Émilie Troille
Université Stendhal - Grenoble 3

L'Analyse de scènes audio-visuelles : un paradigme venu de la *Gestalt*, en plein essor pour l'étude de la multimodalité du langage[1]

Résumé
Dans cet article nous déclinerons langage et images, en parole et visage, en mouvements anticipés, imaginés et en illusions du son par l'image. Ce sera l'occasion pour nous de revisiter la notion de *Gestalt* dont on a pu dire, depuis le structuralisme, qu'elle était définitivement dépassée. En ce qui concerne *Les Structures anthropologiques de l'imaginaire* de Gilbert Durand, on rappellera que la *Gestalt* n'est — même pas implicitement — une approche exclusivement *statique* de la cognition. Bien au contraire, nous montrerons que c'est à partir des mouvements qu'émergent les formes et que se stabilisent en mémoire la morphologie des gestes audibles et visibles de la bouche, saisis au vol dans le décours d'un flux de quelques quatre à six syllabes à la seconde, *via* la perception des coordinations motrices qu'il est nécessaire de maîtriser pour l'expression courante du langage entre humains. Nous appliquerons ici, pour les flux perceptifs du langage, l'*Analyse de scènes*, héritière de la *Gestalt*, y compris pour l'expression gestuelle du visage et de la main chez les sourds qui pratiquent la *Langue française Parlée Complétée*, adaptée du *Cued Speech* du Dr Richard Cornett (1967).

Mots-clés
Analyse de scènes audio-visuelles, anticipation, coarticulation, flux perceptifs, cécité, surdité, Langue française Parlée Complétée (*Cued Speech*), lien production/perception.

Abstract
In this contribution we will approach language and images in different modalities: speech and face, anticipated and imagined movements, illusions on the sound by the image. It will be the opportunity for us to revisit the Gestalt concepts which were considered obsolete since structuralism in Humanities. As instantiated by Gilbert Durand in The Anthropological Structures of the Imaginary *(1999, French 1st ed. 1960), we shall*

1. L'auteur remercie les membres de son jury de thèse, les professeurs Jean-François Bonnot, Éric Truy, Philippe Walter et Marie-Agnès Cathiard, sa directrice. Enfin elle remercie Christian Abry, pour avoir suivi ce travail depuis le début.

*recall that Gestalt is not—even implicitly—an exclusively static approach to cognition. On the contrary we will emphasize that forms can emerge from movements and stabilize in memory, thence morphology. And for speech, we will consider phonology as the stabilized outcome from audible and visible gestures of the mouth—caught in the time-course of a flow of some four in six syllables by second—*via *the perception of motor coordinations: which is necessary to master the most common expression of language between humans. We will adopt here also for the perceptual flows of spoken language,* Multimodal Scene Analysis*, the ongoing legacy of Gestalt Theory, including one of our main research topics: the gestural expression of the face and the hand for Deaf people, who practise as an augment French* LPC *(*Langue française Parlée Complétée*), adapted in 1967 from* Cued Speech *by its originator Dr Richard Cornett.*

Keywords
Audiovisual scene analysis, anticipation, coarticulation, perceptual flows, blindness, deafness, French Cued Speech*, production/perception link.*

La *Gestalt* toujours en plein essor : un paradigme à flux tendu

Accueillie au CRI en 2007, je me suis vite rendu compte que je n'avais pas rencontré dans mon cursus en sciences du langage une vue d'ensemble suffisamment large sur la constellation qui avait pris la linguistique pour pilote des Humanités dans les années d'or du structuralisme. Le livre internationalement célèbre de Gilbert Durand, publié en 1960, faisait incontestablement partie de ce mouvement, ne serait-ce que par son titre : *Les Structures anthropologiques de l'imaginaire* [*SAI*]. C'est en l'occurrence Philippe Walter qui me fit la « commande », à l'occasion de ma soutenance de thèse, de cette contribution à *Iris* pour tenter d'actualiser du point de vue de mes toutes récentes recherches les « soixante-six premières pages » du livre fondateur de G. Durand. Pour répondre à cette demande d'un point de vue qui soit au plus près l'acquis des connaissances mises en œuvre au cours de ma thèse, j'ai choisi de revenir sur une idée qui m'a été présentée de manière surprenante comme « reçue ». La *Gestalt* — toujours continuée dans les courants les plus vivants de la psychologie cognitive et, entre autres, dans l'*Auditory Scene Analysis* d'Albert Bregman (1990), le paradigme expérimental qui s'est révélé le plus adapté pour mon étude de la parole multimodale —, cette *Gestalt* serait-elle depuis le structuralisme définitivement dépassée ?

Rappelons que Gilbert Durand s'était lui-même qualifié de structuraliste « mitigé » (Durand et Chauvin, 1997, p. 88 ; repris d'un article du colloque de Chantilly, 1976, publié en 1978). Ce qu'il affichait dès la fin de son *Introduction*, où *structure* figure parmi le « vocabulaire de l'archétypologie », avec « le schème, l'archétype et le symbole, le mythe, la structure et le régime » (*SAI*, p. 60 et suiv.). Loin de moi l'intention de vouloir tenter une actualisation de l'ensemble de ces concepts et de leurs relations, ce que d'autres ont déjà réalisé en pleine compétence. Je rappellerai simplement pour mon propos que G. Durand notait alors que *structure* est un terme qui traduit en français à la fois l'allemand *Gestalt* « forme intuitive » et

