



ELSEVIER  
MASSON

Disponible en ligne sur [www.sciencedirect.com](http://www.sciencedirect.com)

 ScienceDirect

Psychologie française 55 (2010) 49–58

Psychologie  
française

[www.em-consulte.com](http://www.em-consulte.com)

Article original

# Perception d'attributs des visages parlants par le nouveau-né humain

*Feature perception of speaking faces in newborn infants*

C. Berger<sup>a,\*</sup>, S. Donnadieu<sup>a</sup>, D. Meary<sup>b</sup>, S. Kandel<sup>b</sup>, K. Mazens<sup>b</sup>

<sup>a</sup> LPNC (UMR CNRS 5105), université de Savoie, BP 1104, 73000 Chambéry, France

<sup>b</sup> LPNC (UMR CNRS 5105), université Pierre-Mendès-France, BP 47, 38040 Grenoble, France

Reçu le 18 février 2009 ; accepté le 28 décembre 2009

---

## Résumé

Une méthode de préférence visuelle est utilisée pour étudier comment les nouveau-nés peuvent appréhender certaines caractéristiques internes du visage en contexte multimodal. Dans une première expérience, deux visages mobiles associés à un son de parole sont opposés : l'un net, l'autre filtré passe-bas. Une préférence pour le visage net est observée. Une seconde expérience oppose deux visages nets, présentant des mouvements labiaux respectivement congruents et non congruents avec le son de parole associé. Les nouveau-nés ne manifestent alors aucune préférence. Ces résultats montrent que les nouveau-nés cherchent à appréhender, sur les visages mobiles parlants, plus d'informations que celles transmises uniquement par les très basses fréquences spatiales. Ils ne semblent cependant pas en mesure de détecter les correspondances entre les sons et les mouvements des lèvres.

© 2010 Société française de psychologie. Publié par Elsevier Masson SAS. Tous droits réservés.

*Mots clés* : Perception multimodale ; Nouveau-né ; Préférence visuelle ; Appariement ; Visages

## Abstract

A visual preference method is used for studying how human newborns can process internal characteristics on speaking faces in a multimodal context. In a first experiment, two mobile faces associated with a speech sound are opposed: one is clear and the other is blurred. A preference for the clear face is observed. A second experiment contrasts two clear faces, showing lip movements respectively congruent and non-congruent with the associated speech sound. No preference is observed in this situation. These results show that newborns

---

\* Auteur correspondant.

Adresses e-mail : [Carole.Berger@univ-savoie.fr](mailto:Carole.Berger@univ-savoie.fr) (C. Berger), [Sophie.Donnadieu@univ-savoie.fr](mailto:Sophie.Donnadieu@univ-savoie.fr) (S. Donnadieu), [David.Maery@upmf-grenoble.fr](mailto:David.Maery@upmf-grenoble.fr) (D. Meary), [Sonia.Kandel@upmf-grenoble.fr](mailto:Sonia.Kandel@upmf-grenoble.fr) (S. Kandel), [Karine.Mazens@upmf-grenoble.fr](mailto:Karine.Mazens@upmf-grenoble.fr) (K. Mazens).

process mobile faces by orienting their attention on more information than that corresponding to very low spatial frequencies. It seems however that they cannot detect the correspondences between the lip movements and the associated sounds.

© 2010 Société française de psychologie. Published by Elsevier Masson SAS. All rights reserved.

*Keywords:* Multimodal perception; Newborns; Visual preference; Matching; Face

---

La recherche présentée ici traite de la sensibilité du nouveau-né aux caractéristiques internes du visage, situées en particulier au niveau de la région labiale. L'étude vise à déterminer dans quelle mesure les bébés pourraient être sensibles à la congruence entre les informations visuelles (mouvements labiaux) et auditives (sons de parole) émanant de visages parlants.

