

# Dossier

## Audition-Vision

Nous savons que l'audition et la vision sont en étroite interaction dans la perception de l'environnement, tout comme dans la compréhension du langage.

C'est pourquoi, chez l'enfant sourd, il convient de s'interroger tout particulièrement sur la fonction visuelle. Pour être efficaces, les stimulations plurisensorielles doivent être adaptées aux besoins de chaque enfant. La suppléance visuelle est-elle possible et naturelle chez tous les enfants sourds ? Comment la fonction auditive, partiellement réhabilitée, doit-elle être rééduquée ? Les deux modalités sensorielles sont-elles toujours complémentaires ?

Dans cette perspective, l'ACFOS a invité des spécialistes de l'audition et de la vision lors de ses Journées d'Études pour apporter de nouvelles informations scientifiques aux professionnels de la surdité. La première partie de ce dossier présente les apports théoriques récents réalisés sur le thème audition-vision. Le prochain numéro s'interrogera sur la pertinence des pratiques éducatives, à la lumière des compétences visuelles et auditives de chaque enfant sourd.

## L'intégration audio-visuelle

PR. NAÏMA DEGGOUJ

Les informations visuelles font partie intégrante de la compréhension du discours. Le Pr Naïma Deggouj démontre, au moyen des hallucinations auditives et visuelles, que la perception de la parole est audio-visuelle.

La perception du langage est intégrée : elle se base sur l'utilisation des informations visuelles, entre autres la lecture labiale, et des informations auditives. L'intelligibilité est meilleure en présence de lecture labiale même chez le sujet normo-entendant. L'utilisation de la vision et de l'audition dans la perception du langage n'est pas une simple facilitation mais une réelle intégration, comme cela est démontré par l'effet Mc Gurk.

Nous aborderons quelques données concernant les hallucinations auditives et visuelles, la lecture labiale, les aides à la lecture labiale et l'imagerie de l'intégration audio-visuelle.

### 1. Les hallucinations auditives

Elles se produisent en cas de non-concordance entre ce qui est vu et ce qui est entendu. Deux hallucinations auditives ont été particulièrement étudiées : la ventriloquie et l'effet Mc Gurk.

#### La ventriloquie

Le ventriloque produit de la parole sans bouger les lèvres. La marionnette semble

parler. Il y a confusion entre ce que l'on voit et ce que l'on entend. Même s'il est clair que la marionnette ne peut pas parler, l'origine du son sera attribuée à celle-ci puisque les lèvres du ventriloque ne bougent pas.

Dans la vie quotidienne, cet effet se retrouve lorsqu'on regarde un film avec un son Dolby. Il est clair que le son vient de l'écran, de l'acteur parlant, alors qu'en réalité il vient de l'arrière, de droite ou de gauche.

#### L'effet Mc Gurk

L'effet Mc Gurk est produit en présence de contradictions entre les informations visuelles et auditives sur la place de l'articulation. L'information visuelle donne des informations antérieures alors que le son émis est articulé à l'arrière. La perception globale en sera modifiée. Si un montage vidéo montre une personne articulant un "ba" mais que le son émis correspond à un "ga", la perception par le téléspectateur peut être variable.

Pour certains, l'illusion de perception correspondra à une fusion entre un son articulé en antérieur et en postérieur, cor-

respondant à une perception globale d'un "fa" ou "da", ou à une combinaison des deux phonèmes sous forme d'un "bga" ou "gba".

Quelquefois, la perception du son résultante sera dominée par l'information visuelle (capture visuelle) : les personnes entendent ce qu'elles voient. L'effet Mc Gurk est absent lorsque l'information visuelle ne modifie pas la perception auditive (capture auditive) : le sujet perçoit un "ga".

L'effet Mc Gurk démontre que la catégorisation phonétique est dépendante de la vision et non pas exclusivement de l'information auditive.

#### Facteurs influençant l'effet Mc Gurk

L'hallucination perceptuelle, sous forme de fusion ou de combinaison, dépend de l'hémichamp visuel dans lequel l'image est présente. Les deux hémisphères contribuent différemment aux deux types de réponses.

La capture visuelle sera plus fréquente lorsque les sons émis sont à point d'arti-

culcation antérieure : par exemple, le “ba” dirigera plus facilement la perception. Les informations visuelles sont importantes pour juger de la place de l’articulation alors que les informations acoustiques sont surtout importantes pour la perception du mouvement et de la manière d’articuler.

La capture auditive augmente lorsque l’ambiguïté de l’information visuelle croît : lorsque le point d’articulation n’est pas visible, comme pour les sons postérieurs (“ga”), la perception sera dominée par l’information auditive.