Aufbau « principe organisateur » (*SAI*, p. 65). « Structure » ajoute ainsi, selon lui, à la notion de *forme*, plutôt statique, « un dynamisme transformateur » : autrement dit « structure » vaut d'abord pour « forme transformable » (*SAI*, p. 65-66). Ce qui rejoint sur cet aspect la mathématisation du structuralisme de Claude Lévi-Strauss, et de Jean Piaget qui, pour son épistémologie génétique, insistera tout autant sur cette formalisation mathématique, une structure étant définie par le *groupe de ses transformations*. Pour utiliser un exemple simple, un cube est une structure qui reste invariante quelles que soient les *opérations* de translations et/ou rotations qu'on lui applique. G. Durand, en fin de cette section terminologique, rapprochera plus dynamiquement ces structures du groupement des *symptômes* qui constituent, dans la sémiologie médicale, les *syndromes* ; précisant que c'est ce groupement en syndrome qui en fait des modèles étiologiques dynamiques, plutôt que leur aspect mathématique (*SAI*, p. 66). Pour revenir à l'exemple du cube, G. Durand dans sa critique de *L'Imaginaire* de Jean-Paul Sartre nous rappelle que, pour ce dernier, « seul le cube imaginé a d'emblée six faces ». Soit, vue par la phénoménologie philosophique de J.-P. Sartre, une pauvreté de l'image qui ne nous apprendrait rien qu'on ne sache déjà (*SAI*, p. 17). Mais que nous dit la psychologie de la perception, dont la *Gestalt*, sur ces images imaginées ? Il n'est d'abord pas du tout évident que les six faces sautent instantanément aux yeux dans cette « expérience de pensée » : il y faut même une rotation mentale qui est chronométriquement du même ordre que sa rotation effectuée réellement avec la main (Shepard et Metzler, 1971). Selon les spécialistes de la différence entre l'image auditive et l'image pictoriale (comme Reisberg *et al.*, 1989), cette dernière serait moins sujette en imaginé à la *bistabilité* du *cube de Necker* (<http://fr.wikipedia.org/wiki/Cube_de_Necker>) qu'en perception directe (voir le débat entre Kosslyn et Pylyshyn, 2003). Moins encore que dans le langage pour ce qu'on appelle *transformation verbale* : par exemple *life* => *fly* (Warren, 1961) ou le verlan (« l'envers »). Soulignons qu'un « *effet Necker* » sur des figures ou du langage (de la syllabe à la phrase, *lappe* => *plat, chat curieux* => *curieux chat*, *cent deux* => *deux cents*, *c'est bon, ça ?* => *ça, c'est bon !*) n'est pas obtenu à proprement parler par une *transformation*, comme pour une rotation mentale (voir Abry *et al.*, 2003, pour avoir mis en évidence, par IRMf, le circuit neural de cet hôte de la *boucle articulatoire* dans la mémoire de travail, rebaptisé *Stabil Loop*). Enfin, le système de la profondeur en jeu, entre autres, dans le cube de Necker a amené Ray Jackendoff (1987) à développer l'idée, venue de David Marr (1982), d'une représentation intermédiaire dite $2D^{1/2}$, émergeant avant la vision 3D imaginée. Pris ensemble, ces apports révèlent plutôt, contre sa pauvreté intrinsèque avancée par J.-P. Sartre, une richesse de l'image imaginée en faveur de la position de G. Durand.

Mais contre G. Durand, on rappellera que la *Gestalt* n'est même pas implicitement une approche exclusivement *statique* de la cognition. Ainsi, lorsqu'une image se stabilise soudain sur un autre de ses états possibles (changement de perspective du cube de Necker) et qu'elle reste stable, il n'y a évidemment rien d'autre qu'un processus dynamique des réseaux de neurones en jeu voire, comme a pu le démonter récemment Susana Martinez-Condé *et al.* (2006) pour plusieurs illusions visuelles,

des *micro-saccades* des yeux. De même avec les circuits neuraux de la vision qui permettent de faire de la forme à partir de l'ombre portée — traitement *shape-from-shading* —, et encore davantage pour la forme à partir du mouvement — traitement *shape-from-motion* ou *structure-from-motion* —, il faut définitivement considérer que la *Gestalt Psychologie* n'est pas du côté du statisme (voir, pour la cognition hors langage, Oliver Gapenne et Katia Rovira, 1999 ; pour le langage, Pierre Cadiot et Yves-Marie Visetti, 2001 ; et pour la parole, Marie-Agnès Cathiard et Christian Abry, 2007 ; enfin pour le verbal *versus* le pictorial, Abry *et al.*, 2007).

Nous ajouterons en conclusion de cette entrée en matière sur les apports récents de la psychologie une autre rencontre importante avec G. Durand, celle de la *Linguistique cognitive* défendue par Ronald Langacker (2008) et bien d'autres, avec une grammaire (comprenant sémantique et syntaxe) directement inspirée de la *Gestalt* (illustrée en France par Pierre Cadiot et Yves-Marie Visetti, 2001). Cette approche, qui s'attaque selon les mêmes principes à la métaphore (Lakoff et Johnson, 2003), convient particulièrement bien à la réhabilitation de la rhétorique à laquelle s'est livré G. Durand : « Parti […] d'une prise en considération méthodologique des données de la réflexologie, ce livre aboutit à une prise en considération pédagogique des données de la rhétorique » (*SAI*, p. 498-499) ; « la rhétorique étant le terme ultime de ce trajet anthropologique au sein duquel se déploie le domaine de l'imaginaire » (*ibid.*). Ainsi, depuis que G. Durand saluait G. Bachelard dans son *Introduction* d'un « tout est métaphorique » (*SAI*, p. 26), on comprend mieux comment, au bout du compte il a pu répéter préférer aux formalismes des structuralistes un « structuralisme figuratif » (dès la 4e de couverture de *SAI*), qui convient parfaitement aux dispositifs issus de la *Gestalt*, dans ce qu'on a pu qualifier, il y a déjà près de vingt ans, de retour d'un paradigme « formiste » ou morphologiste, dans lequel René Thom et son disciple Jean Petitot-Cocorda (ce dernier pour ses travaux sur la sémantique, la phonétique et la vision) ont occupé une place fondatrice (Gayon et Wunenburger, 1992).

La *Gestalt* dans l'*Analyse de scènes*

Avant d'exposer notre contribution proprement dite à la *Gestalt* en parole multimodale — dont la parole avec augment manuel, soit le LPC pour les sourds —, nous nous attacherons à présenter les principes de l'analyse de scènes visuelles issue des travaux des théoriciens de la *Gestalt*, puis les mêmes principes appliqués à l'analyse de scènes auditives. La perception de scènes visuelles, comme celle de scènes auditives, nécessite que le sujet soit capable de partitionner les différents plans ou flux qui s'entremêlent au sein de la scène. Cela suppose deux processus majeurs, un processus de *séparation* ou *ségrégation* qui lui permettra de séparer les composantes, visuelles ou auditives, qui ne participent pas du même objet et un processus de *groupement* ou d'*intégration*, le conduisant à regrouper ensemble les parties d'un même objet visuel ou d'une même source sonore. C'est dans le domaine de la vision que se sont développées les premières recherches, en particulier avec les psychologues

de la théorie de la *Gestalt* (ou théorie de la Forme, bien que ce terme de « Forme » soit réducteur par rapport au terme de *Gestalt*), parmi lesquels Max Wertheimer (1923, 1938), Wolfgang Köhler (1929) et Kurt Koffka (1955), qui ont dégagé les lois d'organisation perceptive d'une scène visuelle. Ces lois sont en quelque sorte des primitives visuelles, qui n'ont pas besoin d'être apprises et qui permettent d'interpréter automatiquement ce qui, dans une scène visuelle, relève de la figure ou du fond, ou encore les caractéristiques appartenant à un même objet que celles appartenant à un autre, même si ce dernier cache en partie le premier objet. Par analogie, l'*Analyse de scènes auditives*, lancée en 1990 par le psychologue Albert Bregman, est l'ensemble des processus qui permettent d'agréger ou de ségréger les différents signaux sonores émis, selon qu'ils proviennent d'une même source sonore ou de différentes sources (comme dans « l'effet Cocktail Party » de E. Colin Cherry, 1953). Depuis la publication remarquée de ce livre, nombre de travaux se sont développés (Darwin, 1997 ; McAdams et Bigand, 1994) pour mieux comprendre comment le système auditif traite, du niveau sous-cortical au niveau cortical, l'analyse auditive d'environnements sonores complexes (Pressnitzer *et al.*, 2008). La compréhension de ces principes d'analyse de scènes trouve actuellement une application directe, dans le domaine de la surdité, pour mieux appréhender les difficultés d'intelligibilité que posent les environnements bruyants aux malentendants (Grimault, 2004) et implantés cochléaires (Lancelin *et al.*, 2007).

