

L'observation et l'enregistrement des mouvements dans la parole; problèmes et méthodes

Bernard TESTON

Laboratoire Parole et Langage Université de Provence

29 avenue R. Schuman

13621 Aix en Provence

courriel : teston@lpl.univ-aix.fr

Introduction

Le mouvement des êtres et des choses a été interprété de tout temps par l'homme comme la manifestation du « vivant » naturel ou divin. Depuis l'antiquité il s'est donc attaché à l'observer pour tenter d'en expliquer les mécanismes et les causes. Les premières observations codifiées du mouvement se rapportent à des phénomènes naturels lents autorisant une description aisée en particulier à celles des astres dans le but principal d'en donner des prédictions. Les mouvements naturels des êtres vivants ont été très tôt représentés au moyen de dessins, peintures ou sculptures par des « instantanés » caractéristiques d'une espèce ou d'une attitude sociale souvent schématique, et d'une réalité physiologique le plus souvent erronées. Ceci est dû à l'impossibilité sensorielle de l'homme d'appréhender et mémoriser des mouvements rapides en l'absence de techniques appropriées d'enregistrement et de restitution des images ou des manifestations physiques des mouvements. Pour cette raison, aucune description physiologique des mouvements n'apparaît avant le milieu du 19^{ème} siècle.

Notre préoccupation est l'étude des mouvements mis en jeu dans la communication parlée. Cette étude ne peut donc pas être dissociée d'une étude simultanée des sons de parole.

Nous distinguons deux grandes classes de mouvements. D'une part ceux qui, pendant le discours, apportent dans des proportions variables un complément d'information au message parlé et ne participent pas directement aux mécanismes de sa création. Nous les appelons, par simplification, mouvements d'accompagnement de la parole.

D'autre part, les mouvements des organes articulateurs, qui permettent de produire la parole, que nous étudions dans le cadre d'une phonologie articulatoire et appelons gestes articulatoires.

Les mouvements d'accompagnement de la parole sont généralement des mouvements de mimiques de la face, des mouvements de la tête et des segments des membres supérieurs. Ils suivent surtout les variations prosodiques du discours et sont caractérisés par une vitesse de réalisation relativement faible. La technique de choix pour l'étude de ces mouvements est la vidéocinématographie qui permet de les décomposer image par image et ainsi « d'arrêter le temps ». Les gestes articulatoires sont par contre beaucoup plus rapides car ils conditionnent la réalisation des structures phonémiques. Leurs mouvements doivent être connus avec plus de précision et enfin les principaux organes articulateurs sont invisibles car internes au conduit vocal et outre l'imagerie, de nombreux dispositifs de mesure ont dû être imaginés pour appréhender indirectement leurs mouvements.

Il nous a semblé impossible de présenter les méthodes actuelles d'analyse du mouvement sans mentionner au préalable l'importance des travaux d'Etienne Jules Marey qui fut en son temps un savant d'une envergure mondiale et dont l'apport à l'analyse physiologique du mouvement est considérable. Nous utilisons en permanence les concepts et les méthodes qu'il a imaginées voici plus d'un siècle et qui ont certes été améliorés depuis, au rythme des évolutions technologiques, mais dont les principes restent les mêmes.

1-Etienne Jules Marey et L'analyse du mouvement

Médecin et physiologiste, Etienne Jules Marey s'est intéressé à toutes les manifestations fondamentales de la vie et en particulier aux propriétés des mouvements des êtres et des choses. Mécanicien de génie, il est obsédé par la transcription graphique des phénomènes c'est à dire par leur enregistrement, qui seul permet de les étudier dans le détail. Son unique but est de visualiser sous la forme de lignes sinueuses bien visibles leurs manifestations les plus impalpables.

Bien ancré dans le positivisme de son temps, Marey a pour seule ambition de constater des faits et d'en déduire les seules lois que l'expérience contrôle.

Pour capter le déroulement au cours du temps des phénomènes, il a inventé et perfectionné des méthodes graphiques d'enregistrement avec inscription au moyen de stylets, actionnés par des capsules manométriques, sur du papier noirci de fumée (Marey-1878). Ce sont les fameux cylindres inscripteurs de Marey.

L'apparition des plaques photographiques instantanées au bromure d'argent permet à Marey d'imaginer une méthode d'enregistrement photographique des mouvements, la «chronophotographie sur plaque fixe» (Marey-1882). Cette méthode ingénieuse est basée sur la photographie rapide d'un sujet mobile blanc devant un fond noir. La multiplication des poses est obtenue par la rotation devant la plaque, d'un disque à segments ; « le corps s'écrit lui même, il ne reste qu'à lire ». Cette méthode permet de visualiser jusqu'à 50 poses sur une même plaque avec une résolution temporelle d'une vingtaine de poses par seconde. Elle répond bien au goût de la précision de Marey car elle traduit avec la même fidélité qu'un graphique les rapports d'espace et de temps et les distances parcourues qui sont l'essence du mouvement.

Dès 1883, il en améliore la mesure quantitative grâce à la « chronophotographie géométrique » qui permet d'enregistrer des points et des segments dessinés en blanc sur un mobile noir.

Cependant, pour les mouvements lents, les images qui se superposent contrarient une lecture précise des clichés. En 1888, Marey a l'idée de déplacer la surface photo sensible et conçoit un dispositif qui permet d'enregistrer les images sur une pellicule mobile. Il réalise pour cela un mécanisme de défilement et d'arrêt rapide de la pellicule papier qui autorise 50 photographies par seconde (Marey-1888). En 1893 il présente un projecteur à partir de pellicules « film » qui permet de reproduire le mouvement. Tout le cinéma est déjà là. Mais Marey apprécie peu les projections animées si elles ne correspondent pas à une analyse scientifique du mouvement par le ralenti. Il a été cependant en contact permanent avec les frères Lumière qui lui fournissaient ses plaques photographiques et qui proposèrent le « cinématographe » en 1895.

2-Les grands principes d'enregistrements des mouvements

Le mouvement peut être considéré comme un événement de variation continue que nous ne pouvons appréhender naturellement dans sa totalité qu'à la condition que ses évolutions dans l'espace et le temps soient très lentes. Pour cette raison évidente, il est nécessaire, pour étudier le mouvement, de l'enregistrer afin d'avoir une trace permanente de son évolution permettant son étude dans le détail sans contrainte de temps.