L'approche multimodale a jusqu'ici été peu envisagée dans la littérature traitant de la perception des visages par le nouveau-né. Les visages tels qu'ils apparaissent dans les situations d'interaction quotidienne sont pourtant généralement associés à des informations accessibles par différents canaux sensoriels. Interviennent, en particulier, des informations visuelles incluant des indices de mouvements (indices très présents au niveau de la région labiale) et des informations auditives correspondant à des sons de parole. Ces informations pourraient jouer un rôle important dans l'organisation des perceptions du nouveau-né, étant donné sa sensibilité au mouvement et la fonctionnalité de son système auditif (pour une revue, voir [Lécuyer et al., 1994](#)).

L'étude aborde donc la question, largement discutée dans la littérature, des liens entre modalités sensorielles à la naissance. La conception traditionnelle d'un fonctionnement indépendant des modalités ([Piaget et Inhelder, 1966](#)) laisse place, dans les théories récentes, à l'idée d'un fonctionnement plus unitaire. [Bahrick et al. \(2004\)](#) proposent, par exemple, que les bébés percevaient leur environnement en commençant par détecter des propriétés de type amodal en contexte multimodal.

Le nouveau-né pourrait-il donc bénéficier de la redondance entre des informations provenant de différentes modalités sensorielles lorsqu'il perçoit des visages parlants ? Des études empiriques basées sur une analyse des conduites motrices du bébé suggèrent d'ores et déjà l'intervention d'appariements précoces impliquant l'audition et les données visuelles d'articulation. [Chen et al. \(2004\)](#) ont, par exemple, établi que les nouveau-nés imitaient les mouvements de lèvres (modèle placé à 30 cm, émettant un son) de façon cohérente, en produisant, par exemple, plus de mouvements d'ouverture lorsqu'ils entendaient le son /A/ que le son /M/ (le son /A/ étant naturellement associé à une ouverture de plus forte amplitude que le son /M/). [Aldridge et al. \(1999\)](#) ont, quant à eux, utilisé une procédure de choix opérant. Ils ont montré que les nouveau-nés apprenaient à moduler leur activité de succion non nutritive afin de voir apparaître des stimuli congruents (par exemple, son /A/ associé aux mâchoires ouvertes ou son /U/ associé aux lèvres arrondies) plutôt que non congruents. Il semblerait donc que le bébé soit sensible, à la naissance, à des informations bi-modales impliquant l'audition et l'articulation (voir aussi [Kuhl et Meltzoff, 1996](#)). Il est à noter que les appariements tels qu'ils sont décrits dans les études venant d'être citées impliquent toujours une réponse motrice (par exemple, conduite d'imitation) et sollicitent donc le système d'action. Des auteurs ont suggéré que ce système pouvait fonctionner de façon indépendante du système perceptif ([Mandler, 2000](#)). Il importe donc d'analyser les conduites aussi au niveau perceptif : le bébé pourrait-il tirer profit d'un contexte de perception multimodal pour traiter de façon cohérente des visages humains ?

Les rares études utilisant des indicateurs perceptifs (préférences visuelles) sur des visages humains en contexte multimodal ont rapporté des résultats intéressants chez les très jeunes bébés.

Patterson et Werker (2003) ont, par exemple, montré qu'à deux mois, les bébés peuvent relier des informations phonétiques aux lèvres et à la voix. Ils manifestent à cet âge des préférences pour les mouvements de lèvres synchrones avec le son, plutôt que non synchrones (Dodd, 1979). Des capacités de discrimination d'affects (joie, tristesse ou colère) ont aussi été rapportées chez les bébés de quatre mois en contexte multimodal (Flom et Bahrick, 2007) : le bébé semble détecter à cet âge des relations arbitraires entre le visage (modalité visuelle) et la voix (modalité auditive). Il est intéressant de noter que les discriminations d'affects n'interviennent que plus tardivement, dans cette étude, si ne sont disponibles que des informations unimodales (discriminations intervenant à cinq mois en présence de la voix seule et à sept mois en présence du visage seul).