L'Analyse de scènes visuelles

Lorsque nous observons notre environnement, nous percevons un monde organisé de surfaces et d'objets. Comment composons-nous ou décomposons-nous les informations visuelles qui atteignent notre système perceptif ?

La perception des images bistables

En général, nous n'avons pas de difficulté pour percevoir les limites des objets du monde, en contraste par rapport à un arrière-plan de surface. Il existe cependant des images plus difficiles à analyser pour notre système perceptif : les images bistables. Ces images sont aussi souvent regroupées sous l'appellation « illusions perceptives » ou « illusions d'optique » (longtemps appelées illusions « optico-géométriques »). Elles apparaissent lorsqu'il y a erreur dans l'évaluation d'une certaine propriété d'une figure ou d'un objet. Les psychologues de la *Gestalt* ont apporté à la perception visuelle les concepts de « figure » et de « fond » pour expliquer ces illusions. Confrontés à une image ambiguë, le fond et la figure peuvent être confondus. Nous devons donc séparer une forme dominante, une figure avec une découpe définie, du fond. La célèbre image ambiguë du vase et des visages, conçue par le psychologue danois Edgar Rubin en 1915, en est une bonne illustration. L'ambiguïté de cette image réside dans la difficulté de définir ce qui appartient à la figure et au fond : la figure est-elle un vase noir sur un fond blanc ou des profils blancs sur un fond noir ? En effet, nous pouvons passer d'une perception à l'autre, plus ou moins rapidement selon chaque individu, et parfois même de manière spontanée. En revanche, il est

impossible de percevoir les deux images en même temps : quand nous avons identifié une figure, les découpes semblent lui appartenir, le reste appartenant au fond.

Les lois gestaltistes d'organisation

Le gestaltisme, issu du vocable allemand *Gestalttheorie* et également nommé théorie de la forme, a été fondé dans les années 1920 par Max Wertheimer, puis repris par ses associés Wolfgang Köhler et Kurt Koffka (Boeree, 2000). Cette théorie met en valeur la prééminence de la totalité sur les parties qui la composent : « […] *things are better described as "more than the sum of their parts".* » (Behrens, 1984, p. 49). Ainsi, les psychologues de la *Gestalt* ont décrit les principes ou lois d'organisation de la perception visuelle, qui semblent être des principes fondamentaux et universels (Wertheimer, 1923).

Les principes les plus courants sont le principe de « quadrature » (permettant de regrouper des éléments même très différents parce qu'ils semblent former une figure carrée), le principe de proximité (selon lequel des éléments proches sont plus facilement groupés ensemble), le principe de similitude (où les éléments sont associés d'après leur ressemblance), le principe de bonne continuation (des lignes présentant une continuité régulière seront unifiées en un seul percept, contrairement à des lignes présentant des changements abrupts), le principe de fermeture (la perception privilégie des figures fermées plutôt que des figures ouvertes), ou encore le principe de symétrie. Des illustrations de ces principes sont présentées sur le site du sémioticien visuel Daniel Chandler (2004), ou encore sur le site de George Boeree (2000). Un principe peut l'emporter sur un autre. Ainsi, une diagonale composée de cercles sur un fond de croix sera immédiatement détectée, même si les croix sont entre elles plus proches que les cercles entre eux : le principe de similitude l'emporte ainsi sur le principe de proximité.

La perception d'objets en mouvement

Il est connu de longue date que le mouvement est un indice crucial pour la perception visuelle. La reconnaissance d'objets peut dépendre de la manière dont ceux-ci se déplacent, en particulier lorsqu'il s'agit d'êtres vivants. Nous examinons dans cette partie la façon dont notre système perceptif interprète des informations optiques dynamiques et considérons la perception des événements en tant qu'analyse des transformations du flux optique.

Supposons une scène formée de trois points l'un en dessous de l'autre : le point central se déplace le long d'une diagonale, tandis que les deux points se déplacent horizontalement. À quelle perception cet ensemble de points donne-t-il lieu, une fois mis en mouvement? Tout un chacun perçoit que le point central se déplace de haut en bas, tandis que l'ensemble des trois points se déplacent horizontalement. Bien évidemment, si le point central est présenté seul, son déplacement réel en diagonale est alors immédiatement perçu. La présence des deux points se déplaçant horizontalement modifie la perception du mouvement du premier. Notre système perceptif réalise donc en permanence une interprétation des déplacements, et la

perception que l'on a du mouvement d'un objet ne correspond ainsi pas forcément à son mouvement réel.

Gunnar Johansson (1973, 1975) a proposé une explication de ce phénomène en termes d'« analyse vectorielle perceptive ». Les deux points aux extrémités, en mouvement horizontal, partagent le même vecteur de translation horizontale, dans lequel les autres mouvements peuvent être perçus. Le déplacement du point central est décomposable en deux vecteurs : un vecteur horizontal qui correspond à l'orientation des deux autres points (H), et un vecteur vertical (V). Si l'on enlève la composante commune aux trois points, c'est-à-dire le mouvement horizontal, on voit le point central bouger verticalement. Ainsi, notre cerveau décomposerait automatiquement tout mouvement complexe en mouvements simples (vecteurs), et grouperait ensemble les vecteurs identiques.

Testant la perception du mouvement à l'aide de points lumineux (« *point lights* »), G. Johansson (1973) a prouvé qu'il suffisait de quelques points stratégiquement disposés sur les articulations pour créer la perception immédiate d'un homme en train d'effectuer une activité coordonnée telle que la marche (voir le site du *Biomotion Lab* pour des démonstrations en *point lights* : <http://www.biomotionlab.ca/demos.php>). Dans ses travaux, G. Johansson (1973) a réalisé des films de personnes en train de marcher, de courir et de danser, dans lesquels seules les lampes attachées aux articulations de ces personnes sont visibles (12 points disposés aux épaules, coudes, poignets, hanches, genoux et chevilles). En extrayant seulement une image fixe de l'un des ces films, on n'observe qu'un ensemble de points sans signification. Mais en mouvement, on peut facilement reconnaître un homme en train de courir, etc.