Pour cela ont d'abord été développés des capteurs kinésiographiques capables de transmettre la dynamique d'un mouvement associés à des dispositifs inscripteurs (Marey-1878) dont le premier et non le moindre exemple est le Kymographe utilisé par Rousselot (1897) pour l'étude des mouvements des organes articulateurs. Depuis, les capteurs manométriques ou mécaniques ont gagné en sensibilité, dynamique, fidélité et rapidité grâce aux progrès des techniques de l'électronique de même que les enregistreurs graphiques sur différents supports qui maintenant sont quasiment tous informatisés.

L'information directe du mouvement étant essentiellement visuelle il a été imaginé de l'enregistrer sous la forme d'une succession d'images instantanées. Le mouvement, dont l'évolution est continue, est alors discrétisé en un certain nombre d'images ou poses par unité de temps et les informations de ses variations rapides sont bien évidemment fonction de ce nombre d'image.

Plus un mouvement est rapide et plus le nombre d'images par unité de temps doit être grand. La technique du cinématographe consiste à réaliser une photographie pour chaque image sur un film au moyen d'une caméra. Pour éviter le flou de ces photographies, un obturateur ne prélève la lumière que pendant un court instant. On utilise un projecteur à la même vitesse que la caméra pour restituer le mouvement sous la forme d'une image animée. On connaît l'importance considérable qu'a pris le cinématographe dans l'industrie du spectacle. On connaît moins par contre celle qu'il a eu dans l'étude scientifique du mouvement qui est fondamentale par le fait qu'il permet non seulement l'enregistrement mais la rediffusion du mouvement ainsi que sa décomposition fine avec le ralenti et l'arrêt sur image. Le cinéma standard (qui utilise la persistance rétinienne pour reconstituer le mouvement avec un nombre réduit d'images) s'est fixé rapidement à 24 images par seconde. Pour l'étude de phénomènes rapides on a très vite réalisé des caméras capables de filmer plusieurs milliers d'images par seconde.

La prise et diffusion d'images électroniques (et non plus argentiques sur film photo) par télévision (vidéocinématographie ou par contraction vidéo) est basée sur le même principe et, est en train de remplacer le cinéma dans toutes ses applications. Pour l'étude du mouvement elle s'est développée surtout depuis que l'on a pu réaliser des moyens d'enregistrement et de rediffusion vidéo (magnétoscopes) commodes et efficaces. Parallèlement, les capteurs d'images vidéo ont énormément évolué en sensibilité, résolution et netteté. La vitesse de prise d'image vidéo est de 30 (USA et Japon) ou 25 (reste du monde) images par seconde. Chaque image de télévision est constituée de deux demi images appelées trames composées par un certain nombre de lignes elles même constituées par un certain nombre de points. Le nombre de points par ligne multipliés par le nombre de lignes donne le nombre de points ou pixels d'une trame. Ce nombre total de pixels donne la définition de la trame. C'est le paramètre essentiel de la qualité finale des images vidéo qui sont donc constituées de l'entrelacement des lignes de deux trames consécutives (paire et impaire).

Les magnétoscopes standards (VHS, S-VHS) ont un arrêt sur image. Certains magnétoscopes permettent un arrêt sur trame (U-Matic, BVU, DVC-PRO) ce qui donne une résolution temporelle de 20 millisecondes (50 trames en PAL) pour décomposer le mouvement. Pour l'étude des mouvements rapides il existe également des caméras vidéo à grande vitesse (plusieurs milliers d'images par seconde) (Micquel-1985).

La définition des images est limitée d'abord par le nombre de pixels du capteur d'image de la caméra et ensuite par la bande passante des magnétoscopes. Ces deux dispositifs ont connu durant la dernière décennie des progrès considérables. Tout d'abord, les capteurs d'images de télévision, les tubes « vidicon », ont été remplacés par des matrices de pixels photo sensibles à semi conducteurs appelés CCD (Charge Coupled Devices) qui sont beaucoup mieux adaptées à la décomposition du mouvement (Howels-1986).

En effet, les CCD ont, parmi d'autres avantages, une rémanence quasi nulle (bien plus faible que les vidicons) ce qui permet de ne pas éblouir le capteur), ainsi qu'une plus grande sensibilité associée à la possibilité d'utiliser un obturateur électronique (shutter), qui permet une bien meilleure netteté des images, pour l'étude des mouvements rapides. Les magnétoscopes ont également évolué ces dernières années grâce surtout à l'utilisation de techniques numériques qui sont en train de remplacer progressivement les techniques analogiques. Ils sont également de plus en plus intégrés aux caméras sous la forme de caméscopes qui ont maintenant, pour des prix accessibles, une qualité professionnelle.

3-Les mouvements d'accompagnement de la parole

Nous les avons défini précédemment comme des mouvements qui ne participent pas directement à la production de la parole. Ils intéressent plus particulièrement les mimiques de la face, les mouvements de la tête et des segments des membres supérieurs dans les situations de dialogue et même de l'ensemble du corps dans des situations de scènes (théâtre, art lyrique, conférences).

Les mouvements d'accompagnement sont réalisés pour être vus et fournissent une information complémentaire à la couche verbale. Leur dynamique suit les variations prosodiques du discours (accent, rythme, intonation). Son évolution se situe donc au niveau de la phrase et ne nécessite pas pour leur étude une grande résolution temporelle. Pour ces deux dernières raisons, la vidéocinématographie est l'outil de choix pour les études des mouvements d'accompagnement de la parole.

Selon le type d'études envisagées, on peut définir deux méthodes d'investigations vidéocinématographiques.