Considérons maintenant les compétences perceptives observées spécifiquement chez le nouveau-né. Un point important qui sera développé ici concerne les facilitations dans le traitement des visages liées à la présence d'un contexte multimodal. Nous savons que le bébé s'oriente à la naissance de préférence sur le visage de sa mère (modèle naturel) plutôt que sur un autre visage (Buschnel et al., 1989). Il a été proposé que cette compétence pourrait être liée à la présence d'informations telles que la voix, le sourire, l'odeur, généralement associées au visage de la mère. La préférence pour le visage maternel n'intervient plus, par exemple, si le nouveau-né a été empêché d'entendre la voix de sa mère entre la naissance et le moment du test (Sai, 2005). La présence d'indices tels que la voix maternelle ne semble cependant nécessaire que si les caractéristiques externes du visage sont telles qu'elles n'apportent aucune information exploitable. C'est le cas dans l'étude de Sai (2005) puisque les préférences sont analysées sur deux visages appariés du point de vue de la chevelure (même couleur, même longueur, même coupe de cheveux). Pascalis et al. (1995) ont aussi montré que la préférence pour le visage maternel n'intervenait pas en l'absence d'informations telles que la voix, le sourire ou l'odeur de la mère, si la tête du modèle était recouverte d'un foulard. Lorsque au contraire, le bébé pouvait se baser sur des caractéristiques externes telles que la chevelure, une préférence pour le visage maternel a été observée, même lorsque aucune information complémentaire (voix, sourire, odeur) n'était présente.

Le caractère prépondérant des caractéristiques externes a été confirmé dans des études utilisant des photos de visages (Turati et al., 2006). L'accès à ces caractéristiques semble en fait être associé à une limite du système perceptif qui ne traiterait que les basses fréquences spatiales. Le bébé peut, par exemple, reconnaître des photographies de visages non familiers, suite à une habitude, si ces photographies contiennent des basses fréquences spatiales, inférieures à 0,5 cycles par degré (cpd), plutôt que des hautes fréquences spatiales, supérieures à 1 cpd (Heering et al., 2008). Nous étudions ici si le traitement des caractéristiques externes (informations de contour, correspondant aux basses fréquences spatiales) de visages photographiés pourrait s'accompagner, en contexte multimodal, d'un traitement complémentaire de certaines caractéristiques internes. Cette éventualité est examinée à la naissance dans une situation de perception offrant des indices auditifs et visuels articulatoires (le nouveau-né semblant pouvoir détecter certaines correspondances entre ces informations, tout au moins sur des modèles réels, voir ci-dessus). L'éventuel apport des sons de parole et des mouvements articulatoires est envisagé en regard de ce que nous savons :

- sur la détection de propriétés en contexte multimodal à la naissance (Bahrick et al., 2004) ;
- sur l'efficacité des systèmes perceptifs à la naissance.

Il a été montré qu'en dépit de l'immaturation physiologique du système visuel, le bébé peut appréhender visuellement des informations de mouvement (Haith, 1966). Le système auditif du nouveau-né a aussi été décrit comme étant relativement fonctionnel (Lécuyer et al., 1994). La

question est donc de savoir si, en présence de mouvements articulatoires et de sons de parole, le nouveau-né pourrait manifester des capacités de traitement visuel de certains attributs internes mobiles de visages filmés. Les attributs internes considérés ici sont situés au niveau de la région labiale (siège des mouvements articulatoires).

Deux expériences basées sur une procédure de préférence visuelle seront utilisées. Une première expérience, constituant le point de départ de l'étude, traite de la capacité des nouveau-nés à percevoir, sur un visage mobile, plus d'informations que celles transmises par les basses fréquences spatiales (dans ce cas seulement, nous pouvons penser que le bébé accède à des caractéristiques internes du visage). Deux visages naturels en mouvement associés à un son de parole seront donc contrastés. L'un de ces visages sera non filtré tandis que l'autre sera filtré passe-bas. Nous avons vu que le nouveau-né procédait à une extraction de propriétés visuelles dans une gamme limitée de fréquences spatiales (i.e., basses fréquences), cela lui permettant d'accéder à des propriétés de bas niveau. L'objectif est ici de déterminer si ce fonctionnement empêche toute possibilité d'exploration de caractéristiques internes du visage ou si au contraire, l'accès à certaines de ces caractéristiques demeure possible (lorsque sont présentes des informations visuelles dynamiques et des informations auditives). Le bébé devrait pouvoir percevoir, dans ce dernier cas, les fréquences spatiales disponibles sur le visage non filtré mais indisponibles sur le visage filtré.

