix d'un seul ($30 - 29 = 1$) des 30 sujets du groupe témoin (soit 3,3 %). En revanche, si le sujet répond « A » à l'image de la même série où la contribution de B est de 38 %, sa réponse sera qualifiée de correcte car elle correspond au choix de 80 % des sujets du groupe ($30 - 6 = 24$).

Le point important est donc que ce n'est pas la nature de l'image (plus ou moins de 50 % de contribution de l'image B dans l'hybride) qui est utilisée comme critère, mais la perception par le groupe de référence. Pour la série prise comme exemple, on notera d'ailleurs (Tableau 1) qu'une majorité de réponses « B » est atteinte alors que la contribution de B dans l'hybride n'est que de 44 % : le seuil perceptif est différent du seuil physique. De toute manière, au même titre que l'analyse des réponses « B », le choix des expressions de référence A et B est arbitraire, puisque les images sont présentées dans un ordre aléatoire : la série colère (A)–tristesse (B) aurait pu être la série tristesse (A)–colère (B).

3.2. Pondération de l'erreur et de la réponse correcte

Dans l'exemple d'erreur qui vient d'être utilisé, il s'agit d'une erreur sévère. En effet, notre sujet de 48 ans produit une réponse différente de celle de 96,7 % des sujets du groupe témoin : il s'agissait d'une image « facile ». En revanche, s'il avait répondu « A » à l'image de la même série où la contribution de B est de 44 %, c'eût été une erreur légère, puisque 19 des 30 sujets seulement (63,3 %) ont répondu B : il s'agit d'une image « difficile ».

Tableau 1

Résultats et premières étapes du traitement des résultats pour le groupe 41–50 ans auquel on a présenté la série « colère–tristesse » de la blonde

Image	Contribution de B dans l'image (%)	Nombre de sujets (N)	Nombre de réponses « B »	Réponse majoritaire	Proportion de choix majoritaires	Proportion corrigée	Odd ratio de la réponse minoritaire
1	0	30	0	A	1,00	0,9833	0,0169
2	10	30	0	A	1,00	0,9833	0,0169
3	20	30	1	A	0,97	0,9667	0,0345
4	30	30	2	A	0,93	0,9333	0,0714
5	35	30	1	A	0,97	0,9667	0,0345
6	38	30	6	A	0,80	0,8000	0,2500
7	41	30	2	A	0,93	0,9333	0,0714
8	44	30	19	B	0,63	0,6333	0,5789
9	47	30	23	B	0,77	0,7667	0,3043
10	50	30	25	B	0,83	0,8333	0,2000
11	53	30	26	B	0,87	0,8667	0,1538
12	56	30	23	B	0,77	0,7667	0,3043
13	59	30	26	B	0,87	0,8667	0,1538
14	62	30	29	B	0,97	0,9667	0,0345
15	65	30	30	B	1,00	0,9833	0,0169
16	70	30	30	B	1,00	0,9833	0,0169
17	80	30	30	B	1,00	0,9833	0,0169
18	90	30	30	B	1,00	0,9833	0,0169
19	100	30	30	B	1,00	0,9833	0,0169

Rappelons que le sujet produit une réponse parmi deux possibles, et que la réponse est obligatoire. Il faut donc à présent pondérer l'ampleur de l'erreur ou de la réponse correcte d'un sujet particulier par le choix majoritaire de son groupe témoin. Nous allons tenir compte de la proportion de sujets du groupe qui a produit la même réponse que le sujet. Si c'est la réponse minoritaire, cette proportion sera considérée comme un estimateur de la probabilité de produire cette réponse sous l'hypothèse que le sujet est « normal ». Une correction sera toutefois apportée pour les proportions extrêmes (0 et 100 % : unanimité du groupe), pour trois raisons. Tout d'abord, par définition, des probabilités nulles ou égales à 1 sont impossibles, a fortiori lorsqu'elles servent d'estimateurs. Ensuite, des valeurs nulles ou égales à 1 poseront des difficultés lorsque nous calculerons un score global pour la série (voir plus loin) : en effet, il s'agira de calculer le produit des 19 probabilités de la série (si une seule est nulle, le produit sera nul) ou la somme de leurs logarithmes (le logarithme de 0 est impossible). Enfin, comme on va le voir, il s'agit de probabilités faibles : leur produit sera donc une valeur fort proche de zéro. Pour ces trois raisons, la transformation suivante sera appliquée aux cas extrêmes : si, pour une image i donnée, $p_i = 0$, alors $p_i = \frac{1}{2N}$ (où N est le nombre de sujets du groupe) ; de même, si $p_i = 1$, alors $p_i = 1 - \frac{1}{2N}$ (une telle transformation est commune : Macmillan et Kaplan, 1985).

Par ailleurs, nous n'allons pas considérer la simple proportion (éventuellement corrigée), mais l'exprimer par rapport à la proportion de la réponse alternative (majoritaire). Il s'agit d'une transformation qui exprime la « bizarrerie » (le caractère rare) de la réponse produite (odds). Cet odd ratio est souvent utilisé en épidémiologie, dans l'analyse de l'occurrence de pathologies rares : $\frac{p_i}{1-p_i}$. Ainsi, si la réponse « B » n'est produite que par 20 % des sujets du groupe témoin (p_i vaut 0,2), la proportion de réponses « A » vaut $1 - p_i$ (soit 0,8), et le odd ratio 0,25 : le sujet a quatre fois moins de chances de donner la bonne réponse en choisissant B qu'en choisissant A. Dans le

langage des parieurs, on dira que la réponse « B » est mise à quatre contre un : un parieur sur cinq recevrait donc quatre fois sa mise initiale. Afin de préserver les propriétés des proportions (ou probabilités) et compte tenu de la définition du odd ratio, le numérateur du rapport est donc la probabilité de l'événement rare, la réponse minoritaire (et le dénominateur celle de la réponse majoritaire : $\frac{p(\min)}{1-p(\min)} = \frac{p(\min)}{p(\max)}$).