3-1-La vidéocinématographie qualitative

C'est la méthode la plus simple à mettre en œuvre. On peut l'utiliser s'il n'est pas nécessaire de connaître avec précision les amplitudes et les directions dans l'espace des mouvements mais uniquement leur enveloppe de variation associée à leur développement temporel. La résolution temporelle de 25 images par seconde est considérée comme suffisante et l'évaluation de l'amplitude des mouvements ne nécessite pas une grande définition des images. Cependant, malgré cette simplicité apparente, il est toujours recommandé de s'entourer de précautions élémentaires pour l'enregistrement des mouvements. Tout d'abord, il est nécessaire de prendre garde à avoir bien en vue les parties du corps dont les mouvements nous intéressent, si possible face à la caméra et avec un cadrage qui utilise son plein champ. Ensuite, il est conseillé de filmer toujours avec le maximum de lumière afin d'avoir de la profondeur de champ et de pouvoir travailler avec une exposition courte (shutter) pour avoir un maximum de netteté des images. Ceci même au détriment d'une bonne définition des couleurs dont nous n'avons rien à faire pour l'étude du mouvement. Enfin, pour aider le suivi de points particuliers on peut utiliser des repères visuels de référence et des maquillages variés.

Il est utile de préciser que pour l'étude de mouvements dans l'espace, (des segments supérieurs ou de la tête par exemple) il est nécessaire d'utiliser deux caméras car une seule ne permet d'apprécier correctement que les mouvements situés dans son plan de vue et cela même pour des études qualitatives. La finalité de nos études de mouvements étant de les mettre en relation avec les sons de parole, leur enregistrement est réalisé simultanément à celui des images sur une piste son du magnétoscope.

La plus grande partie des études sur les mouvements d'accompagnement de la parole sont réalisées qualitativement et au moyen d'une unique caméra ou camescope car c'est de loin la méthode la plus simple et la moins onéreuse donc la plus accessible. Il en est tout autrement des études de mouvement quantitatives

3-2-La vidéocinématographie quantitative

3-2-1 Principe général

On entend sous ce qualificatif des études menées à partir de mesures géométriques de longueur de segments, angles de rotation, translation et élévation entre des points caractéristiques du mouvement de différentes parties anatomiques du corps. Ces points doivent être très contrastés par rapport à l'absorption lumineuse générale pour être bien visibles. Dans ses «Chronophotographies géométriques» Marey utilise des points blancs sur un fond noir. Ces points peuvent être marqués par maquillage ou par des pastilles ou des sphères de diverses tailles fixées correctement aux parties anatomiques dont on veut étudier le mouvement.

La mesure précise des distances entre les marqueurs ne peut être réalisée qu'à la condition qu'ils se trouvent et se déplacent dans un plan orthogonal à l'axe de la caméra (cad de face). La mesure de rotation n'est également fiable que si son axe se situe également dans ce plan.

Pour des mesures dans l'espace il est donc nécessaire d'utiliser deux caméras dont les axes seront orthogonaux. Les mesures sont alors effectuées directement sur les images enregistrées sur les supports propres à chaque technique : manuellement, sur les photographies de Marey, et en arrêt sur l'image sur des écrans de projection de cinéma puis sur écran de télévision associé à un magnétoscope enfin, en procédures automatiques sur ordinateur après numérisation des images vidéo.

Seules sont utilisées de nos jours les techniques vidéo. Les mesures automatisées sur ordinateur ne sont accessibles que depuis peu d'années. Elles représentent une amélioration considérable par le fait qu'elles suppriment toute intervention manuelle dans les mesures, leur transcription et leur traitement. Quelle que soit la technique utilisée, il est nécessaire pour disposer de mesures fiables, de procéder à des calibrages spatiaux des enregistrements.

3-2-2-Les procédures de calibrage

Elles sont toujours identiques à celles qui ont été imaginées et mises en œuvre pour la première fois par Marey. Elles consistent à placer dans le champs des images et dans le plan des marqueurs, une règle ou un volume dont les cotes géométriques servent de référence de distance. A partir de ces références, dont on connaît les dimensions exactes, on peut reporter une échelle sur la surface de projection des images et ainsi procéder à des mesures vraies en centimètres par exemple. Dans les procédures automatiques sur ordinateurs après enregistrement vidéo, la longueur de la règle de référence est convertie en nombre de pixels sur les axes vertical et horizontal. Ainsi, en comptant les pixels entre deux points on mesure précisément leur distance. Pour cela, il est nécessaire d'avoir une bonne définition des images (grand nombre de pixels).

3-3-Les problèmes de synchronisations son-image et de caméras multiples

Nous avons déjà mentionné l'obligation d'enregistrer simultanément les images du mouvement et le signal acoustique de la parole. Cela ne pose pas de problèmes particuliers lorsque l'on utilise une seule caméra ou caméscope moderne. Les choses se compliquent très vite dès que l'on va utiliser plusieurs caméras et plusieurs sources sonores dans le cas d'études de mouvement dans le dialogue entre plusieurs locuteurs par exemple ou de plusieurs parties anatomiques d'un même locuteur.

Dans ce cas, il est nécessaire de synchroniser toutes les caméras entre elles. On utilise pour cela des signaux de « genlock » qui permettent de synchroniser les fréquences image de chaque caméra à partir d'une horloge unique. Cette horloge permet également de numérotter les images de chaque caméra. Pour permettre un repérage aisé il est conseillé de segmenter les enregistrements en séquences de quelques dizaines de secondes au moyen de signaux de clap qui remettent le comptage des images à zéro. Le numéro du clap est associé au numéro d'image. Ainsi, toutes les images de chaque caméra peuvent être très précisément repérées. Les signaux d'horloge et de clap peuvent être enregistrés simultanément aux signaux de parole ce qui permet une synchronisation parfaite du son et des images si l'on utilise des moyens d'enregistrements multiples et différents. Cela peut être le cas pour des études de gestes articulatoires

Ces systèmes de synchronisation ne sont pas standards par le fait qu'ils n'existent pas dans le commerce, si ce n'est dans la classe « broadcast » inaccessible aux moyens habituels de la recherche, et sont « bricolés » pour les besoins expérimentaux (Simon et al-1984, Teston-1991). De plus, ils nécessitent des matériels peu classiques tels que des caméras synchronisables.

4-4-Les systèmes de vidéométrie intégrés pour l'étude des mouvements

Nous avons juste mentionné leur existence précédemment et il est nécessaire de développer leur description car ils commencent à se diffuser lentement grâce à leurs performances et leur facilité d'utilisation qui s'améliore tandis que leur prix de revient diminue. Seuls ces systèmes peuvent permettre un futur développement des mesures quantitatives.