L’effet Mc Gurk diminue lorsque la qualité de l’image est dégradée. Toutefois, cet effet se maintient même lorsque l’image est très dégradée, ne permettant plus de reconnaître le visage ou les lèvres. L’influence de la vision sur l’audition est donc probablement très globale. Elle ne se base pas sur la détection fine des mouvements péri-oraux, elle se base sur l’utilisation d’informations dynamiques, temporelles et non sur des détails. La vision périphérique intervient plus que la vision fovéale dans cette intégration.

L’effet Mc Gurk est également très résistant à une désynchronisation entre l’information visuelle et l’information auditive. Lorsque nous visionnons des films doublés, nous supportons une désynchronisation entre le son et l’image atteignant jusque 250 msec pour la parole. Par contre, la désynchronisation audio-visuelle pour les bruits est mal tolérée, dès 20 msec. Le système perceptuel tolère donc bien la désynchronisation de la parole mais pas des bruits.

L’effet Mc Gurk est également influencé par la familiarité du visage parlant. L’effet est plus manifeste avec des visages non familiers. Par exemple, cet effet est observable lorsqu’il y a non-concordance entre le sexe du locuteur et de la voix émise, si le visage n’est pas familier.

Par contre, lorsque le visage est familier, les performances chutent. Il n’y a pas d’indépendance complète entre le traitement de la reconnaissance du visage et la reconnaissance de la parole sur un

visage. La parole n’est pas traitée de la même façon s’il s’agit d’un visage familier ou non.

L’effet Mc Gurk n’est présent que chez 50% à 80% des sujets adultes. Les enfants y sont par contre moins sensibles puisqu’il n’est observé que chez 15% d’entre eux. Les jeunes enfants sont donc plus auditifs et sont moins perturbés par les informations visuelles contradictoires.

Les asiatiques sont moins sensibles à l’effet Mc Gurk, probablement en raison des facteurs culturels : il est grossier de regarder le visage parlant.

L’intégration audio-visuelle demande également une bonne connaissance de la langue. L’effet Mc Gurk n’est présent que lorsque la langue est bien dominée et non au début de l’apprentissage d’une langue étrangère.

L’effet Mc Gurk est moins fréquent chez les enfants avec troubles d’apprentissage scolaire, surtout lorsqu’ils ont des anomalies aux potentiels évoqués auditifs précoces.

## Catégorisation des perceptions phonétiques

L’effet Mc Gurk est un effet phonétique et non pas lexical. La non-congruence entre la place d’articulation et la place d’articulation entendue crée la modification de catégorisation phonétique. Cette non-congruence peut modifier la perception phonétique sans en modifier la catégorisation.

En effet, Mc Donald (2000) a démontré chez des sujets ne présentant pas d’effet Mc Gurk (c’est-à-dire sans modification des catégorisations phonétiques) que la perception du phonème est modifiée au niveau qualitatif. Il est perçu comme moins clair. La dégradation de l’image augmente la clarté du phonème perçu. Ces sujets sont donc bien perturbés dans leur perception phonétique par une information visuelle non congruente mais sans changement de catégorisation phonétique.

Il existe donc une différence entre la perception, qui est une intégration sensorielle précoce, et la catégorisation qui est une décision phonétique plus tardive et basée sur la première étape.

Rosenblum a montré chez des patients présentant un effet Mc Gurk de type fusion que le phonème perçu, en situation non congruente, donne une sensation subjective différente du phonème produit en congruence audio-visuelle : le “da” perçu lorsque le sujet voit un “ba” et entend un “ga”, est différent du “da” perçu lorsque le sujet entend et voit un “da”.

La catégorisation phonétique se base sur ce que le sujet perçoit et ce qu’il a en mémoire. Les entrées sont comparées à des prototypes phonétiques qui seraient en mémoire (théorie Fuzzy Logical, Massaro 1984).

La présence de prototypes de phonétiques dans notre cerveau est argumenté par des données électrophysiologiques. La Mismatch Negativity (MMN) est par exemple retrouvée chez des sujets ayant un effet Mc Gurk, c’est-à-dire ayant une catégorisation phonétique modifiée alors que l’information auditive est restée identique. La MMN indexe le système automatique cérébral, détectant un changement d’information phonétique. L’information phonétique reste donc bien en mémoire.

En résumé, les hallucinations auditives se produisent lorsqu’il y a contradiction entre les informations auditives et visuelles, portant sur l’information spatiale.

## 2. Les hallucinations visuelles

Les hallucinations visuelles se produisent lorsque les informations visuelles et auditives ne sont pas concordantes au niveau temporel et fréquentiel.

Si un flash est envoyé en même temps qu’un son, il est perçu comme unique. Si par contre, ce flash est envoyé en même temps que deux ou plusieurs sons, il sera perçu comme double. La vision est donc modifiée par l’information auditive temporelle.

Si des sujets sont soumis à des trains différents de stimulations visuelles et auditives, la perception visuelle sera modifiée par les changements de fréquence des stimulations auditives. La perception globale audio-visuelle est dominée par les informations auditives pour le traitement fréquentiel.