Une seconde expérience analysera spécifiquement les caractéristiques de la perception au niveau de la région labiale. Deux visages nets seront cette fois-ci contrastés, présentant des mouvements de lèvres respectivement congruents et incongruents avec le son. Nous examinerons dans quelle mesure la détection visuelle d'une congruence entre le son et le mouvement des lèvres pourrait intervenir dès la naissance. La congruence sera ici particulièrement saillante puisqu'elle se définira par la forme des lèvres (correspondant au son produit) et par des critères temporels (exacte synchronie entre le mouvement des lèvres et le son produit). Se pose donc la question de savoir si les appariements mis en évidence à la naissance par le biais des conduites d'imitation pourraient se généraliser, en conditions favorables, à une procédure de préférence visuelle (les résultats positifs dont nous disposons actuellement avec cette procédure étant établis à partir de deux mois, [Dodd, 1979](#) ; [Patterson et Werker, 2003](#)).

## 1. Expérience 1 : visage non filtré versus visage filtré passe-bas

Les nouveau-nés peuvent-ils traiter certaines caractéristiques internes du visage lorsque sont disponibles à la fois des informations auditives (sons de parole) et de mouvement (mobilité au niveau des lèvres) ? Cette question est examinée par le biais d'une analyse des préférences entre un visage mobile non filtré et le même visage filtré passe-bas (filtrage de type gaussien, ayant pour conséquence de masquer les traits internes).

### 1.1. Participants

L'échantillon est constitué de 22 nouveau-nés (14 garçons et huit filles) de la clinique mutualiste de Grenoble. Les participants sont tous nés à terme ( $\pm 15$  jours) et sont en bonne santé (score d'Apgar égal à 10). Leur poids moyen à la naissance est de 3180 g (minimum : 2340 g, maximum : 4415 g) et leur moyenne d'âge est de 100 heures (minimum : 72 heures, maximum : 152 heures). Les nouveau-nés éveillés sont testés le matin, dans une salle calme de la clinique. L'un des parents accompagnant le bébé donne son consentement écrit après avoir été informé du but et du déroulement de l'expérience.



Fig. 1. Stimuli proposés au cours des deux expériences conduites. Le point positionné entre deux images (figurant les écrans) représente la position de la caméra filmant le regard des nouveau-nés.

Expérience 1 (en haut). Les films du visage parlant sont opposés à leur double après filtrage des hautes fréquences spatiales. Sur ces films, la bouche et les yeux ne sont plus visibles : seuls subsistent le contour et le mouvement du visage.

Expérience 2 (images du bas). Les films (ouverture des lèvres versus avancée/ protrusion) sont présentés en parallèle mais le son associé est toujours le son /A/.

## 1.2. Stimuli

Les stimuli (Fig. 1) correspondent à des vidéos (en format avi) présentant un visage féminin animé de mouvements de lèvres. À chaque image « originale » est associée une image « filtrée passe-bas » (fonction `fspecial` et `imfilter`: Matlab®). Le filtrage, de type gaussien, élimine les hautes fréquences spatiales, diminuant ainsi la netteté de l'image. Plus précisément, l'image filtrée ne préserve que les fréquences spatiales les plus basses (<0,15 cpd) de l'image originale. Les traits internes du visage, la bouche et les yeux ne sont alors plus visibles. Seul subsiste le contour du visage et les informations dynamiques.