3.3. Calcul du score global pour l'ensemble de la série

Le score global pour la série devrait être, en principe, le produit des 19 odds ratios : en effet, la probabilité d'obtenir exactement la série des 19 réponses observées est égale au produit des probabilités associées à chacune des 19 réponses indépendantes. Cela entraîne deux difficultés. D'une part, nous obtiendrons des valeurs très proches de 0, en particulier parce que nous utilisons la probabilité de la réponse minoritaire : dans l'exemple du Tableau 1, le produit des probabilités vaut environ $5 \cdot 10^{-23}$. D'autre part, l'interprétation d'un produit de probabilités n'est pas chose aisée.

Nous avons vu que la distribution des probabilités de réponses « B » pour une série donnée obéit à une fonction sigmoïde (Fig. 2), plus particulièrement à une fonction logistique. L'expression la plus simple d'une fonction logistique est la suivante : $p = \frac{e^a}{1+e^a} = \frac{1}{1+e^{-a}}$. Par conséquent, toute différence entre deux distributions i et j (deux séries pour un même sujet, ou une même série soumise à deux sujets) sera uniquement fonction du paramètre a , puisque les autres composantes de la formule (1 et e) sont des constantes : l'important est donc la différence entre a_i et a_j . Or, on peut montrer que $a_i = \log_e \left(\frac{p_i}{1-p_i} \right)$ et que $a_j = \log_e \left(\frac{p_j}{1-p_j} \right)$: a_i et a_j sont donc les logarithmes respectifs des odds ratios. Si, dans les odds ratios, on remplace p_i par $\frac{1}{1+e^{-a_i}}$ et p_j par $\frac{1}{1+e^{-a_j}}$, nous obtenons $p_i = e^{a_i}$ et $p_j = e^{a_j}$ dont le produit $p_i p_j = e^{a_i+a_j}$. Puisque le logarithme naturel de ce produit est $a_i + a_j$, il en découle que

Tableau 2

Calcul du score global de sept sujets virtuels illustrant différents patterns de réponses (les valeurs en gras correspondent aux erreurs)

Image	Réponse majoritaire	Odd ratio de la réponse minoritaire	S1 : aucune erreur		S2 : une erreur légère		S3 : une erreur grave		S4 : deux erreurs légères		S5 : une erreur grave et une légère		S6 : deux erreurs graves		S7 : pas de réponse correcte	
			Réponse	Log	Réponse	Log	Réponse	Log	Réponse	Log	Réponse	Log	Réponse	Log	Réponse	Log
1	A	0,017	A	-4,078	A	-4,078	B	4,078	A	-4,078	B	4,078	B	4,078	B	4,078
2	A	0,017	A	-4,078	A	-4,078	A	-4,078	A	-4,078	A	-4,078	A	-4,078	B	4,078
3	A	0,034	A	-3,367	A	-3,367	A	-3,367	A	-3,367	A	-3,367	A	-3,367	B	3,367
4	A	0,071	A	-2,639	A	-2,639	A	-2,639	A	-2,639	A	-2,639	A	-2,639	B	2,639
5	A	0,034	A	-3,367	A	-3,367	A	-3,367	A	-3,367	A	-3,367	A	-3,367	B	3,367
6	A	0,250	A	-1,386	A	-1,386	A	-1,386	A	-1,386	A	-1,386	A	-1,386	B	1,386
7	A	0,071	A	-2,639	A	-2,639	A	-2,639	A	-2,639	A	-2,639	A	-2,639	B	2,639
8	B	0,579	B	-0,547	A	0,547	B	-0,547	A	0,547	A	0,547	B	-0,547	A	0,547
9	B	0,304	B	-1,190	B	-1,190	B	-1,190	A	1,190	B	-1,190	B	-1,190	A	1,190
10	B	0,200	B	-1,609	B	-1,609	B	-1,609	B	-1,609	B	-1,609	B	-1,609	A	1,609
11	B	0,154	B	-1,872	B	-1,872	B	-1,872	B	-1,872	B	-1,872	B	-1,872	A	1,872
12	B	0,304	B	-1,190	B	-1,190	B	-1,190	B	-1,190	B	-1,190	B	-1,190	A	1,190
13	B	0,154	B	-1,872	B	-1,872	B	-1,872	B	-1,872	B	-1,872	B	-1,872	A	1,872
14	B	0,034	B	-3,367	B	-3,367	B	-3,367	B	-3,367	B	-3,367	B	-3,367	A	3,367
15	B	0,017	B	-4,078	B	-4,078	B	-4,078	B	-4,078	B	-4,078	B	-4,078	A	4,078
16	B	0,017	B	-4,078	B	-4,078	B	-4,078	B	-4,078	B	-4,078	B	-4,078	A	4,078
17	B	0,017	B	-4,078	B	-4,078	B	-4,078	B	-4,078	B	-4,078	B	-4,078	A	4,078
18	B	0,017	B	-4,078	B	-4,078	B	-4,078	B	-4,078	B	-4,078	B	-4,078	A	4,078
19	B	0,017	B	-4,078	B	-4,078	B	-4,078	B	-4,078	B	-4,078	A	4,078	A	4,078
		- somme des log		53,59		52,49		45,43		50,12		44,34		37,28		-53,59
		Scores		100,00		98,98		92,39		96,76		91,37		84,78		0,00