Ils ont été conçus à l'origine pour des études sur la motricité en général (bio-mécanique, ergonomie, gestes sportifs etc..) et des utilisations cliniques telles que l'aide au diagnostic et la rééducation fonctionnelle motrice. Leur existence est relativement récente (une dizaine d'années) et leur diffusion a suivi le développement de la micro informatique.

Leur principe est basé comme nous l'avons mentionné précédemment sur le suivi de marqueurs de référence disposés sur les parties anatomiques à étudier.

Il existe deux types de systèmes vidéométriques: D'une part des systèmes complets comprenant une ou plusieurs caméras connectées à un ordinateur. Ces caméras sont équipées de stroboscopes à émission infrarouge et sont sensibles à ce rayonnement. Les marqueurs sont des réflecteurs infrarouges. La définition temporelle peut être très supérieure à la vidéo standard car elle n'est pas assujettie à un enregistrement sur magnétoscope. L'émission sous la forme de flashes du rayonnement infrarouge d'éclairement du sujet est synchronisée sur l'ouverture du « shutter » de la caméra. Cette technique permet d'avoir un très bon contraste et une bonne précision dans la détection des marqueurs pour analyser des phénomènes rapides avec de grandes vitesses de prise d'image. Les CCD des caméras étant à haute définition et leurs optiques particulièrement linéaires, ces systèmes sont parfaitement homogènes pour réaliser des enregistrements précis. Ce sont de véritables instruments de mesure.

Il est à noter que les images vidéo de ces systèmes ne sont pas utilisables pour étudier une scène car n'y apparaissent que les marqueurs. Pour nos applications il est nécessaire de doubler la prise d'image par des caméscopes standards

Sans faire une énumération exhaustive de ces systèmes de vidéométrie nous mentionnons le système « ELITE » d'origine italienne et le système « VICON » d'origine anglaise qui ont des caractéristiques proches bien que ce dernier soit actuellement d'une utilisation beaucoup plus aisée. Il peut, dans sa version la plus complète, gérer 7 caméras et filmer jusqu'à 240 images par seconde. Ces systèmes peuvent également enregistrer simultanément des signaux électrophysiologiques mais dont les largeurs de bandes en fréquence sont très inférieures à la parole. Ceci représente pour nos applications une contrainte qui ne peut être surmontée que par des « bricolages de laboratoire » nécessitant un staff technique bien sensibilisé à ce type de manipulations.

Enfin il est très important de prendre conscience que ces systèmes ne peuvent donner des informations synthétiques, donc directement accessibles des mouvements, que pour des séquences relativement courtes. La quantité d'information de ces systèmes peut être considérable et n'est restituée que sous la forme de tableaux de coordonnées qui peuvent vite atteindre des tailles gigantesques. Après l'acquisition des mouvements et le calcul des grandeurs définies par le protocole d'expérience, tout le traitement des données reste à faire.

Ces systèmes complets sont toujours très onéreux et nécessitent un accompagnement technique relativement lourd. Il existe cependant des solutions qui consistent à enregistrer avec des moyens vidéo classiques les mouvements et à les traiter sur ordinateur avec des programmes identiques à ceux des systèmes intégrés. De tels programmes ont été souvent développés dans des laboratoires pour leurs propres besoins. On en trouve dans le commerce, parfois diffusés par les mêmes sociétés conceptrices de systèmes de vidéométrie.

3-5-Les programmes de mesures vidéométriques

Dans le paragraphe 4-2 nous avons mentionné les procédures de mesure des mouvements à partir d'enregistrements. Elles sont basées sur des mesures géométriques entre des marqueurs qui doivent se différencier facilement des autres parties de l'image par un bon contraste. Une procédure de détection automatique des marqueurs doit pouvoir être effectuée avec le moins d'erreurs possibles. Pour cela, les systèmes intégrés de vidéométrie, utilisent des marqueurs infrarouges.

Dans une image vidéo standard on peut obtenir un bon contraste avec des marqueurs blancs. Le programme va donc détecter ces points après que l'utilisateur leur ait donné un indice de reconnaissance. A partir de cela, la machine va faire tous les calculs de mesure, selon un protocole prédéfini par l'utilisateur, entre les différents marqueurs dont elle connaît les coordonnées pour chaque image. Outre ces mesures « brutes », les programmes de mesure vidéométriques permettent une grande variété de traitements plus ou moins complexes qui vont de l'élémentaire suivi de points à la reconstruction en trois dimensions (3D).

Pour utiliser de tels programmes, il est nécessaire auparavant d'acquérir les images vidéo sur l'ordinateur. C'est le rôle de la carte d'acquisition vidéo. Les fabricants de programme de vidéométrie proposent tous de telles cartes mais il faut prendre garde à notre contrainte d'acquisition simultanée du signal de parole. Certes, tous les micro-ordinateurs sont équipés de cartes son mais il est vraisemblable qu'elles ne fonctionneront pas correctement en acquisition simultanée avec une carte d'acquisition vidéo. Il est préférable d'utiliser des cartes multimédia capables de gérer l'acquisition du signal vidéo et de deux signaux acoustiques (ex MATROX RAINBOW ou MIRO DC30).

Là encore en dehors de toute exhaustivité, nous mentionnons le programme STATIS surtout utilisé en posturographie. D'origine française, il n'est pas très puissant mais il est simple et peu onéreux. Le programme VISION CONCEPT est beaucoup plus puissant, et très modulaire. Du même constructeur que le système VICON (Oxford Metrics) il nous paraît être le mieux adapté à nos préoccupations.

4- Les gestes articulatoires

La production de la parole est certainement un des actes neuromoteurs les plus complexe du monde vivant connu. En effet, elle met en jeu un très grand nombre de muscles, caractérisés par de très nombreuses unités motrices et dont la synchronisation doit être parfaitement contrôlée pour créer l'objet sonore linguistique.

La production de la parole est un système dynamique, dont le comportement à un moment donné dépend de ses états antérieurs. Le système est donc dépendant d'une variable paramétrable fonction du temps qui dans ce cas est un geste articulatoire. Notre préoccupation est de savoir comment évoluent ces gestes dans le temps lorsque les paramètres de contrôle varient pour aboutir au son porteur de sens.