Les vidéos sont accompagnées d'une bande son au format wav (44 100 Hz) correspondant à la prononciation du son /A/ (synchronisé avec le mouvement des lèvres). Les stimulations visuelles/ auditives sont :

- congruentes lorsqu'au son /A/ est associé un mouvement d'ouverture complet de la bouche ;
- non congruentes lorsqu'au son /A/ est associé un mouvement d'avancée des lèvres avec protrusion (ce type de mouvement labial étant en principe associé au son /OU/).

Les films avec mouvement d'ouverture ou d'avancée des lèvres ont une durée de trois secondes et sont montés en boucle pour obtenir des films de 12 secondes. Des séquences comprenant six

Tableau 1

Séquences de visages (F: visage filtré, NF: visage non filtré) et mouvements articutoires associés: ouverture (ouv.), correspondant au son /A/, ou avancée (av.), correspondant au son /OU/.

Expérience 1				Expérience 2			
Écran gauche		Écran droit		Écran gauche		Écran droit	
F/ouv.	12 s	NF/ ouv.	12 s	NF /ouv.	24 s	NF /av.	24 s
NF/av.	12 s	F/av.	12 s				
F /ouv.	12 s	NF/ouv.	12 s	NF /av.	24 s	NF /ouv.	12 s
NF/av.	12 s	F/av.	12 s				
F /ouv.	12 s	NF/ouv.	12 s				
NF/av.	12 s	F/av.	12 s				
Total	72 s		72 s		48 s		48 s

Seul le mouvement d'ouverture (ouv.) est congruent avec le son diffusé (de type /A/).

Les positions droite/gauche des stimuli sont contrebalancées.

films de 12 secondes (donc d'une durée totale de 72 secondes) ont été construites et présentées (Tableau 1).

### 1.3. Procédure

Les séquences apparaissent sur deux écrans (37,6 × 30,2 cm) séparés de 5 cm l'un de l'autre et situés devant le nouveau-né. Les positions droite/ gauche des visages filtré et non filtré sont alternées toutes les 12 secondes. L'ordre de succession des vidéos (mouvement d'ouverture ou de protrusion des lèvres) d'un écran à l'autre est contrebalancé.

Les nouveau-nés sont installés sur un transat confortable, incliné à 30° et placé en face et au centre des écrans. La distance entre les yeux du bébé et les écrans est de 40 cm environ. Une caméra numérique placée entre les écrans permet d'enregistrer l'orientation du regard. Deux paires de haut-parleurs, positionnés sous les écrans, diffusent les sons de parole. Chaque écran et paire de haut-parleurs est piloté par une unité centrale indépendante.

Le visage des nouveau-nés est filmé pendant toute la présentation des séquences, les films étant ensuite analysés à l'aveugle par un juge (D.M.) qui estime la direction du regard (écran de droite versus écran de gauche).

### 1.4. Résultats

Les pourcentages moyens de regard sur l'écran présentant le visage filtré et sur l'écran présentant le visage non filtré sont représentés (Tableau 2). L'écran présentant le visage non filtré est plus

Tableau 2

Pourcentages de fixation des deux écrans lors de l'expérience 1 (visage filtré versus visage non filtré) et de l'expérience 2 (visage non filtré non congruent avec le son versus visage non filtré congruent avec le son).

Expérience 1		Expérience 2	
Visage filtré	Visage non filtré	Visage non filtré non congruent	Visage non filtré congruent
38 %	62 % dont 29 % congruent 33 % non congruent	51 %	49 %

regardé que celui présentant le visage filtré ( $M = 62\%$  vs  $M = 38\%$ ,  $t_{21} = 2,17$ , par paires, unilatéral,  $p < 0,05$ ). Les pourcentages de temps de regard sur les visages non filtrés ne sont pas différents selon que ceux-ci sont congruents ou non congruents avec le son (29 % et 33 % respectivement,  $t_{21} = 1,97$ ,  $p = 0,06$ ).