S : sujet.

deux distributions différeront par la somme des logarithmes des odds ratios. Nous pouvons donc calculer un score global par la somme de ces logarithmes, plutôt que par le produit des probabilités initiales. Dans la mesure où le logarithme d'une proportion est négatif, on prendra en considération le négatif de

cette somme. Ainsi, le score global observé (S_{obs}) s'obtient par $s_{obs} = -\sum_{i=1}^n \log_e \left(\frac{p_i}{1-p_i} \right)$. Le Tableau 2 présente les résultats de sept sujets virtuels, depuis le sujet parfait qui ne commet aucune erreur jusqu'à celui qui ne produit que des erreurs, en passant par quelques cas intermédiaires illustratifs.

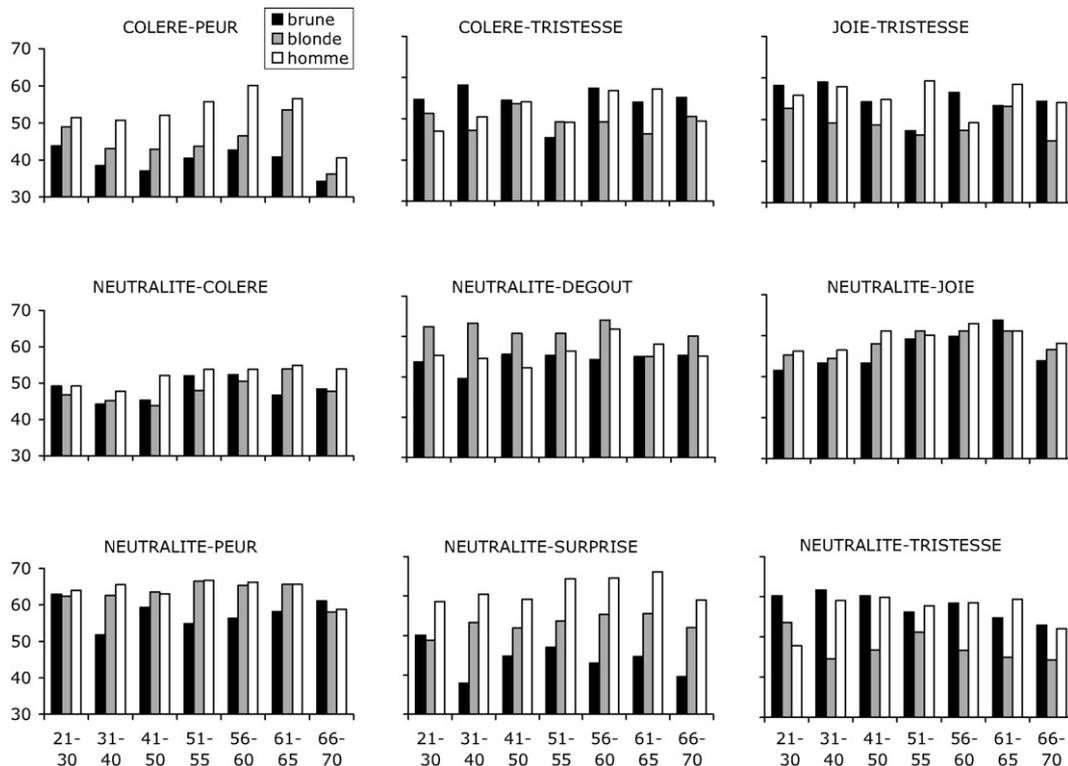


Fig. 3. Score maximal possible pour chacune des séries et chaque groupe d'âge.

3.4. Calcul du score global relatif pour la série

Le score maximal possible est celui qu'obtiendrait un sujet qui ne produit aucune erreur. Il est toutefois différent d'une série à l'autre, compte tenu du mode de calcul (certaines séries sont plus faciles que d'autres). Ainsi, pour l'exemple qui nous occupe, ce maximum est 53,59 (Tableau 2 : sujet S₁), mais pour le groupe 21–30 ans, le score maximum de la même série (colère–tristesse chez la blonde) devient 51,35 (Annexe A). Par conséquent, afin de pouvoir comparer des séries, des groupes ou des sujets individuels, le score obtenu par un sujet donné sera exprimé en pourcentage du maximum possible pour la série considérée et la tranche d'âge à laquelle il appartient. Dans notre exemple, le sujet 2 du Tableau 2 obtiendrait un score relatif de $100 \cdot \frac{52,49}{53,59} = 97,95\%$.

Une seconde difficulté doit être résolue. Comme on le voit dans le Tableau 2, le sujet 7 qui ne commet que des erreurs obtient un score négatif. Le score minimum n'est pas 0 mais le négatif du score maximum. Après transformation en pourcentage, nous obtenons donc une échelle de scores relatifs bornée par –100 et +100. Afin de faciliter la compréhension des résultats et les analyses comparatives ultérieures, il convient donc de ramener cette échelle à l'intervalle 0–100. Par conséquent, le score relatif en pourcentage est à augmenter de 100 (ce qui conduit à un intervalle 0–200) et le résultat est à diviser par 2. Au total, le score relatif pour une série et un sujet donné devient $\frac{(s_{obs} \cdot 100) + 100}{2} = \frac{50 \cdot (s_{obs} + s_{max})}{s_{max}}$, en pourcentage du maximum. Ces scores ont été calculés pour les sept sujets virtuels du Tableau 2. On peut constater la relative sensibilité de ces résultats au nombre et à la gravité des erreurs produites.