4-1- La phonologie articulatoire

La phonologie articulatoire est dérivée du modèle de la dynamique des tâches dont le but principal est l'étude des mouvements, de leurs compensations et de la chronologie de ces compensations. Dans la production de la parole, les mouvements sont caractérisés par des gestes discrets et abstraits. Les gestes sont définis en termes de constriction dans le conduit vocal et sont, pour la dynamique des tâches, des mouvements ayant une trajectoire vers une cible spatiale. Pour générer ces trajectoires, il faut connaître pour chaque variable pertinente l'état instantané du système, la position de la nouvelle cible et les valeurs d'amortissement du système (Demolin-1998).

Le modèle de phonologie articulatoire proposé par Browman et Goldstein (1989) est le seul qui introduit explicitement un aspect dynamique en phonologie. Dans ce modèle, les gestes qui sont des unités d'action, sont également des unités de base de contraste dans les items lexicaux, les atomes de la description, les unités de base de la phonologie. Ils sont des caractérisations abstraites d'événement articulatoires ayant chacun une durée intrinsèque.

Les gestes ne correspondent pas à des traits ou à des segments phonologiques. Si les segments sont définis par des matrices de traits qui leur sont propres, pour Browman et Goldstein, les segments sont définis par des constellations de gestes. Ainsi, la phonologie peut être considérée comme un ensemble de relations parmi les gestes, événements physiques réels, qui caractérisent les systèmes et les modèles de production.

L'introduction du geste comme élément de base des systèmes phonologiques a introduit l'aspect dynamique en phonologie par sa nature spatio temporelle. Le modèle offre des solutions élégantes et naturelles à des problèmes qui se résolvaient auparavant par la formulation de règles « ad hoc » (Demolin-1998). Dans cette perspective, il est donc fondamental de pouvoir observer, enregistrer et mesurer les gestes articulatoires.

4-2-Les méthodes d'observation des gestes articulatoires

Les moyens d'observation des gestes sont très divers dans leurs principes et leurs techniques et diffèrent en fonction de la nature des articulateurs.

On peut observer les gestes à trois niveaux distincts.

D'abord, au niveau du contrôle neuromoteur des muscles qui actionnent les organes articulatoires principalement par des méthodes d'analyse électromyographiques qui donnent des informations sur l'activité musculaire.

Ensuite, au niveau du mouvement proprement dit de ces organes articulatoires (lèvres, langue, mandibule, voile, larynx) observés directement au moyen d'images vidéocinématographiques ou indirectement au moyen de capteurs de mouvement.

Enfin, au niveau des phénomènes que ces mouvements induisent; d'abord les paramètres aérodynamiques (débits d'air oral et nasal et pressions intra orale et sous glottique), modulés par les constriction du conduit vocal , ensuite le signal acoustique qui est l'ultime conséquence des gestes articulatoires.

4-3-Les méthodes d'observation électromyographiques

Les techniques d'électromyographie (EMG) consistent à capter les potentiels d'action des fibres musculaires qui constituent le signal électro physiologique EMG lorsqu'elles se contractent, et fournissent un travail. Son évolution dans le temps permet de suivre le travail musculaire. Ces techniques à priori séduisantes sont utilisées de routine en clinique pour les affections neuromotrices mais ne sont pas accessibles aux chercheurs non médecins. De plus elles sont difficiles à maîtriser tant au plan technique de la capture des potentiels EMG que de leur interprétation. Il existe deux méthodes de capture des potentiels EMG. D'une part au moyen d'électrodes de surface qui sont sensibles à l'activité globale du muscle le plus proche ou le plus énergétique situé à proximité sous la peau. D'autre part au moyen d'électrodes implantées directement dans le muscle dont on veut mesurer l'activité. Cette dernière méthode beaucoup plus sélective est plus délicate à mettre en œuvre. Bien que l'amplitude des signaux EMG (valeur efficace) soit proportionnelle en première approximation à l'énergie développée par le muscle, cette relation n'est pas linéaire, elle dépend trop de l'état des électrodes. On peut donc difficilement interpréter les variations du signal EMG d'une manière autre que qualitative et surtout en fonction toute relative avec l'amplitude d'un geste. Cependant, les variations de potentiels EMG peuvent donner des informations très riches sur la chronologie des événements musculaires et sur les relations agonistes-antagonistes des acteurs gestuels. Même s'ils sont la manifestation de l'origine de tous gestes, les potentiels EMG ne peuvent constituer une méthode d'étude des gestes articulatoires en soit. Ils ne doivent être considérés que comme un complément aux méthodes traditionnelles d'observation des mouvements.

4-4 Les méthodes d'observation des mouvements des organes articulateurs

4-4-1 Les méthodes vidéocinématographiques

Ce sont les mêmes méthodes d'observation d'images cinématographiques en vidéo que celles décrites précédemment pour l'étude des mouvements d'accompagnement de la parole. Elles sont utilisées pour étudier les mouvements des organes articulatoires visibles sur la face des sujets ou cachés dans le conduit vocal.

En vidéo traditionnelle on enregistre les mouvements des lèvres de face et de profil. On peut en mesurer précisément l'aperture avec des techniques particulières (Lallouache- 1990). Il est possible également d'enregistrer les mouvements du voile ou des parois pharyngales au moyen d'un endoscope souple par voie nasale (Autesserre et al-1989).

La vidéocinéradiographie a remplacé le radio cinéma pour l'étude de profil des mouvements des organes articulateurs sous rayons X. Pour des raisons juridiques et d'éthique qu'il est illusoire d'espérer remettre en cause à l'avenir, il n'est plus possible maintenant d'utiliser cette technique d'imagerie basée sur l'utilisation de radiations ionisantes. Ceci est bien dommage car cette technique est une méthode de choix pour l'observation et la mesure des gestes articulatoires. Il faut donc utiliser d'autres méthodes et exploiter au maximum les anciens enregistrements.