Ce paradigme des points lumineux a été utilisé en parole dès 1979 par Q. Summerfield, puis largement repris par Lawrence Rosenblum, Judith Johnson et Helena Saldaña (1996) pour augmenter la perception de stimuli audio bruités, et par L. Rosenblum et H. Saldaña (1996) pour tester la primauté de l'information dynamique sur l'information statique. Il faut retenir de ces études que c'est l'augmentation du nombre de points sur le visage qui permet d'aboutir aux meilleurs résultats, puisqu'elle permet de préciser d'autant mieux la position des articulateurs visibles (lèvres, langue, mâchoire), soit au final une information de forme à partir du mouvement (par une procédure « *shape-from-motion* », démontrée depuis Shimon Ullman [1979], reprise dans une approche neurocomputationnelle par Martin Giese et Tomaso Poggio [2003]). Ainsi, ce que l'on récupèrerait pourrait être à la fois les primitives cinématiques, mais également une reconstruction de la forme à partir du mouvement. En fait, selon l'échantillonnage plus ou moins fin en points, la forme sera plus ou moins bien récupérée (pour une discussion, voir Cathiard, 1994).

L'Analyse de scènes auditives

L'analyse de scènes auditives (« *Auditory Scene Analysis* », ou ASA), théorisée par Albert Bregman (1990) à partir de l'analyse de scènes visuelles, doit être comprise comme la structuration par le système auditif d'un environnement sonore complexe. A. Bregman illustre le processus de l'analyse auditive à partir de deux exemples

sonores : (i) celui d'un mélange sonore composé (ia) d'une voix d'homme prononçant le mot « choux », (ib) d'une voix qui chantonne et (ic) d'un fond musical ; et (ii) celui du mot « choux » prononcé isolément, dans le silence. Un auditeur parvient à détecter sans difficulté le mot présent dans le mélange sonore, qu'il ait été ou non prévenu du mot à entendre. Pourtant, si l'on utilise un système de décomposition spectrale des fréquences contenues dans le mélange sonore et dans le mot, il sera bien difficile de séparer, dans le mélange sonore, les fréquences appartenant au mot de celles qui constituent le reste du mélange. La décomposition spectrale du signal ne fournit pas les différentes sources sonores que repère pourtant aisément à l'oreille n'importe quel auditeur.

A. Bregman (1990) explique que l'analyse primaire de scènes auditives n'est pas liée à la connaissance spécifique des sons qui composent les voix, les instruments de musique, etc., mais plutôt à des propriétés acoustiques générales qui permettent de décomposer n'importe quel type de mélange sonore. Caroline Bey (2002) ajoute que l'analyse primaire est de type préattentif, puisque le cerveau découpe automatiquement, et de manière non consciente, le mélange sonore en entités perceptives distinctes, à partir de régularités acoustiques générales.

La première des régularités concerne la nature harmonique des sons. Les vocalisations humaines en sont un bon exemple. La vibration des cordes vocales génère une onde complexe dont la fréquence la plus basse représente la fréquence fondamentale (F0). Les autres fréquences composant cette onde sont des multiples entiers de F0, ou harmoniques. L'oreille est capable de regrouper les fréquences qui sont des harmoniques d'une fréquence fondamentale commune, pour ainsi percevoir qu'il s'agit d'un même son ou source. L'exemple le plus pertinent de cette stratégie est notre capacité à distinguer la hauteur de deux sons complexes simultanés. Pour déduire correctement la hauteur de chacun des sons, le système auditif doit sélectionner les fréquences qui lui appartiennent, en utilisant les relations harmoniques entre les composantes d'une même onde complexe. Brian Moore, Brian Glasberg et Robert Peters (1986) ont étudié un son complexe dont toutes les composantes sont des harmoniques de F0. Le *percept* est un seul son, d'une hauteur unique. En revanche, en désaccordant l'un des harmoniques de basse fréquence, les sujets perçoivent deux sons de hauteur différente. Cette première régularité peut se rapprocher du principe de *destin commun* en vision.

Une seconde régularité est la synchronisation. Il est peu probable que des sons différents débutent et se terminent tout à fait en même temps. Généralement, un des sons est déjà actif lorsque l'autre démarre, celui-ci pouvant s'arrêter avant ou après le premier. A. Bregman (1990) nomme cette stratégie « ancien-plus-nouveau ». Ainsi, un spectre qui devient plus complexe tout en conservant ses fréquences initiales est perçu comme deux signaux : un ancien auquel s'ajoute un nouveau. Richard Warren (1984) a démontré le phénomène de continuité à travers l'exemple d'un son long et pur en alternance avec des bruits plus forts. Il s'avère que l'on perçoit le prolongement du son à travers le bruit. Il semble que l'on préfère considérer qu'il s'agit d'un signal sonore continu, auquel s'ajoutent des bruits, plutôt qu'une alternance son-bruit. C'est ici le principe visuel de fermeture qui s'applique.

Enfin, notre système auditif est sensible à la progression de la transformation. Les propriétés d'un son tendent à évoluer de manière continue plutôt que soudaine tout au long du signal. Deux règles permettent d'analyser la progression de la transformation. La première règle est celle de la « transformation soudaine », qui annonce le départ d'un nouveau signal. La « continuité homophonique » (Warren, 1982) illustre cette règle. L'auteur donne l'exemple d'un son qui se maintient pendant quelques secondes à une intensité déterminée, puis devient soudainement plus fort pendant un instant, pour enfin revenir à son intensité originale. La période de plus forte intensité étant courte, elle est perçue comme un deuxième son identique au premier, et non comme une transformation du signal de départ. Notre système auditif considère que deux sons identiques sont ajoutés, c'est pourquoi l'intensité augmente. Le premier son est donc perçu tout au long de la diffusion. A. Bregman (1991) précise que cette expérience ne fonctionne pas si le changement d'intensité n'est pas soudain : dans ce cas, on percevra un changement du niveau d'intensité d'un même son, et non deux sons (ce changement doux correspondant au principe visuel de bonne continuation). La seconde règle est celle du groupement par similitude. A. Bregman (1990) indique que des similitudes au niveau des fréquences, de la localisation spatiale et du contenu spectral influencent le groupement des sons. D'après cette règle, le système auditif réunit les sons en fonction de leurs propriétés semblables pour en faire des groupes perceptifs. Chaque groupe provient ainsi d'une même source sonore. Il convient de rappeler ici l'étude de Léon Van Noorden (1975). Des sons aigus (A) et graves (G) sont alternés, avec un important intervalle de hauteur. Tant que la succession des sons est lente (c'est-à-dire que les sons sont éloignés temporellement), le système auditif perçoit une seule source ; si la séquence s'accélère, il perçoit deux sources distinctes, qui émettent presque en même temps si le tempo est très rapide. Le groupement est donc sensible au tempo, et à la différence de fréquence entre sons graves et aigus.

L'analyse des flux multimodaux en parole

La question clé que nous allons poser ici, par le biais d'un bref compte rendu d'une partie de notre recherche de thèse, est celle de la perception des flux acoustiques et optiques dans la parole, et dans la parole coordonnée avec le code manuel de Richard Cornett (1967) pour la Langue française Parlée Complétée.