3.5. Application

Nous avons calculé le score maximal possible (brut) pour chacune des séries et chacun des groupes d'âge. La Fig. 3 en présente le résultat. Sur un plan purement descriptif, on peut observer des variations, puisque ce score maximal varie entre 34,54 (groupe 66–70 ans, colère–peur chez la brune) et 66,77 (groupe 51–55 ans, neutralité–peur chez l'homme), la moyenne étant 53,74. Intuitivement, on peut interpréter ce score maximal comme une expression du degré d'unanimité des sujets du groupe témoin sur l'ensemble de la série, et donc la « facilité » de celle-ci.

3.6. Illustrations

Un échantillon indépendant de 13 sujets sains (dont six femmes), âgés de 29 à 41 ans et sélectionnés selon les mêmes critères que pour les sujets des sept groupes initiaux, a été soumis aux 27 séries, avec cinq présentations de chaque série. Le score a été calculé pour chaque sujet, passation et série, en prenant appui sur le score maximal du groupe d'âge correspondant (21–30 ans pour quatre sujets ; 31–40 pour huit sujets ; 41–50 pour un sujet). La Fig. 4 présente les moyennes des 13 sujets, par série. Une Anova 3 (personnage) × 9 (série) × 5 (passation) [les trois facteurs sont répétés] a montré un effet du personnage ($F[2,24] = 5,65$, $p < 0,01$), de la série ($F[8,96] = 5,08$, $p < 0,0001$) et de l'interaction

personnage × série ($F[16,192] = 3,98$, $p < 0,0001$). En revanche, il n'y a pas d'effet significatif de la passation, ni d'interaction significative entre la passation et les autres facteurs. Cela montre la bonne stabilité des performances des sujets au cours des passations. Par ailleurs, nous voulons simplement illustrer ici les potentialités et la sensibilité du score à diverses variables indépendantes (ici le personnage et la série, mais pas la passation), sans en faire une interprétation détaillée.

Les 27 séries émotionnelles ont été soumises à 35 patients psychotiques, âgés de 21 à 70 ans. Vingt patients présentaient la forme paranoïde de la psychose (huit en phase aiguë et 12 en phase chronique), et 15 patients la forme hétéroforme (quatre en phase aiguë et 11 en phase chronique). Les scores de chacun ont été calculés par référence au groupe d'âge correspondant. Une Anova 2 (chronicité) × 2 (forme de psychose) × 3 (personnage) × 9 (série), avec les deux derniers facteurs répétés, a montré des effets principaux significatifs de la chronicité ($F[1,31] = 3,84$, $p < 0,06$), de la série ($F[8,248] = 12,93$, $p < 0,0001$) et du personnage ($F[2,62] = 4,71$, $p < 0,0125$), ainsi qu'une interaction significative chronicité × forme × série ($F[8,248] = 2,42$, $p < 0,016$) qui est illustrée par la Fig. 5. À nouveau, nous n'entrons pas dans le détail de l'interprétation de cette interaction Fig. 5.

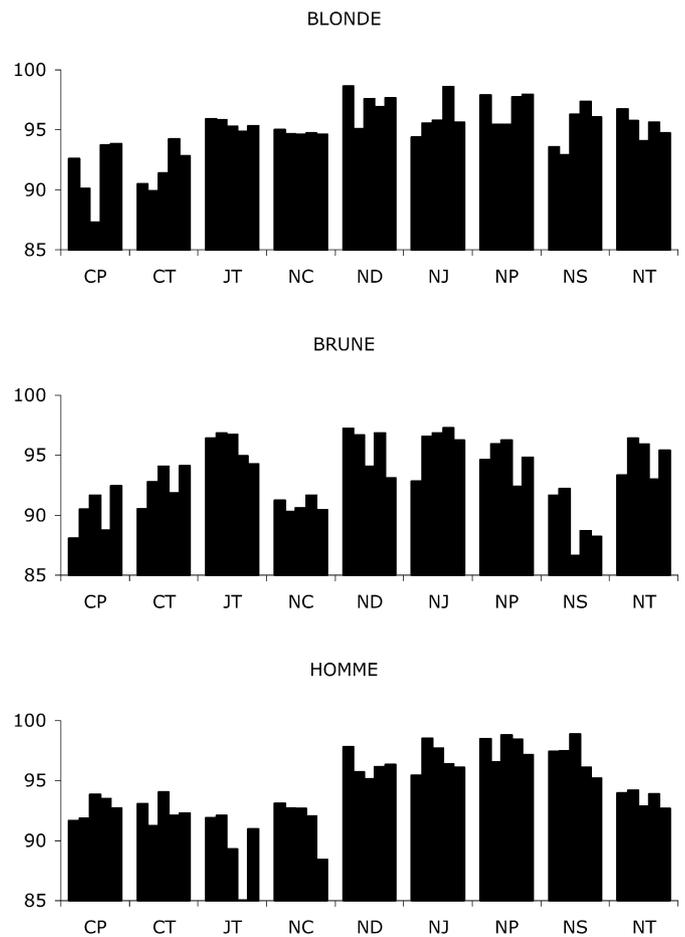


Fig. 4. Score moyen (en %) de 13 sujets sains au cours de cinq passations, pour chacune des neuf séries et chacun des trois personnages. N : neutralité ; C : colère ; P : peur ; T : tristesse ; J : joie ; D : dégoût ; S : surprise.