Les problèmes d'enregistrement, de synchronisation et de mesures vidéométriques sont les mêmes que pour les mouvements d'accompagnement, cependant pour l'observation des gestes articulatoires il est nécessaire de filmer les images avec la meilleure résolution temporelle possible (50 trames par seconde). Nous mentionnons dans cette rubrique la technique qui est appelée à remplacer en imagerie la vidéocinéradiographie. Il s'agit de l'imagerie par résonance magnétique nucléaire plus communément appelée IRM ou scanner RMN. Ces techniques d'investigations médicales « radiologiques » permettent des mesures d'une très grande précision sur l'anatomie du conduit vocal non seulement sur le plan sagittal médian, mais également frontal et transversal ce qui permet de définir pour la première fois les volumes précis des différentes cavités du conduit. Le problème essentiel de cette technique non invasive pour les sujets se situe pour l'instant dans sa faible résolution temporelle qui est au maximum d'une dizaine d'images par seconde associée à la grande difficulté de disposer de telles machines car elles sont rares et occupées à plein temps pour leurs missions hospitalières. Cependant, cette technique est en train d'évoluer rapidement et déjà elle donne de remarquables informations (Demolin-1997).

4-4-2 Les méthodes kynésiographiques

Ces méthodes consistent à transcrire sur un graphique les variations d'un mouvement après en avoir appréhendé la manifestation au moyen d'un capteur. Pour répondre aux mêmes préoccupations qui sont les nôtres, Rousselot (1897) a développé plusieurs capteurs kynésiographiques mécaniques qui lui ont permis de décrire de nombreux phénomènes. Depuis, on en a imaginé beaucoup mais seules, l'électropalatographie et la magnéto-métrie sont d'une utilisation fréquente.

4-4-2-1 L'électropalatographie

L'électropalatographie ou EPG est une technique très particulière qui permet de visualiser les contacts de la langue sur le palais en phonation. Elle donne donc la topographie des lieux d'articulation lingo-palatins. On doit à Hardcastle (1989) un rôle tout à fait majeur dans le développement de cette technique. Son principe consiste à disposer des électrodes sur un palais artificiel réalisé pour chaque sujet après une prise d'empreinte. Un dispositif électronique permet de détecter le contact de la muqueuse linguale avec les électrodes et un programme informatique permet de visualiser en temps réel les contacts palatins et de faire des traitements tels que suivis de contacts, symétrie, antériorité, postériorité, centre de gravité, etc. De mise en œuvre aisée, L'EPG est une très bonne méthode pour l'étude des gestes linguaux bien qu'en dehors de tout contact, elle ne donne aucune information sur la trajectoire de la langue. Il existe deux systèmes d'EPG, le premier proposé par la société Kay Elemetric est surtout diffusé aux USA et le second de la société Milgrants qui utilise le système de Hardcastle est plus simple et plus efficace bien que moins onéreux. L'inconvénient majeur qui empêche une plus grande diffusion de cette technique se situe dans la nécessité de réaliser un palais artificiel. La fabrication de ces dispositifs avec des techniques de prothèses odontologiques coûte cher. De plus, dès qu'un problème dentaire apparaît (plombage, dents de sagesse, croissance des enfants etc...) il est nécessaire de les refaire. Ceci explique sa faible diffusion pour la rééducation articulatoire chez l'enfant malgré l'intérêt de la méthode.

4-4-2-2 Les méthodes magnétoométriques

La magnétoétrie est une technique de mesure de distance au moyen de champs magnétiques. Son principe est utilisé dans de nombreux domaines industriels et scientifiques. Les premières tentatives peu convaincantes d'utilisation dans notre spécialité datent de plus de vingt ans. Cependant, la prohibition des investigations d'imagerie sous rayons X lui a donné une impulsion nouvelle depuis une décennie en Europe et aux USA si bien que deux systèmes de magnétoétrie sont proposés à nos recherches; le Movetrack de Botronic d'origine suédoise et l'Articulograph de Carstens d'origine allemande. Le principe de base de ces machines repose sur la mesure des distances entre des bobines qui créent un champ magnétique et un petit capteur que l'on fixe sur l'organe en mouvement. Les bobines fixées sur un casque solidaire du crâne sont placées orthogonalement au plan sagittal de la tête du sujet ainsi que les capteurs. Ces derniers permettent de mesurer les mouvements des lèvres, de la mandibule, de la langue et même du voile. Les deux dispositifs sont assez différents. Le Movetrack dispose de deux bobines son casque est très léger, bouge peu et ne gêne pas le sujet. Il est cependant sensible aux mouvements parasites des capteurs (tilts). L'Articulograph est beaucoup plus lourd et a des problèmes de positionnement. Equipé de trois bobines qui définissent mieux le plan sagittal il est moins sensible aux tilts. Dans les meilleures conditions ces appareils ont une précision de mesure des mouvements l'ordre du millimètre et peuvent donner également des indications cinétiques telles que l'accélération. On peut utiliser autant de capteurs que nécessaire. Dans la pratique on dépasse rarement cinq capteurs. Ces derniers sont de petites bobines (2 mm) qui doivent être solidarisées avec l'organe mouvant par collage. Si c'est relativement simple sur les lèvres, et les dents, il n'en est pas de même sur la langue. Les sujets supportent mal les fils des capteurs en bouche et surtout, sur la langue on n'est jamais sûr de la bonne orthogonalité des capteurs par rapport au plan sagittal. L'utilisation de ces systèmes réclame une bonne pratique et un support technique compétent (FIPKM-1993) ce qui explique que seuls quelques centres maîtrisent ce type de mesure. A notre avis, cette technique donne de bons résultats quantitatifs sur certains organes mais manque de certitude sur d'autres. Elle peut donner en tout cas de très riches informations sur la chronologie des événements articulatoires. Il est certain que de nombreux chercheurs ont trop attendu de cette méthode et un certain désintérêt paraît insensiblement gagner le microcosme. Pourtant elle peut, à notre avis, être encore bien améliorée.