Le fait que les bébés dirigent préférentiellement leur regard vers le visage non filtré suggère qu'ils peuvent discriminer les stimuli. Le traitement des visages avait été précédemment décrit comme s'effectuant chez le nouveau-né à partir de propriétés extraites dans une gamme limitée de fréquences spatiales (fréquences comprises entre 0 et 0,5 cpd, Heering et al., 2008). Nous démontrons ici qu'au sein de cette gamme de fréquences accessibles, le bébé manifeste une préférence pour des fréquences supérieures à 0,15 cpd. Il ne semble donc pas se contenter de traiter des propriétés de bas niveau. Le fait que les bébés cherchent à appréhender plus d'informations que celles offertes par les très basses fréquences spatiales est compatible avec l'idée qu'ils pourraient s'intéresser aux caractéristiques internes du visage (cela lorsque d'autres informations, telles que des informations dynamiques et auditives, sont disponibles).

Par ailleurs, les nouveau-nés ne semblent pas percevoir la correspondance entre l'information visuelle et auditive, les stimulations congruentes n'étant pas préférées aux stimulations non congruentes. Il est néanmoins possible que dans la situation proposée, l'attention ait été capturée par le visage non filtré (versus filtré), masquant ainsi d'éventuels effets de congruence. Il est aussi possible qu'un rythme d'alternance des films toutes les 12 secondes, bien qu'ayant conduit à une préférence pour le visage non filtré, ait été insuffisante pour traiter de façon efficace une information aussi précise que la forme labiale. Pour ces raisons, nous opposerons dans l'expérience suivante deux visages non filtrés, présentant des mouvements labiaux respectivement congruents et non congruents avec les sons diffusés. Les temps d'exposition avant l'alternance seront aussi prolongés (présentation de deux films consécutifs de 12 secondes avant l'alternance).

## 2. Expérience 2 : visage non filtré congruent versus visage non filtré non congruent

Cette expérience a pour but d'analyser de façon plus précise comment le nouveau-né appréhende la mobilité labiale en lien avec le son perçu. Deux visages non filtrés sont cette fois-ci opposés, produisant des mouvements de lèvres respectivement congruents et non congruents avec les sons diffusés en boucle.

### 2.1. Participants

L'échantillon est constitué de 35 nouveau-nés (19 garçons et 16 filles), nés à terme ( $\pm 15$  jours) et en bonne santé (score d'Apgar égal à 10). Leur poids moyen à la naissance est de 3260 g (minimum : 2135 g, maximum : 4715 g) et leur moyenne d'âge est de 95 heures (minimum : 75 heures, maximum : 135 heures). Les conditions de test correspondent à celles décrites dans l'expérience 1.

### 2.2. Matériel, procédure et stimuli

Les stimuli (Fig. 1) apparaissant sur les écrans droit et gauche sont cette fois-ci deux visages non filtrés (de la même personne). L'un de ces visages (positions droite/ gauche contrebalancées) présente des mouvements de lèvres correspondant à une ouverture des lèvres tandis que l'autre visage présente des mouvements d'avancée et de protrusion des lèvres. Ces mouvements labiaux sont congruents ou non congruents avec les sons de type /A/ diffusés (naturellement associés à une

ouverture des lèvres). En plus d'être définie spatialement (par une forme des lèvres appropriée), la congruence est aussi définie temporellement puisque les mouvements de lèvres interviennent en exacte synchronie avec le son (en condition congruente) ou en léger décalage (en condition non congruente). Chaque séquence est composée de deux films identiques de 12 secondes (diffusés l'un à la suite de l'autre sans nette transition) et a donc une durée de 24 secondes (Tableau 1). Deux séquences de 24 secondes sont donc présentées l'une à la suite de l'autre sur chaque écran. L'une de ces séquences présente des mouvements articulatoires congruents avec le son /A/ (mouvements d'ouverture de la bouche). L'autre séquence présente des mouvements articulatoires incongruents avec le son /A/ (mouvement d'avancée des lèvres et de protrusion, correspondant à l'articulation du son /OU/). L'ordre est inversé pour la moitié des sujets (contrebalancement des positions droite/gauche).