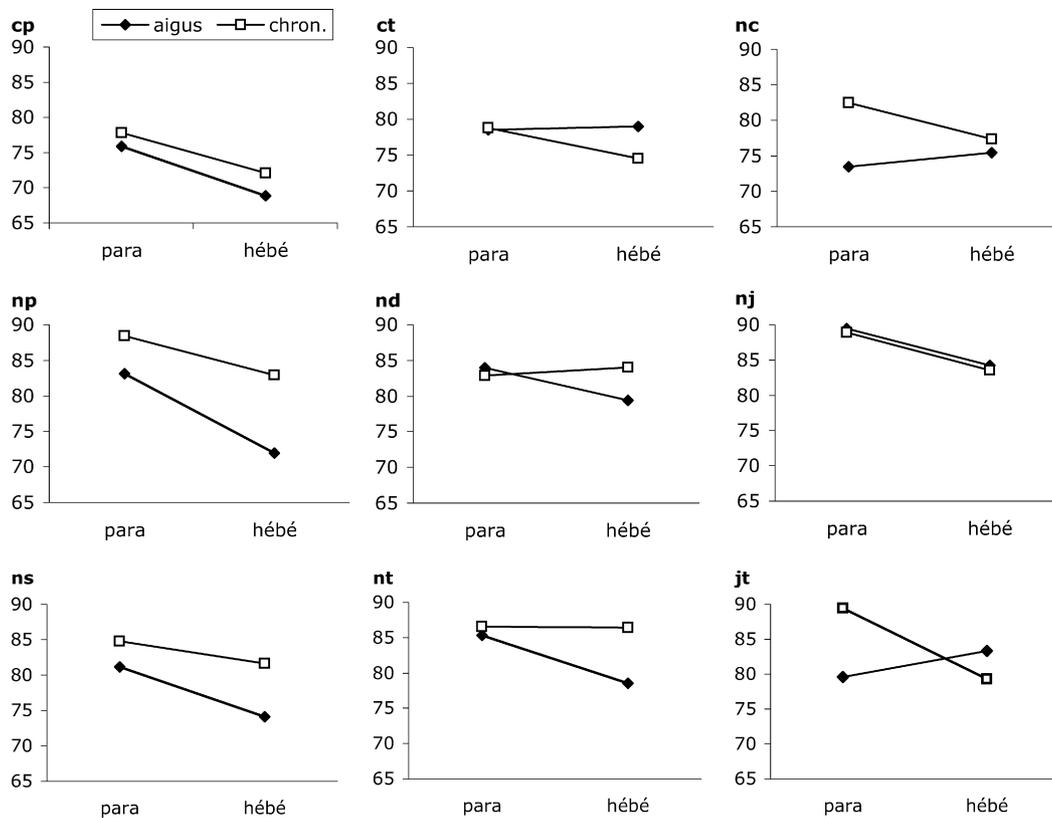


Fig. 5. Score moyen (en %) de 35 sujets psychotiques pour chacune des neuf séries, en fonction de la forme de psychose (hébéphrénique, paranoïde) et de sa chronicité (phase chronique, phase aiguë). n : neutralité ; c : colère ; p : peur ; t : tristesse ; j : joie ; d : dégoût ; s : surprise.

Les neuf séries émotionnelles exprimées par la blonde ont été présentées à 33 patients atteints d’une démente d’Alzheimer débutante. Nous n’avons pas pu calculer le score de 22 patients, car ils étaient âgés de plus de 70 ans (nous n’avons pas de groupe de référence pour cette tranche d’âge). Les 11 autres patients (six femmes) étaient âgés de 51 à 70 ans, avec une moyenne de 67,2 ; ils ont été comparés à huit sujets sains appariés. La Fig. 6 présente les moyennes des deux groupes pour chacune des neuf séries. Une Anova 2 (groupe) × 9 (série), dont le second facteur est répété, a montré un effet significatif de la série ($F[8,136] = 3,79, p < 0,0005$, sans effet significatif du groupe et sans interaction significative entre la série et le groupe.

Enfin, un petit échantillon de six patients cérébrolésés (quatre ruptures d’anévrisme de l’artère communicante antérieure, un infarctus bilatéral du *globus pallidum*, un AVC frontal), comparé à huit sujets sains appariés en âge et sexe, a été soumis aux 27 séries. La Fig. 7 présente les scores moyens des deux groupes, qui ont été comparés par le test non-paramétrique de Mann-Whitney, vu la taille des groupes. Une différence significative en faveur des sujets sains est observée pour la série joie–tristesse produite par la blonde ($U = 11, p < 0,05$), la série neutralité–colère produite par la blonde ($U = 11, p < 0,05$) et par l’homme ($U = 7, p < 0,035$) et la série neutralité–tristesse produite par la brune ($U = 9, p < 0,064$: tendance), mais un avantage des patients sur les sujets sains pour la série joie–tristesse produite par la brune ($U = 9, p < 0,064$: tendance).