4-5 Les méthodes aérodynamiques

Le domaine de l'aérodynamique dans la production de la parole est fondamental par le fait qu'il est à l'origine de toutes ses manifestations sonores. C'est en effet la colonne d'air pulmonaire qui, modulée par les différentes constriction du conduit vocal est la source du signal vocal. Les paramètres aérodynamiques sont constitués par les débits à la bouche (oral) et aux narines (nasal) et les pressions intraorale et sous glottique. La connaissance des variations de ces paramètres en fonction des segments phonémiques prononcés donne des informations très riches sur les mouvements des organes articulatoires du conduit vocal (Warren-1996). On doit là encore à Rousselot (1997) d'avoir codifié les principaux paramètres aérodynamiques. Il est cependant illusoire d'attendre une relation linéaire entre les variations de débits et de pression et les gestes articulatoires. Cependant, ils permettent une remarquable description chronologique des gestes articulatoires et sont indispensables dans l'étude des phénomènes consonantiques complexes et de la coarticulation. (Demolin-1996a et b). Ces méthodes, dont la pratique n'est pas très compliquée n'ont pas l'importance qu'elles méritent à notre avis. Les systèmes de mesures aérodynamiques ne sont pas très nombreux. On peut citer le masque de Rothemberg très diffusé aux USA et l'Aérophone de Frokjer Jensen commercialisé par Kay Elemetric. Ces systèmes sont entachés de plusieurs défauts. Seul le système EVA proposé par la société SQLab permet des mesures nasales et une faible gêne articulatoire (Teston-1994).

4-6 Les méthodes acoustiques

Le signal de la parole est la résultante acoustique d'une certaine forme du conduit vocal elle même fonction des différentes constriction provoquées par les gestes des organes articulateurs. Les relations entre les gestes et le signal peuvent être formalisées au moyen de modèles articulatoires qui, à partir de certaines positions des organes produisent un son correspondant. C'est la synthèse vocale articulatoire. Il existe ainsi plusieurs modèles de production. Depuis quelques années on a tenté de réaliser la méthode inverse, c'est à dire de reconstituer les gestes articulatoires à partir du signal acoustique en utilisant les modèles de production « à l'envers ». C'est ce que l'on appelle l'inversion des modèles. A partir du signal acoustique, on tente de reconstituer les lieux et les grandeurs des constriction du conduit vocal. Pour cela il est nécessaire de disposer de modèles robustes peu sensibles aux variations inter locuteurs, ainsi que des données fiables sur un corpus de mouvements associés aux signaux acoustiques correspondant, pour valider l'inversion du modèle (Demolin et Soquet-1996-c). On peut se demander pourquoi chercher à observer des gestes à partir du signal lorsque l'on dispose de toutes les méthodes que nous venons de présenter. En fait il existe beaucoup de langues pour lesquelles il est difficile d'obtenir des informations autres qu'acoustiques et dont on serait condamné à ignorer les systèmes articulatoires. Il ne faut pas attendre de ces méthodes des résultats extraordinaires car les modèles de production sont trop schématiques, et les données de validations trop parcellaires ou peu précises. Mais, les bases de données articulatoires s'enrichissent progressivement ce qui permet d'affiner les modèles et ainsi améliorer l'inversion. Cela devrait finalement permettre de se passer, de toutes les techniques d'observation des mouvements des organes articulateurs dans un avenir plus ou (surtout) moins proche.

4-7 Les systèmes d'exploitation de données articulatoires intégrés

Nous venons de décrire une grande variété de systèmes d'observations directes ou indirectes des mouvements des organes articulateurs. De principes et de techniques différentes ils sont souvent très difficiles à faire fonctionner ensemble et surtout en parfaite synchronisation pour que leurs informations complémentaires puissent être efficacement utilisées et exploitées.

Les gestes articulatoires sont rapides, nombreux et complexes. Il est important de multiplier les mesures simultanées sur un même événement au moyen de plusieurs techniques. Ainsi, aux paramètres aérodynamiques sont associées l'EPG, des signaux électrophysiologiques, des images ou encore la magnétométrie. L'acquisition simultanée de tous ces paramètres ainsi que du signal de parole et autres signaux acoustiques (glottographe) doit être parfaitement synchronisée. Seuls quelques rares systèmes de laboratoire peuvent réaliser de tels enregistrements. C'est dans ce but que le système Physiologia a été développé à Aix (Teston-1991) qui de plus permet de synchroniser des images vidéo.

Pour exploiter toutes ces données de manière commode et fiable, il est indispensable également de disposer d'éditeurs de signaux capables d'exploiter des données autres qu'acoustiques. Pour cela, l'éditeur Phonedit est le complément du système Physiologia.

Un autre problème apparaît très vite avec la multiplication des paramètres enregistrés. C'est la quantité considérable de données qui s'accumulent et encombrant vite les ordinateurs dans l'attente de leurs traitements. Nous conseillons de ne pas multiplier à l'infini le nombre de paramètres sous prétexte que « plus on en a meilleurs seront les résultats ». On doit bien établir le protocole expérimental en fonction d'un objectif scientifique très clairement défini sous peine de risquer d'être très vite submergé par une masse de données inexploitable.

Conclusion

Il peut apparaître au terme de cet exposé qu'il existe deux grandes catégories d'études sur les mouvements dans la parole. La première, sur les mouvements d'accompagnement et la seconde sur les gestes articulatoires. La première, peut paraître moins complexe au plan méthodologique, mais cela est trompeur, car en fait il n'y a pas tellement de différences dans les principes, méthodes et précautions

expérimentales mis en œuvre dans ces études. Ils ne se distinguent principalement que dans la précision des mesures (d'espace et le temps). Il n'a jamais été simple ou facile d'observer, enregistrer et interpréter des mouvements. En revanche, une différence existe dans les moyens d'investigation. L'apparition des caméscopes modernes avec un standard unique, permet depuis peu de temps un développement de ce type d'études avec un investissement modique. Mais, même pour un matériel (qui semble modeste) il est toujours nécessaire de travailler avec méthode. Cet avantage disparaît totalement pour les études quantitatives des mouvements d'accompagnement qui nécessitent des investissements aussi importants que pour celles des gestes articulatoires et dont les quantités d'information à traiter sont du même ordre.