La flexibilité de l'anticipation dans les flux audibles et visibles de la parole

Nous avons établi que la parole bimodale est flexible même dans les structures ConsonneVoyelle-ConsonneVoyelle (CV-CV) les plus simples, non seulement entre locuteurs mais chez un même sujet (Troille *et al.*, 2010). En effet, depuis que le phénomène de l'anticipation visuelle et son extension maximale a été établi par M.-A. Cathiard *et al.* (1991), il a souvent été répété que *la parole peut être vue avant d'être entendue*. Mais, grâce à la variabilité individuelle des performances de nos locuteurs, nous avons pu aussi montrer que *la parole peut être vue aussitôt qu'elle est entendue*.

Et dernier cas, *qu'elle peut être entendue avant d'être vue*. C'est ce cas que nous allons illustrer plus particulièrement. Nous avons choisi de nous focaliser sur un phénomène majeur en parole, celui de l'anticipation, plus particulièrement l'anticipation d'arrondissement des lèvres dans la transition intervocalique, d'une voyelle non arrondie [i] à une voyelle arrondie [y], entre lesquelles peuvent s'insérer une ou plusieurs consonnes, permettant au geste d'arrondissement de se déployer par avance. La séquence qui a retenu notre attention est une séquence « …zizu… » [zizy] (fig. 1). Parce que ce stimulus présente des propriétés structurelles qui pouvaient offrir les opportunités suivantes : (i) comporter un geste d'arrondissement de voyelle à voyelle [iy], qui soit principalement le responsable du changement acoustique pertinent pour la langue étudiée, en l'occurrence le français ; (ii) avec un déroulement de cette transition durant une consonne fricative voisée [z] ; (iia) ceci afin d'inter-

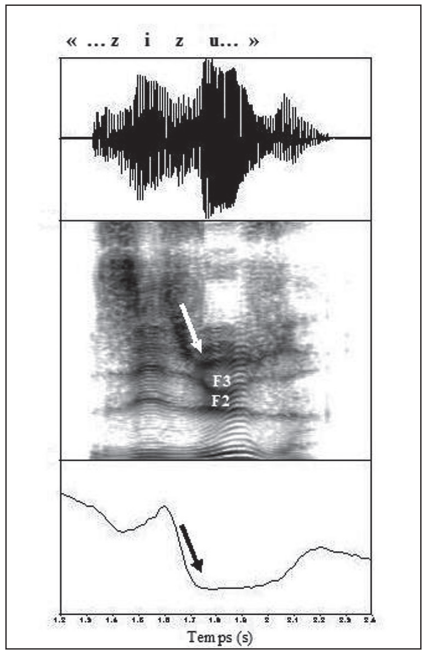

Figure 1. – Gestalt *soit* Analyse de scène *pour un* pattern *bimodal, percept audiovisuel test « …zizu… » [zizy].*

En haut, signal acoustique avec, au centre, l'analyse de ses composantes harmoniques et formantiques (spectrogramme de 0 à 8 000 Hz). En bas, évolution de l'aire aux lèvres (de 0 à 2 cm²). La flèche descendante en blanc sur le spectrogramme indique l'abaissement du bruit de friction de la consonne, synchrone du mouvement de diminution de l'aire aux lèvres (flèche noire). Le rapprochement des deux formants F2 et F3, groupement focalisant caractéristique de la voyelle [y], n'a lieu qu'après ce mouvement du bruit de friction. En perception, le *flux* audio de la voyelle dans la consonne — le fait qu'on puisse à proprement parler entendre le mouvement des lèvres vers la voyelle à travers le bruit consonantique — permet d'anticiper l'identification de cette voyelle [y] à venir ; tandis que la structure acoustique formantique caractéristique de [y] n'est pas encore présente.

rompre le moins possible le flux acoustique ; (iib) tout en offrant un bruit de friction dans les hautes fréquences, suffisant pour qu'il puisse porter les changements de résonance correspondant au geste d'arrondissement des lèvres, bien au-dessus de la gamme des formants caractérisant le changement des voyelles. Ainsi, le but est bien de comparer l'efficacité perceptive des flux unimodaux (auditif et visuel) et bimodal (audiovisuel) avec un paradigme permettant de suivre, dans son décours temporel naturel, le flux lié à la voyelle à travers le flux de la consonne, avec une consonne perméable aux effets de coarticulation. Le stimulus [zizy] et son contrôle [zizi], tronqués à partir du début de la consonne toutes les 20 ms, de manière à dévoiler plus ou moins d'informations sur la voyelle à percevoir (paradigme du *gating* ; François Grosjean, 1980), sont présentés aux sujets en ordre aléatoire : leur tâche est d'identifier à chaque séquence la voyelle [y] ou [i] à venir.

Les résultats d'identification de la voyelle [y] en condition auditive (fig. 2) ont démontré une perception de la voyelle arrondie dès la fin du premier tiers de la consonne [z] intervocalique : soit une avance d'identification de la voyelle de plus de 90 ms. Elle est suivie de l'identification audiovisuelle, avec un retard d'environ 20 ms, alors que la perception visuelle est la plus tardive, montrant un retard de 45 ms sur l'audition. Pour comprendre ces résultats, nous avons pu bénéficier de la possibilité de relier nos données en perception à nos données en production (avec le suivi du décours de l'aire aux lèvres). Nous avons ainsi pu observer que le mouvement du bruit de friction précédait celui du troisième formant dans la transition de [i] vers [y], et que les mouvements d'aire aux lèvres et de bruit de friction étaient

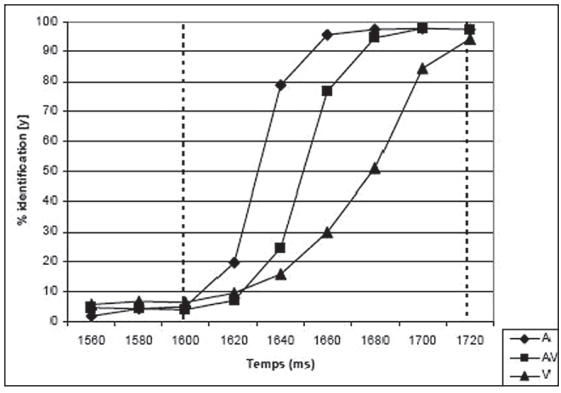

Figure 2. – Résultats d'identification perceptive [y] des sujets entendants-voyants pour le même stimulus test « …zizu… » [zizy] qu'en figure 1 (les lignes verticales pointillées délimitent la tenue de la consonne [z] devant [y]).