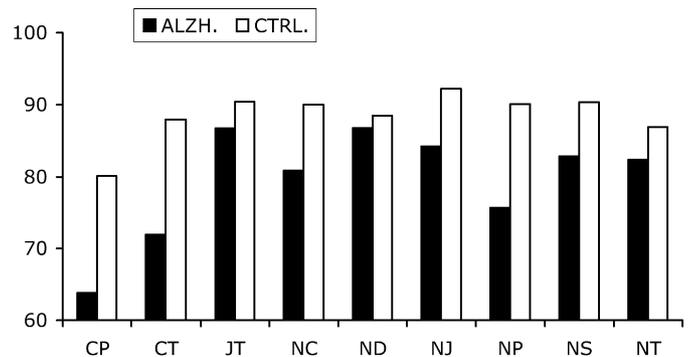


Fig. 6. Score moyen (en %) de 11 patients atteints de la maladie d’Alzheimer débutante comparés à huit sujets sains appariés pour chacune des neuf séries d’expressions produites par la blonde. N : neutralité ; C : colère ; P : peur ; T : tristesse ; J : joie ; D : dégoût ; S : surprise.

L’étape empirique suivante, qui dépasse l’objectif du présent article, sera donc de soumettre de nouveaux groupes indépendants de sujets sains à ces séries émotionnelles et de calculer les scores relatifs individuels. Cela permettra d’abord de comparer statistiquement les effets de l’âge, de la série et du personnage chez des sujets sains. Cela nous conduira ensuite à des normes, à partir des moyennes et écarts-types de chaque série pour chaque groupe, et ces normes permettront d’apprécier les performances d’un sujet individuel (ou d’un groupe de sujets) susceptible de présenter une déficience suite, par exemple, à une atteinte neurologique, à une pathologie psychiatrique

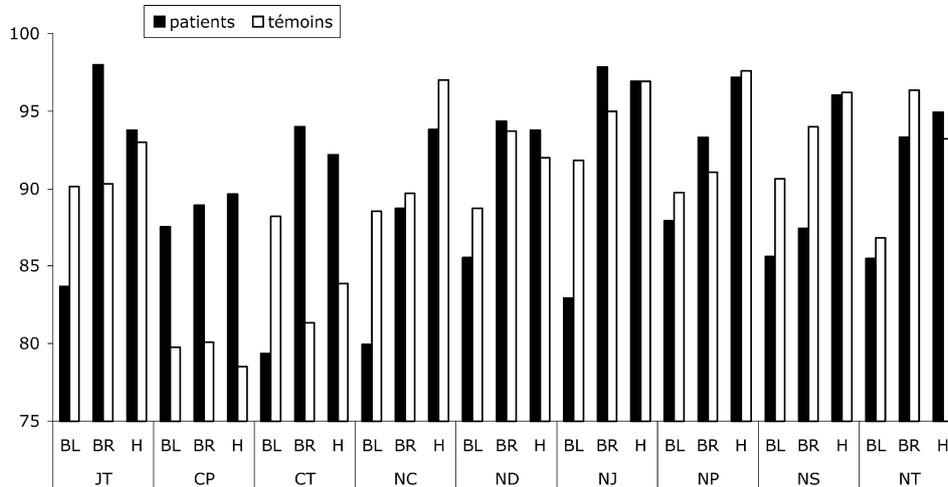


Fig. 7. Score moyen (en %) de six patients cérébrolésés et huit sujets sains appariés pour chacune des neuf séries et chacun des trois personnages. N : neutralité ; C : colère ; P : peur ; T : tristesse ; J : joie ; D : dégoût ; S : surprise ; BL : blonde ; BR : brune ; H : homme.

ou aux effets du vieillissement normal. En effet, tant que ces normes ne sont pas disponibles, nous ne pouvons décider si le score d'un sujet donné est « normal » ou « pathologique ». En revanche, comme on vient de le voir par quatre exemples, il est dès à présent possible de comparer les séries, les personnages, ou encore des groupes entre eux, même sans disposer de normes.

4. Discussion et conclusion

La technique du morphing est aujourd'hui abondamment utilisée pour établir la nature catégorielle de la perception des identités faciales ou des expressions faciales émotionnelles. Elle pourrait évidemment être étendue à d'autres types de stimuli visuels complexes, voire, en l'adaptant, à d'autres modalités sensorielles (la voix, l'écriture, l'odeur, etc.).

Nous montrons ici que les épreuves, tests ou expériences mis au point pour évaluer cette perception catégorielle peuvent également être exploités pour mesurer la perception, indépendamment de sa nature catégorielle ou non : c'est le score relatif. Inévitablement, comme tout score composite, cette démarche entraîne une perte d'information.

En effet, par définition, on semble perdre toute information sur le profil général de la courbe de réponses, en particulier sa nature sigmoïdale (catégorielle) ou non. C'est évidemment inhérent à la notion même de « score global ». Soulignons toutefois que l'allure de la courbe est implicitement prise en compte par l'introduction des proportions (p_i) de sujets du groupe de référence dans le calcul du score par image, dont le score relatif final n'est que la synthèse ajustée et exprimée par rapport à un maximum. En outre, l'allure sigmoïdale de la courbe est explicitement prise en compte, puisque ce sont les propriétés mêmes de la courbe logistique qui permettent de passer d'un produit de probabilités, difficilement interprétable,

à une somme de logarithmes. Un article en préparation porte sur cette allure des courbes et montre, en particulier, que c'est la fonction logistique qui conduit au meilleur ajustement des données empiriques, pour chacune des courbes.