La différence principale qui existe entre ces deux catégories d'études réside dans la difficulté inhérente aux études des gestes articulatoires dont la plupart des méthodes se situent dans le cadre juridique régissant les expérimentations humaines. Les enregistrements doivent alors être réalisés sous contrôle médical dans un lieu habilité (centre hospitalier) après accord d'un comité d'éthique (loi Huriet). Ceci est une contrainte à laquelle les chercheurs ont quelques difficultés à s'habituer, mais qui est une sécurité pour les sujets autant que pour eux même. Ce type de recherches impose une collaboration obligatoire avec des médecins et renforce les contacts déjà incontournables pour l'accès aux techniques d'imagerie et d'électrophysiologie. Pour cela, et d'autres raisons développées précédemment, les investigations sur les gestes articulatoires sont réservées à quelques centres de recherche qui disposent ou ont accès à tous les moyens nécessaires. Ces centres sont généralement ouverts aux chercheurs extérieurs pour qu'ils puissent y effectuer leurs enregistrements de données avant de les exploiter dans leurs laboratoires d'origine. De plus, on voit apparaître des bases de données certes parcellaires et de tailles modestes mais qui s'enrichissent de jour en jour.

Les études sur les mouvements en général n'ont jamais été faciles à mener. Elles ont toujours été tributaires de techniques spécifiques qui ont souvent fait évoluer la technologie. Ces dernières années ont connu des améliorations considérables. Le vidéo cinéma d'abord en est la plus importante, associée maintenant à la vidéo numérique (celui qui a connu le cinéma argentique appréciera mieux que quiconque). La vidéométrie quantitative permet d'éviter les travaux de «bénédictins» qui étaient associés aux dépouillements et aux mesures manuelles. Il en est de même pour les systèmes intégrés d'études articulatoires qui ainsi permettent de multiplier les travaux pour améliorer les modèles de production.

Il y a vingt ans, les chercheurs n'imaginaient même pas pouvoir un jour disposer de telles techniques qui restaient pour eux du domaine du rêve. Les études sur les mouvements qui sont fondamentales et passionnantes ont donc les moyens de se développer. Alors, profitons en !

Bibliographie

- Autesserre D., Galindo B., Teston B. et Vigouroux N., (1989), « Movement of the lips and velum in speech: variations in aerodynamic parameters », *2nd Eurospeech-ESCA*, Paris September 1989, 437-440.
- Browman, & C., Goldstein, L., (1989), « Articulatory gestures as phonological units », *Phonology*, vol. 6, 201-251.
- Demolin, D. & Teston, B., (1996 a), « Aspects articulatoires et aérodynamiques des occlusives labio vélaires », *XXIèmes Journées d'Etude sur la Parole de la Société Française d'Acoustique*, Avignon, Juin 1996, 111-114.
- Demolin, D. & Teston, B., (1996 b), « Labiodental flaps in Mangbetu », *Journal of the international Phonetic Association*, vol. 26, n2, 103-111.

- Demolin, D. & Soquet, A., (1996 c), « Speech production models as experimentation tools in phonology », *5th Laboratory Phonology Conference*, Evanston.
- Demolin, D., George, M., Lecuit, V., Metens, T., Raeymaeckers, B., & Soquet, A., (1997), « Coarticulation and articulatory compensations studied by dynamic MRI », *5th Eurospeech*, Rhodes, September 1997, vol. 1, 43-46.
- Demolin, D., (1998), « Dynamique des systèmes phonologiques- Aspects formels et expérimentaux », *XXIIèmes Journées d'Etude sur la Parole de la Société Française d'Acoustique*, Martigny, 15-19 Juin 1998, 53-70.
- FIPQM-31, (1993), *Proceedings of the ACCOR Workshop on Electromagnetic Articulography in Phonetic Research*, Institut für Phonetik und Sprachliche Kommunikation, München, vol. 31, 272 p.
- Gouret, J. P., Paillé, J., Boé, L. J., Descout, R., 1985, Système d'exploitation automatique de labiofilms, *14èmes Journées d'Etude sur la Parole, du Groupement des Acousticiens de Langue Française*, Paris, 61-64
- Hardcastle, W. J., Jones, W., Knight, C., Trudgeon, A., & Calder, G., (1989). « New developments in electropalatography : a state of the art report », *Clinical Linguistic and Phonetics*, vol. 3, 1-38.
- Howells, J.W., (1986), « Most of what you wanted to know about charge coupled devices », Tutorial presented to ITVA, Toronto, Nov. 1986, *PANASONIC Doc.*, 13 p.
- Lallouache, M.T. (1990), « Un poste visage-parole; Acquisition et traitement de contours labiaux », *18èmes Journées d'Etude sur la Parole de la Société Française d'Acoustique*, Montréal, Mai 1990, 282-285.
- Marey, E. J., (1878), *La méthode graphique dans les sciences expérimentales et principalement en physiologie*, Welter, Paris, XXIV, 673 p.
- Marey, E. J., (1882), « Analyse du mécanisme de la locomotion au moyen d'une série d'images photographiques recueillies sur une même plaque et représentant les phases successives du mouvement », *CRAS*, 3 juillet 1882, 14-16.
- Marey, E. J., (1888), « Décomposition des phases d'un mouvement au moyen d'images photographiques recueillies sur une bande de papier sensible qui se déroule », *CRAS*, 29 octobre 1888, 677-678.
- Micquel, J.C., (1985), *L'observation en vidéo rapide*, Lavoisier, Paris. 471 p.
- Rousselot, P. J., (1897-1901), *Principes de Phonétique expérimentale*, Welter, Paris, 638 p.
- Simon, P., Wolf, F., Nadjafizadeh, H., & Boock, G. (1984), « Numérotation automatique et codage synchrone graphique et phonétique pour l'exploitation des films radiologiques », *Travaux de l'Institut de Phonétique de Strasbourg*, n°16, 181-190.

- Teston B., (1991), « Un système d'acquisition d'images simultanées pour l'étude des mouvements des organes articulateurs », *12ème Congrès International des Sciences Phonétiques*, Aix-en-Provence, août 91, vol. 5, 454-457.
- Teston, B., (1993), " A new air flowmeter design for the investigation of speech production", *Proc. Eurospeech 3*, European Speech Communication Association, Berlin, september 93, 405-409.
- Teston B. & Galindo, B., (1990), « Design and Development of a workstation for speech production analysis », *Verba 90*, Alcatel France, Rome, January 90, 400-408.
- Warren, D. W., (1996), « Regulation of speech aerodynamics », in N. J. Lass, *Principles of Experimental Phonetics*, College Hill Press, San Diego, 71-76.