### 2.3. Résultats

Les pourcentages moyens de regard sur l'écran présentant le visage net non congruent et sur l'écran présentant le visage net congruent sont représentés (Tableau 2). Ces pourcentages (51 % et 49 % respectivement) ne sont pas significativement différents ( $t_{34} = 0,89, p = 0,37$ ).

Bien que répondant à des caractéristiques dépassant le niveau des très basses fréquences spatiales (expérience 1), les nouveau-nés ne semblent donc pas en mesure d'établir des correspondances entre les sons perçus et les informations visuelles disponibles au niveau de la région labiale.

## 3. Discussion

Les résultats obtenus ici montrent que les nouveau-nés fixent plus longtemps un visage non filtré qu'un visage filtré passe-bas. Bien que leur système perceptif soit adapté au traitement des basses fréquences spatiales, il semble donc qu'ils cherchent à répondre à plus d'informations que celles offertes par les très basses fréquences (dépassement du niveau des caractéristiques externes du visage).

Il est difficile de savoir précisément quels sont les attributs que le bébé perçoit sur le visage non filtré lorsqu'il manifeste une préférence pour celui-ci. Nous pouvons penser qu'un accès à certaines caractéristiques visuelles a été rendu possible par la présence des mouvements labiaux et/ou par celle d'indices auditifs. Nous avons cumulé la présence de ces deux indices afin d'optimiser la possibilité d'une analyse perceptive visuelle sur le bas du visage. Les influences respectives des sons et des mouvements labiaux devront maintenant être analysées séparément. Cela est important car les explications proposées pour rendre compte des compétences du nouveau-né ne recouvrent bien évidemment pas la même signification selon les facteurs invoqués. Une explication en termes de sensibilité au mouvement conduira à analyser si la présence d'indices dynamiques peut éventuellement favoriser le repérage de certains attributs, présents en particulier dans la partie basse du visage (siège des mouvements articulatoires). Une explication basée sur l'apport du contexte multimodal (Bahrck et al., 2004) conduira plutôt à analyser si, à partir des indices visuels et auditifs qui lui sont fournis, le bébé peut accéder à certaines propriétés, de type amodal. L'accès à ces propriétés, rendu possible par la redondance inter-sensorielle, pourrait constituer les prémisses d'une perception cohérente du visage humain.

Une telle explication s'avère être a priori difficilement compatible avec l'observation d'une absence d'effet de congruence entre le signal auditif et le signal visuel (apparaissant au niveau des lèvres). Rappelons qu'aucune préférence pour la stimulation congruente n'a pu être mise en

évidence dans notre étude, y compris lorsque deux visages non filtrés ont été opposés (expérience 2). Cela suggère que le nouveau-né ne peut réagir que de façon imprécise aux formes et mouvements labiaux. Il a été précédemment montré qu'à l'âge de deux mois, le bébé pouvait relier les informations phonétiques aux lèvres et à la voix (Patterson et Werker, 2003). Au vu des observations effectuées ici, il semblerait que de tels liens ne puissent intervenir à la naissance. Certaines études ont pourtant révélé la présence de capacités multimodales chez le nouveau-né en situation de perception de visages. En couplant des visages articulant des voyelles simples à des sons congruents ou non congruents, Aldridge et al. (1999) ont, par exemple, fait état d'une intégration entre données auditives et visuelles. Il est possible qu'il faille passer par la modalité orale (suction non nutritive dans les travaux de Aldridge et al., 1999) pour voir émerger des capacités précoces de perception audio/visuelle de la parole. Chen et al. (2004) avaient aussi relaté, à la naissance, des capacités d'appariement en situation de perception de parole lorsque la modalité orale était impliquée (les auteurs analysant cette fois-ci les conduites d'imitation des mouvements de lèvres). Ces données sont importantes car elles contribuent à préciser les conditions de l'existence d'un traitement multisensoriel à la naissance.