Bien que présentées dans un ordre aléatoire, les 19 images d'une même série ne sont pas simplement 19 stimuli équivalents d'une même catégorie. En effet, leur composition est telle qu'ils varient en termes de proportions de pixels de deux images sources. Par conséquent, les méthodes classiques reposant sur un comptage du nombre d'erreurs puis la dérivation éventuelle d'indices plus élaborés issus, par exemple, de la *théorie de la détection du signal*, ne sont guère appropriées dans ce contexte. Notre utilisation du terme « erreur » est d'ailleurs quelque peu ambiguë, comme nous l'avons expliqué plus haut (nous n'avons pas trouvé de terme plus approprié). Quoi qu'il en soit, une comparaison directe de la méthode proposée ici à des méthodes plus habituelles n'aurait guère de sens, à moins de négliger les caractéristiques des stimuli utilisés. Cependant, lorsque des scores normatifs seront disponibles, il sera sans doute possible de comparer cette méthode à d'autres en termes de détection de performances pathologiques (sensibilité et spécificité).

Il nous semble donc que l'intérêt principal de cette démarche est de pouvoir définir, en le nuancant, le caractère correct ou erroné d'une réponse, en prenant appui sur la réponse produite par la majorité des sujets du groupe témoin (p_i). En effet, une difficulté habituelle des tests, épreuves ou expériences où on demande au sujet de choisir librement est qu'on ne peut ensuite décider si sa réponse est correcte ou non : il s'agit plutôt d'une « préférence ». L'adaptation proposée ici nous semble par ailleurs généralisable à toute épreuve destinée à mesurer la perception catégorielle, y compris dans les modalités non visuelles.

Annexe A

Nombre de sujets, par groupe, personnage et paire d'expressions, qui ont choisi l'expression B, pour chacune des 19 images.

	% de B	21-30; n=30		31-40; n=30		41-50; n=30		51-55; n=30		56-60; n=30		61-65; n=30		66-70; n=24	
		brune	blonde	homme	brune	blonde	homme	brune	blonde	homme	brune	blonde	homme	brune	blonde
0	0	0	0	1	1	1	0	0	1	0	0	1	0	0	0
10	0	0	0	1	2	0	1	0	1	2	0	2	1	0	1
20	1	0	1	1	0	0	1	3	1	1	0	1	1	0	0
30	2	0	4	3	2	5	3	1	2	0	3	2	1	0	2
35	5	3	7	8	4	6	9	3	4	4	3	4	6	3	7
38	16	10	13	12	9	14	10	8	16	8	8	12	11	14	14
41	12	9	19	12	4	10	13	9	13	8	5	9	10	7	6
44	22	21	22	18	18	27	18	23	24	18	15	27	19	23	24
47	23	21	26	23	20	23	26	19	27	20	24	25	24	23	28
50	19	20	25	23	21	27	23	25	27	21	21	29	23	27	27
53	21	22	27	25	21	28	22	21	28	22	25	30	23	26	30
56	21	24	27	22	25	27	24	20	27	20	22	26	19	22	30
59	28	27	30	27	29	29	23	23	30	23	24	29	29	28	30
62	29	30	29	29	29	29	25	27	28	27	29	28	27	28	29
65	27	30	30	26	29	30	28	29	30	28	30	30	29	30	28
70	29	27	30	27	28	29	27	29	30	28	27	29	27	28	30
80	30	30	30	29	28	30	29	29	29	29	30	29	29	29	30
90	29	29	30	28	29	29	29	30	30	28	30	30	29	29	30
100	30	30	29	29	28	29	28	28	29	30	27	30	29	29	28
0	0	0	0	0	0	1	0	0	1	1	1	1	0	1	0
10	0	0	0	0	0	0	0	0	0	1	0	2	1	0	0
20	1	1	3	5	2	1	6	1	1	3	1	2	1	0	0
30	5	3	4	8	2	2	14	2	3	7	1	3	2	0	3
35	14	5	3	16	5	6	21	1	3	15	1	6	11	4	2
38	25	6	10	28	12	17	27	6	17	27	7	17	23	5	15
41	18	5	5	20	7	13	23	2	5	21	4	7	14	3	5
44	27	23	20	28	22	24	27	19	28	26	12	21	27	14	26
47	26	20	20	30	17	25	29	23	21	26	15	21	28	19	24
50	29	24	25	30	23	26	28	25	26	28	20	26	29	19	29
53	28	27	27	28	29	26	29	26	29	18	30	29	22	27	29
56	27	25	27	29	23	29	29	23	29	28	26	24	30	22	29
59	30	27	28	30	28	28	28	26	29	29	27	30	29	25	30
62	28	30	29	30	26	30	30	29	28	29	29	30	29	27	30
65	30	30	29	29	30	30	29	30	30	28	30	29	29	29	30
70	30	30	28	30	28	29	29	30	28	27	29	29	30	30	30
80	30	29	30	30	30	30	30	29	29	30	30	30	30	30	30
90	30	30	28	30	30	29	30	30	30	30	30	30	29	30	30
100	30	30	30	29	30	29	30	30	27	30	30	30	30	29	29
0	0	0	0	0	0	1	0	1	4	0	2	2	1	1	3
10	1	0	1	1	2	3	0	1	2	2	0	2	2	0	0
20	1	0	0	2	1	3	3	0	0	2	1	1	1	1	0
30	7	0	1	3	0	1	5	1	2	3	0	1	3	0	4
35	9	3	3	5	0	0	13	1	4	7	0	4	3	0	3
38	27	5	9	26	3	21	24	2	15	22	2	18	24	2	15
41	15	2	3	8	1	4	16	1	3	11	3	5	11	1	7
44	30	14	21	29	14	25	28	10	26	25	8	28	28	9	23
47	25	15	21	25	16	26	24	10	23	24	9	26	28	10	25
50	30	21	29	30	15	30	28	14	29	25	15	27	27	13	25
53	30	25	26	29	24	28	29	20	29	27	17	29	28	19	26
56	27	22	25	28	15	26	28	13	26	28	15	28	27	14	28
59	30	30	30	30	24	29	30	26	30	28	26	30	30	23	29
62	30	29	30	29	28	30	30	26	30	28	25	30	29	23	28
65	29	29	30	30	30	30	30	28	30	29	29	30	29	29	29
70	29	30	30	30	30	30	30	30	30	29	28	30	30	30	30
80	30	30	29	30	30	30	29	30	29	30	30	30	30	29	30
90	30	29	29	29	30	30	30	30	30	30	30	30	29	30	30
100	30	30	30	30	29	30	30	30	30	29	30	30	30	29	30