En résumé, lorsqu'un sujet est soumis à une stimulation bimodale, la perception globale sera dirigée par la modalité la plus performante en fonction du type d'analyse. Le système visuel est plus performant pour l'analyse des informations spatiales et le système auditif plus performant pour l'analyse des informations temporelles fréquentielles.

Si la discordance porte sur les informations spatiales, la perception globale sera dirigée vers le système visuel. Par contre, en cas de discordance portant sur l'aspect temporel, la perception sera plutôt dirigée par le système auditif.

### 3. La perception audio-visuelle de la parole

La perception audio-visuelle de la parole est plus que de la lecture labiale. En effet, nous utilisons les informations visuelles à notre disposition pour écouter quelqu'un : le contexte, les gestes, les mimiques, la posture, les changements dans le regard liés à l'articulation. La vision des mouvements articulatoires correspond en une translation en codes phonologiques d'informations purement optiques.

Au niveau du visage, la lecture se porte sur la région des yeux, de la bouche, de la région nasale surtout du côté droit. Le pourcentage de temps passé à regarder l'une ou l'autre région dépend des sujets et des tâches. Lorsqu'il s'agit d'un monologue, le regard portera surtout sur les yeux. Cela sera également le cas lors de l'analyse de l'intonation, de la prosodie. Par contre, les yeux seront moins regardés lors des tâches de lecture labiale de mots. Dans des conditions d'écoute plus difficile, la bouche sera plus investie.

Dans toutes les conditions, toutes les informations du visage sont utilisées.

L'intelligibilité dans le bruit est meilleure lorsqu'on peut voir tout le visage et non pas que la bouche. Le terme de lecture labiale n'est donc pas adéquat.

L'aide apportée par la "lecture labiale" est très connue chez les sourds puisqu'ils peuvent détecter les mots avec un pourcentage de succès de 90%. Cependant, l'entendant a également une intelligibilité améliorée dans le bruit et dans le calme s'il peut réaliser de la lecture labiale (Summerfield).

La vision est porteuse d'informations sur la place d'articulation. L'audition donne des informations sur le voisement et sur la manière d'articuler.

Les aides à la lecture labiale, tel que le LPC ou la AKA, permettent de donner une représentation phonologique plus précise du langage parlé. Il a été démontré que cette information visuelle est traitée comme une information phonétique (Leybaert et Charlier).

### 4. Lieu d'intégration audio-visuelle

L'intégration audio-visuelle se produit déjà à un niveau bas du système nerveux : dans le colliculus supérieur. Dans cette structure, des neurones répondent à différentes modalités sensorielles avec pour certains une réponse de type intégrative.

Ces neurones multimodaux sont distribués dans plusieurs réseaux, dans le cerveau, entre autres dans la région pariétale, autour du sulcus temporal supérieur et dans les loges frontaux.

Samson 2001 - a décrit une aire auditive spécifique pour le traitement du langage activé lors d'une analyse acoustique de mots, en modalité auditive ou visuelle.

Il s'agit d'une région multisensorielle d'intégration intervenant après le cortex auditif primaire. Cette zone s'active en présence de langage mais ne s'active pas en présence de bruit. Elle se retrouve de part et d'autres du gyrus temporal supérieur, dans le sulcus temporal supérieur, bilatéralement.

Le cortex auditif sensible pour le langage semble correspondre à une structure de décodage de la parole. Elle élaborerait des représentations neuronales d'objets sonores qui sont spécifiques à la voix et au langage.

Ces objets sonores intègrent dans cette structure multimodale des informations visuelles liées à l'analyse des mouvements de lecture labiale. Il s'agirait d'une étape indispensable pour mettre en route les réseaux neuronaux du traitement ultérieur du langage dans l'hémisphère gauche. L'expérience auditive précoce est nécessaire au développement d'un réseau bien structuré, bien cohérent.

Giraud a démontré, chez les adultes sourds porteurs d'un implant cochléaire et chez les normo-entendants, que l'étude de la parole active le cortex visuel et non pas uniquement le cortex auditif. Il y a d'emblée activation audio-visuelle.

L'activité des aires auditives et visuelles est modulée par l'interaction des stimulations auditives et visuelles.

### 5. Conclusion

L'intégration audio-visuelle favorise les informations auditives ou visuelles en fonction du type d'analyse réalisée. Lorsque la non-congruence porte sur les informations spatiales, les informations visuelles vont dominer. Lorsqu'elle porte sur les informations fréquentielles, les informations auditives vont dominer.

Il s'agit d'une réelle intégration et non pas une facilitation avec modification des activités cérébrales modulées.

L'intégration audio-visuelle du langage concerne les entendants comme les sourds. ♦

*La bibliographie de cet article est disponible à Acfos sur simple demande.*

*Pr Naïma Deggouj  
UCL Saint Luc, Bruxelles*