À mesure que les sujets ont accès (au hasard) à une portion plus longue de la suite du signal, leurs résultats d'identification des trois *flux* croissent selon une fonction en S bien connue. On voit qu'avant même l'arrivée de la voyelle proprement dite, les anticipations grâce aux informations auditives (première courbe, A) gagnent sur les informations visuelles (dernière courbe, V). Les identifications bimodales audiovisuelles (AV) étant intermédiaires entre les deux modalités simples : au bout du compte, la perception monomodale auditive gagne significativement sur toutes les autres.

synchrones et descendants dans la première partie de la consonne [z] (fig. 1). La voyelle [y] était identifiée en audio aussitôt que le mouvement du bruit de friction descendait au cours de la consonne [z] ; tandis qu'en vision seule, les sujets avaient besoin d'attendre un degré de constriction labiale plus important avant d'identifier décisivement la voyelle aux lèvres, ce qui correspond à des effets différents de l'arrondissement, plus précoce sur la fricative que sur la structure des formants vocaliques. Le fait que la courbe d'identification auditive soit en avance de 20 ms sur la perception audiovisuelle révèle l'intégration par nos sujets des deux modalités auditive et visuelle, avec un effet de ralentissement de l'identification auditive par la modalité visuelle.

Avec un même stimulus [zizy], enregistré chez le même locuteur près de dix ans auparavant, Pierre Escudier *et al.* (1990) avaient observé le résultat classique de l'avance de la vision sur l'audition (avec un geste d'arrondissement perçu visuellement 40 à 60 ms avant sa perception acoustique). L'analyse articulatoire du stimulus révèle que le geste labial d'arrondissement des lèvres pour la voyelle [y] est très anticipé dans la production de ce locuteur : l'établissement du geste de constriction de [y] se fait en effet pendant la deuxième partie de la voyelle [i] précédente, et le minimum de constriction caractéristique de [y] est atteint en début de consonne [z]. L'anticipation articulatoire est donc particulièrement ample, ce qui explique cette avance de la vision sur l'audition.

Nous avons à nouveau utilisé le même paradigme d'anticipation en contexte CV-CV, en enregistrant cette fois la production de la séquence [zizy] par une locutrice. Les tests perceptifs n'ont montré aucune différence entre la perception des trois modalités auditive, visuelle et audiovisuelle. Pour ce stimulus, l'information visuelle portée par la diminution de l'aire aux lèvres est seulement un peu en avance sur l'information auditive portée par l'abaissement du bruit de friction. Ce mouvement d'aire aux lèvres débute en effet dans la fin de la voyelle [i], quelques 40 ms avant l'abaissement du bruit de friction de la consonne : ce qui aboutit à un cas d'anticipation intermédiaire, entre celui où le geste articulatoire est extrêmement précoce, se déployant dès la voyelle non arrondie [i] (comme dans l'étude de P. Escudier *et al.*, 1990) et le cas où le geste articulatoire est synchrone du décours acoustique, comme dans notre première expérience. Ainsi, pour ce troisième cas, la parole peut être entendue aussi précocement qu'elle est vue.

Ceci démontre la nécessité de mettre en relation les données de perception et les données de production, ce qui comprend la possibilité de rendre compte des effets de la production sur le visible et l'audible. Ainsi, la perception de la parole dépend fortement du timing d'anticipation de la coordination des gestes articulatoires en production. En ce qui concerne l'information vocalique, la coarticulation consonantique peut porter plus précocement l'information auditive que la voyelle elle-même, en fonction de la structure du stimulus. En ce qui concerne le timing des informations auditives et visuelles, il devient de plus en plus clair que les dates de délivrance de chaque composante du flux perceptif peuvent changer drastiquement les résultats obtenus.

Quelle leçon générale peut être tirée de ces expériences en ce qui concerne l'organisation des flux acoustico-optiques de la parole et de leur contrôle ? L'occasion que nous avons eue de tester le même locuteur à plus de dix ans d'intervalle a fourni un démenti clair de l'opinion, aujourd'hui courante, selon laquelle la parole serait systématiquement vue avant d'être entendue. Nous proposons de reformuler cette proposition de la façon suivante : l'organisation temporelle des flux auditif et visuel doit être considérée comme le résultat d'une importante flexibilité de leurs coordinations chez un même locuteur et entre locuteurs.

Les coordinations perception-production de la Langue française Parlée Complétée ou quand le corps gagne sur le code

La Langue française Parlée Complétée, issue du *Cued Speech* de R. Cornett (1967), a été créée pour améliorer la perception de la langue orale par les malentendants, en complétant manuellement les formes labiales ambiguës (les positions de la main sur le visage codant les voyelles et les configurations digitales codant les consonnes, avec recodage de la chaîne parlée en une suite de syllabes CV). Grâce à ce système, combinant le geste de la main à celui des lèvres, la totalité de l'information phonologique peut être récupérée par le sujet sourd décodeur au travers de la seule modalité visuelle. L'efficacité du code a été démontrée depuis longtemps et par plusieurs études, que ce soit aux niveaux des représentations phonologiques, du développement lexical et morphosyntaxique, du développement des habiletés métaphonologiques, de la mémoire et de la lecture (LaSasso, Crain et Leybaert, 2010).

Mais il a fallu attendre l'étude de Virginie Attina (2005) pour que sa production soit étudiée. Analysant la coordination des clés avec le son dans la syllabe CV, l'auteure a démontré que l'information de la main anticipe l'information des lèvres en LPC. Plus exactement, l'information vocalique transmise par la position de la main devance l'information vocalique transmise par les lèvres ; la configuration consonantique de la main, quant à elle, pointe la position du visage en synchronie avec la consonne labiale.

Nous avons repris, avec nos sujets malentendants décodeurs, le paradigme de l'anticipation du flux vocalique au travers du flux consonantique. Nous avons adapté nos stimuli [zizy] et [zizi] utilisés dans nos expériences avec des sujets entendants-voyants, en leur associant un codage LPC par incrustation d'une main, permettant de désambiguïser le mouvement labial. L'image d'une main a été pour cela artificiellement synchronisée sur le visage du locuteur, selon les valeurs d'anticipation de la main sur les lèvres établies pas V. Attina (2005 ; Attina *et al.*, 2004). Et nous avons soumis des sujets sourds décodeurs aux mêmes tests d'identification [y]/[i], mais cette fois en modalité visuelle (lecture labiale seule) et en modalité visuelle accompagnée de code LPC. Les résultats permettent de dégager plusieurs constats : (i) les performances en condition LPC apparaissent supérieures par rapport à la condition visuelle seule, même si dans les deux cas la voyelle [y] a été identifiée précocement dans la consonne [z], avec un gain obtenu en condition LPC de 66 ms (fig. 3) ; l'avance de la main sur les lèvres, ici recréée comme augment, est bien utilisée

perceptivement par les sujets, comme au naturel ; (ii) la comparaison des résultats en LPC de nos sujets sourds aux résultats de nos sujets entendants nous a permis de constater que les sujets sourds en LPC peuvent être aussi performants que les sujets entendants le sont dans la modalité dans laquelle leur identification est la plus précoce, ici la perception auditive. Autrement dit, la main peut à temps suppléer au son en Langue française Parlée Complétée.