Références

- Beale, J.M., Keil, F.C., 1995. Categorical effects in the perception of faces. *Cognition* 57, 217–239.
- Bornstein, M.H., Kessen, W., Weiskopf, S., 1976. Color vision and hue categorization in young human infants. *Journal of Experimental Psychology: Human Perception and Performance* 2, 115–129.
- Bruyer, R., Granato, P., 1999. Categorical effects in the perception of facial expressions: MARIE, a simple and discriminating clinical tool. *European Review of Applied Psychology* 49, 3–10.
- Bülthoff, I., Newell, F.N., 2004. Categorical perception of sex occurs in familiar but not unfamiliar faces. *Visual Cognition* 11, 823–855.
- Calder, A.J., Young, A.W., Perrett, D.I., Ectoff, N.L., Rowland, D., 1996. Categorical perception of morphed facial expressions. *Visual Cognition* 3, 81–117.
- Campanella, S., Hanoteau, C., Dépy, D., Rossion, B., Bruyer, R., Crommelinck, M., et al., 2000. Right N₁₇₀ modulation in a face discrimination task: an account for categorical perception of familiar faces. *Psychophysiology* 37, 796–806.
- Campanella, S., Chrysochoos, A., Bruyer, R., 2001. Categorical perception of facial gender information: behavioural evidence and the face-space metaphor. *Visual Cognition* 8, 237–262.
- Campanella, S., Quinet, P., Bruyer, R., Crommelinck, M., Guérit, J.M., 2002. Categorical perception of happiness and fear facial expressions: an ERP study. *Journal of Cognitive Neuroscience* 14, 210–227.
- Campanella, S., Hanoteau, C., Seron, X., Joassin, F., Bruyer, R., 2003. Categorical perception of unfamiliar facial identities and the face-space metaphor. *Visual Cognition* 10, 129–156.
- De Gelder, B., Teunisse, J.P., Benson, P.J., 1997. Categorical perception of facial expressions: categories and their internal structure. *Cognition and Emotion* 11, 1–23.
- Ekman, P., 1992. An argument for basic emotions. *Cognition and Emotion* 6, 169–200.
- Ekman, P., Friesen, W.V., 1976. *Pictures of facial affect*. Consulting Psychologists Press, Palo Alto.
- Ectoff, N.L., Magee, J.J., 1992. Categorical perception of facial expressions. *Cognition* 44, 222–240.
- Granato, P., Bruyer, R., 2002. Measurement of the perception of facially expressed emotions by a computerized device. *European Psychiatry* 17, 339–348.
- Granato, P., Bruyer, R., Révillion, J.J., 1996. Étude objective de la perception du sourire et de la tristesse par la méthode d'analyse et de recherche de l'intégration des émotions « MARIE ». *Annales Médico-Psychologiques* 154, 1–9.
- Harnad, S., 1987. In: *Categorical Perception: the groundwork of cognition*. Cambridge University Press, Cambridge.
- Hsu, S.M., Young, A.W., 2004. Adaptation effects in facial expression recognition. *Visual Cognition* 11, 871–899.
- Kiffel, C., Campanella, S., Bruyer, R., 2005. Categorical perception of faces and facial expressions: the age factor. *Experimental Aging Research* 31, 119–147.
- Levin, D.T., Beale, J.M., 2000. Categorical perception occurs in newly learned faces, other-race faces, and inverted faces. *Perception and Psychophysics* 62, 386–401.
- Lieberman, A.M., Harris, S., Hooman, H.S., Griffith, B.C., 1957. The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology* 54, 358–368.
- Macmillan, N.A., Kaplan, H.L., 1985. Detection theory analysis of group data: estimating sensitivity from average hit and false-alarm rates. *Psychological Bulletin* 98, 185–199.
- Plutchik, R., 1980. *Emotion: a psychoevolutionary synthesis*. Harper and Row, New York.
- Shah, R., Lewis, M.B., 2003. Locating the neutral expression in the facial-emotion space. *Visual Cognition* 10, 549–566.
- Stevenage, S.V., 1998. Which twin are you? A demonstration of induced categorical perception of identical twin faces. *British Journal of Psychology* 89, 39–57.
- Young, A.W., Rowland, D., Calder, A.J., Ectoff, N.L., Seth, A., Perrett, D.I., 1997. Facial expression megamix: tests of dimensional and category accounts of emotion recognition. *Cognition* 63, 271–313.