Les limites observées dans notre travail se traduisent essentiellement par le fait que le bébé ne semble pas pouvoir percevoir le mouvement labial en lien avec le son. Au vu des données venant d'être relatées, il semble peu probable que ces limites puisse être liées à des difficultés générales d'intégration de l'information provenant de différentes modalités sensorielles. Nous suggérons qu'elles pourraient être plus vraisemblablement attribuables aux potentialités des systèmes perceptifs. En dépit du mouvement et du contraste accentué au niveau des lèvres, il est possible que le système n'ait pas été en mesure d'appréhender de façon suffisamment efficace les informations visuelles fournies. Bien que le nouveau-né ait manifesté une préférence pour le visage non filtré (versus filtré), il se pourrait qu'il n'ait pas eu accès à des informations suffisantes lui permettant de repérer des régularités dans le contexte multimodal. Des études ultérieures devront maintenant analyser si, en présence d'informations plus saillantes et provenant éventuellement aussi d'un plus grand nombre de sources, le nouveau-né pourrait manifester des capacités de perception des liens entre mouvements labiaux et sons de parole.

## Conflit d'intérêt

Aucun.

## Références

- Aldridge, M.A., Braga, E.S., Walton, G.E., Bower, T.G.R., 1999. The intermodal representation of speech in newborns. *Developmental Science* 2, 42–46.
- Bahrick, L.E., Lickliter, R., Flom, R., 2004. Intersensory redundancy guides the development of selective attention, perception and cognition in infancy. *Current Directions in Psychological Science* 13, 99–102.
- Buschnel, I.W.R., Sai, F., Mullin, J.T., 1989. Neonatal recognition of the mother's face. *British Journal of Developmental Psychology* 7, 3–15.
- Chen, X., Striano, T., Rakoczy, H., 2004. Auditory-oral matching behaviour in newborns. *Developmental Science* 7, 42–47.
- Dodd, B., 1979. Lip reading in Infants. Attention to speech presented in- and out-of-synchrony. *Cognitive Psychology* 11, 478–484.
- Flom, R., Bahrick, L.E., 2007. The development of infant discrimination of affect in multimodal and unimodal stimulation: the role of intersensory redundancy. *Developmental Psychology* 43 (1), 238–252.
- Haith, M.M., 1966. The response of the human newborn to visual movement. *Journal of Experimental Child Psychology* 3, 235–243.

- Heering, A., Turati, C., Rossion, B., Bulf, H., Goffaux, V., Simion, F., 2008. Newborns' face recognition is based on spatial frequencies below 0.5 per degree. *Cognition* 106, 444–454.
- Kuhl, P.K., Meltzoff, A.N., 1996. Infant vocalizations in response to speech: vocal imitation and developmental change. *Journal of the Acoustical Society of America* 100, 2425–2438.
- Lécuyer, R., Pêcheux, M.-G., Streri, A., 1994. Le développement cognitif du nourrisson. Nathan.
- Mandler, J.M., 2000. Perceptual and conceptual processes in infancy. *Journal of Cognition and Development* 1, 3–36.
- Patterson, M.L., Werker, J.F., 2003. Two-month-old infants match phonetic information in lips and voice. *Developmental Science* 6, 191–196.
- Pascalis, O., DeSchonen, S., Morton, J., Deruelle, C., Fabre-Grenet, M., 1995. Mother's face recognition by neonates: a replication and an extension. *Infant behaviour and Development* 18, 79–85.
- Piaget, J., Inhelder, B., 1966. *La psychologie de l'enfant*. PUF, Paris.
- Sai, F.Z., 2005. The role of the mother's voice in developing mother's face preference: Evidence for intermodal perception at birth. *Infant and Child Development* 14, 29–50.
- Turati, C., Cassia, V.M., Simion, F., Leo, I., 2006. Newborns' face recognition: role of inner and outer facial Features. *Child Development* 77, 297–311.