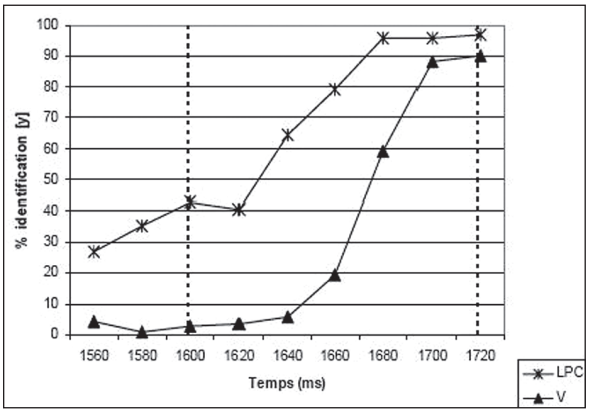

Figure 3. – Résultats d'identification perceptive [y] des sujets sourds décodeurs en Langue française Parlée Complétée (LPC) pour le même stimulus test « …zizu… » [zizy] qu'en figure 1, lorsqu'on leur permet de bénéficier d'un augment ou flux *manuel LPC, soit du mouvement d'une main naturelle (image incrustée).*

La courbe de droite en condition visuelle seule (V) présente la même anticipation (avant l'arrivée de la voyelle [y], ligne pointillée de droite) que celle des sujets entendants-voyants de la figure 2. Noter par contre l'avance qu'apporte la main (courbe LPC), qui rejoint l'avance donnée par le son (A en figure 2). Autrement dit, la main peut à temps suppléer à l'oreille en Langue française Parlée Complétée.

Ainsi, en examinant soigneusement la structure des stimuli testés, nous avons pu montrer que les patrons perceptifs résultants sont « rivés » (*locked*) à la production oro-faciale de la parole. Ce qui se démontre en tenant compte des relations articulatori-acoustiques. Nos expériences de *gating*, menées avec des entendants et des sourds, nous ont permis de tester la gamme de flexibilité que peut permettre cette coordination phonémique unique de la face et de la main. Ces résultats viennent renforcer la conception avancée depuis V. Attina *et al.* (2004), dans laquelle le comportement anticipatoire dans la Langue française Parlée Complétée repose sur la mise en phase des types de contrôles les plus compatibles, ceux des contacts de la main avec le visage pour les voyelles et ceux des constrictions de la bouche pour les consonnes. C'est le constat régulier de cette avance, et le phasage bien particulier des gestes vocaliques et consonantiques qui montrent un contrôle inversé en LPC par rapport à la parole, qui a conduit V. Attina *et al.* (2004) à parler d'une « *topsy-turvy vision of the Cued Speech landscape* » (p. 209), soit l'inverse d'un langage labial com-

plété par la main. La fenêtre qui nous a été ainsi ouverte par le code de R. Cornett — surtout par la tournure qui lui a permis d'être neuralement incorporé (*embodied and "embrained"* selon les termes de Tim Van Gelder, 1999) dans une habileté linguistique — nous a ainsi permis, de manière *a priori* surprenante, d'apporter des réponses plus décisives sur la nature des contrôles des segments dans la phonologie du langage que par la seule observation des actes de parole.

Nous espérons avoir réussi — sinon à démontrer — du moins à faire entrevoir les prolongements des intuitions de ce structuralisme figuratif si spécifique à G. Durand, dans les travaux les plus récents en psychologie cognitive issus directement de la *Gestalt*. On retiendra notre insistance aussi bien sur la dynamique de cette *Gestalt*, que sur la stabilité des formes linguistiques issues de ce traitement des flux perceptifs multimodaux (du son, du visage et de la main).

Bibliographie

Abry Christian, Cathiard Marie-Agnès et Diaferia Marie-Laure, « Enactive Art: Parietal and Frontal Brain Art? From Pictorial to Speech Evidence », *Proceedings of the 4th International Conference on Enactive Interfaces* (19-24 novembre, Grenoble, France, 2007), p. 25-28.

Abry Christian, Sato Marc, Schwartz Jean-Luc, Lœvenbruck Hélène et Cathiard Marie-Agnès, « Attention-based maintenance of speech forms in memory: The case of verbal transformations », *Commentary on target paper "Working Memory Retention Systems: A State of Activated Long-Term Memory"* by Ruchkin, Grafman, Cameron and Berndt, *Behavioral and Brain Sciences*, vol. 26, nº 6, 2003, p. 728-729.

Attina Virginie, *La Langue française Parlée Complétée (LPC) : Production et Perception*, thèse de sciences cognitives, Institut national polytechnique de Grenoble, 2005.

Attina Virgine, Cathiard Marie-Agnès, Beautemps Denis et Odisio Matthias, « A pilot study of temporal organization in Cued Speech production of French syllables: Rules for a Cued Speech synthetizer », *Speech Communication*, vol. 44, nos 1-4, 2004, p. 197-214.

Behrens Roy, *Design in the visual arts*, Englewood Cliffs, Prentice-Hall, 1984.

Bey Caroline, « Rôle des connaissances dans la construction de la scène auditive : ce que j'entends dépend-il de ce que je cherche à entendre ? », *Fondation Fyssen, Annales*, nº 17, 2002, p. 65-73.

Boeree C. George, *Gestalt Psychology*, <http://www.ship.edu/~cgboeree/gestalt.html>, 2000.

Bregman Albert S., *Auditory Scene Analysis: The Perceptual Organization of Sound*, Cambridge, Massachusetts, MIT Press, 1990.

—, « Using quick glimpses to decompose mixtures », dans J. Sundberg, L. Nord et R. Carlson (éds), *Music, language, speech, and brain*, Londres, MacMillan, 1991, p. 284-293.

Cadiot Pierre et Visetti Yves-Marie, *Pour une théorie des formes sémantiques – motifs, profils, thèmes*, Paris, Presses universitaires de France, 2001.

Cathiard Marie-Agnès, *La perception visuelle de l'anticipation des gestes vocaliques : cohérence des événements audibles et visibles dans le flux de la parole*, thèse de doctorat de psychologie cognitive, Université Pierre Mendès France (Grenoble), 1994.

Cathiard Marie-Agnès et Abry Christian, « Speech structure decisions from speech motion coordinations », *Proceedings of the XVIth International Congress of Phonetic Sciences* (6-10 août, Saarbrücken, Allemagne, 2007), p. 291-296.

Chandler Daniel, *Gestalt principles of visual organization, visual perception*, <www.aber.ac.uk/media/Modules/MC10220/visper07.html>, 2004.

Cherry E. Colin, « Some experiments on the recognition of speech, with one and with two ears », *Journal of the Acoustical Society of America*, vol. 25, nº 5, 1953, p. 975-979.

Cornett Richard Orin, « Cued Speech », *American Annals of the Deaf*, vol. 112, 1967, p. 3-13.

Darwin Chris J., « Auditory grouping », *Trends in Cognitive Sciences*, vol. 1, nº 9, 1997, p. 327-333.

Durand Gilbert, *Les Structures anthropologiques de l'imaginaire*, Paris, Dunod, 1960.

Durand Gilbert et Chauvin Danièle, *Les champs de l'imaginaire*, Grenoble, Ellug, 1997.

Escudier Pierre, Benoit Christian et Lallouache Tahar, « Identification visuelle de stimuli associés à l'opposition /i/-/y/ : étude statique », *Actes du premier congrès d'Acoustique* (10-13 avril, Lyon), *Suppl. au Journal de Physique*, nº 2, 1990, p. 541-544.

Gapenne Oliver et Rovira Katia, « Gestalt Psychologie et cognition sans langage. Actualité d'une figure historique », *Intellectica*, nº 1, 1999, p. 229-250.

Gayon Jean et Wunenburger Jean-Jacques (dir.), *Les Figures de la forme*, Paris, L'Harmattan, 1992.

Giese Martin A. et Poggio Tomaso, « Neural Mechanisms for the recognition of biological movements », *Nature Reviews Neuroscience*, nº 4, 2003, p. 179-192.

Grimault Nicolas, « Analyse séquentielle des scènes auditives chez le malentendant », *Revue de Neuropsychologie*, nº 14, 2004, p. 25-39.

Grosjean François, « Spoken word recognition processes and the gating paradigm », *Perception & Psychophysics*, nº 28, 1980, p. 267-283.

Jackendoff Ray, *Consciousness and the Computational Mind*, Bradford Books / MIT Press, 1987.

Johansson Gunnar, « Visual perception of biological motion and a model for its analysis », *Perception and Psychophysics*, nº 14, 1973, p. 201-211.

—, « Visual motion perception », *Scientific American*, nº 232, 1975, p. 76-88.

Koffka Kurt, *Principles of Gestalt Psychology*, Londres, Routledge & Kegan Paul, 1955.

Köhler Wolfgang, *Gestalt psychology*, New York, Liveright. Traduction française (1964) : *Psychologie de la forme*, Paris, Gallimard, 1929.

Kosslyn Stephen M., Ganis Giorgio et Thompson William L., « Mental imagery: against the nihilistic hypothesis », *Trends in Cognitive Sciences*, vol. 7, nº 3, 2003, p. 109-110.

Lakoff George et Johnson Mark, *Metaphors We Live By*, University of Chicago Press, 2003 edition contains an 'Afterword'.

Lancelin Denis, Gnansia Dan et Lorenzi Christian, « Démasquage de la parole dans le bruit chez les sujets entendants, malentendants et implantés cochléaires », *Les Cahiers de l'Audition*, vol. 20, nº 3, 2007, p. 54-57.

Langacker Ronald W., *Cognitive Grammar: A Basic Introduction*, New York, Oxford University Press, 2008.

LaSasso Carole, Crain Kelly et Leybaert Jacqueline, *Cued Speech and Cued Language for deaf and hard of hearing childen*, San Diego, Plural Publishing Inc, 2010.

McAdams Stephen et Bigand Emmanuel, *Penser les sons : Psychologie cognitive de l'audition*, Paris, PUF, 1994.

Marr David, *Vision*, San Francisco, Freeman editors, 1982.

Martinez-Conde Susana, Macknik Stephen L., Troncoso Xoana G. et Dyar Thomas A., « Microsaccades counteract visual fading during fixation », *Neuron*, nº 49, 2006, p. 297-305.

Moore Brian C. J., Glasberg Brian R. et Peters Robert W., « Thresholds for hearing mistuned partials as separate tones in harmonic complexes », *Journal of the Acoustical Society of America*, nº 80, 1986, p. 479-483.

Pressnitzer Daniel, Sayles Mark, Micheyl Christophe et Winter Ian M., « Perceptual organization of sounds begins in the auditory periphery », *Current Biology*, vol. 18, nº 15, 2008, p. 1124-1128.

Pylyshyn Zénon, « Explaining mental imagery: now you see it, now you don't. Reply to Kosslyn *et al.* », *Trends in Cognitive Sciences*, vol. 7, nº 3, 2003a, p. 111-112.

—, « Return of the mental image: are there really pictures in the brain? », *Trends in Cognitive Sciences*, vol. 7, nº 3, 2003b, p. 113-117.

Reisberg Daniel, Smith J. David, Baxter David A. et Sonenshine Marcia, « Enacted auditory images are ambiguous; pure auditory images are not », *The Quarterly Journal of Experimental Psychology*, vol. 41, nº 3, 1989, p. 619-641.

Rosenblum Laurence D., Johnson Judith A. et Saldaña Helena M., « Visual kinematic information for embellishing speech in noise », *Journal of Speech and Hearing Research*, vol. 39, nº 6, 1996, p. 1159-1170.

Rosenblum Lawrence D. et Saldaña Helena M., « An audiovisual test of kinematic primitives for visual speech perception », *Journal of Experimental Psychology: Human Perception and Performance*, vol. 22, nº 2, 1996, p. 318-331.

Shepard Roger et Metzler Jacqueline, « Mental rotation of three dimensional objects », *Science*, vol. 171, nº 3972, 1971, p. 701-703.

Troille Émilie, *De la perception audiovisuelle des flux oro-faciaux en parole à la perception des flux manuo-faciaux en Langue française Parlée Complétée. Adultes*

et enfants : entendants, aveugles ou sourds, thèse de sciences du langage, Université Stendhal-Grenoble 3, 2009.

Troille Émilie, Cathiard Marie-Agnès et Abry Christian, « Speech face perception is locked to anticipation in speech production », *Speech communication*, vol. 52, nº 6, 2010, p. 513-524.

Ullman Shimon, *The interpretation of visual motion*, Cambridge, Massachusetts, MIT Press, 1979.

Van Gelder Tim, « Dynamic approaches to cognition », dans R. Wilson et F. Keil (éds), *The MIT Encyclopedia of Cognitive Science*, Cambridge, The MIT Press, 1999, p. 244-246.

Van Noorden Léon, *Temporal coherence in the perception of tone sequences*, Doctoral dissertation, Eindhoven University of Technology, 1975.

Warren Richard M., « Illusory changes of distinct speech upon repetition – the verbal transformation effect », *British Journal of Psychology*, vol. 52, nº 3, 1961, p. 249-258.

—, *Auditory perception: A new synthesis*, Elmsford, Pergamon Press, 1982.

—, « Perceptual restoration of obliterated sounds », *Psychological Bulletin*, vol. 96, 1984, p. 371-383.

Wertheimer Max, « Untersuchungen zur Lehre von der Gestalt II », *Psychological Research*, vol. 4, nº 1, 1923, p. 301-350.

—, *A Source Book of Gestalt Psychology*, New York, Harcourt, Brace and Co., 1938.