

UN SYSTEME D'ACQUISITION D'IMAGES SIMULTANÉES POUR L'ÉTUDE DES MOUVEMENTS DES ORGANES ARTICULATEURS.

Bernard TESTON

URA 261 CNRS, Institut de Phonétique, Aix en Provence, France.

ABSTRACT

This experimental device, is particularly valuable for obtaining a perfect correlation between articulatory and acoustic events. It allows the simultaneous recordings:

- Labio videofilms: frontal and side view.
- Radioscopic videofilms of the complete vocal tract.
- Endoscopic videofilms of velum, pharynx, larynx, oral cavity etc...

Speech signals and images are synchronized. The various images are recorded simultaneously on video recorders at 50 frames per second.

1. INTRODUCTION

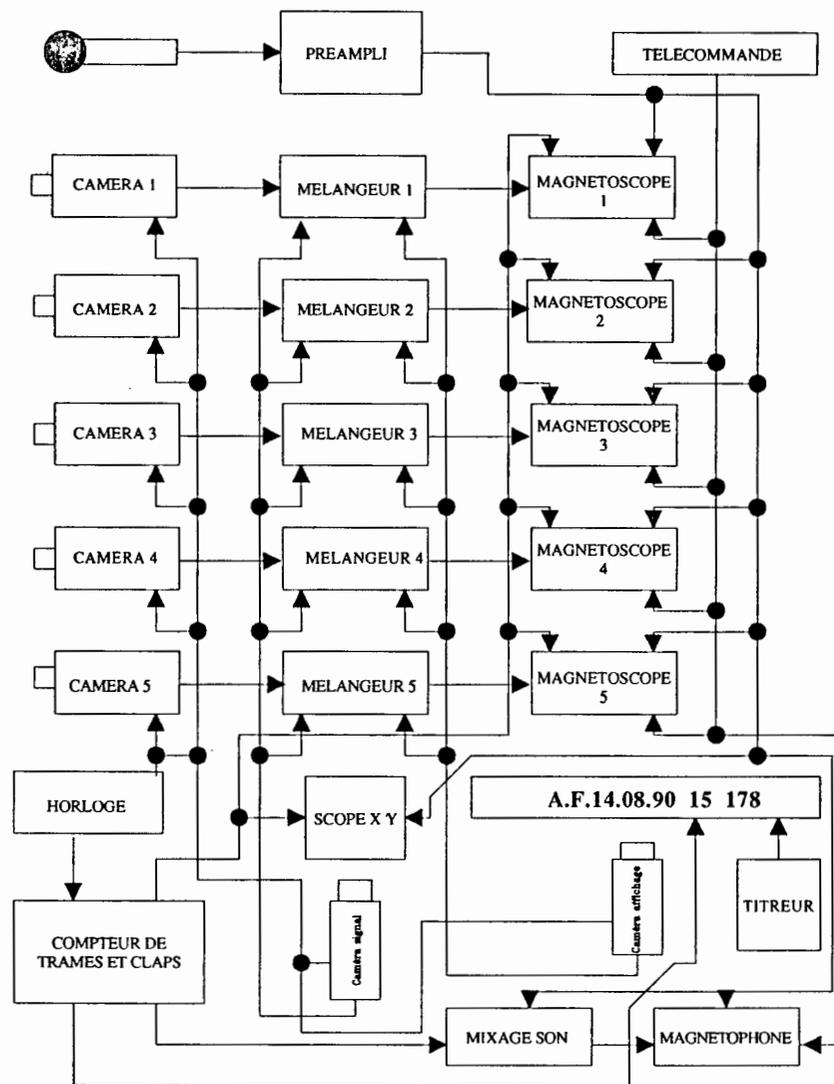
Les études sur les processus d'encodage de la parole qui permettent de passer du niveau linguistique au niveau phonétique, revêtent depuis plusieurs années un intérêt croissant, et en particulier celles qui portent sur les mécanismes articulatoires. Les phénomènes de coproduction et de coarticulation, ainsi que la variabilité intra et inter locuteurs, sont les deux pôles de cet intérêt avec comme objectif, l'amélioration des systèmes de synthèse et reconnaissance vocale.

Les dispositifs d'exploration des mouvements des organes articulatoires, sont nombreux et de principes variés. Une des méthodes les plus directes pour mettre en évidence leurs mouvements, est de décomposer leur cinématique image par image, selon le concept «du récit chronophotographique avec arrêt sur l'image», imaginé par MAREY [1]. En 1986, nous avons réalisé un dispositif d'acquisitions vidéo simultanées sur des images du voile, des lèvres et de la coupe sagittale du conduit vocal [2]. Ce dispositif qui nous donnait satisfaction, a dû évoluer pour les raisons suivantes. Tout d'abord, ce système était constitué par du matériel vidéo standard, et en particulier par des consoles de mixage utilisées au minimum de leurs possibilités et dont nous ne disposons pas en permanence. Nous devions l'emprunter ou le louer auprès de différents fournisseurs, il était de ce fait hétéroclite et onéreux. Enfin, le dispositif de synchronisation manquait de précision, même s'il nous a permis de ne

pas perdre d'images. La nécessité de réaliser de multiples enregistrements sur un grand nombre de locuteurs, dans le cadre d'un contrat de recherche avec la Région Provence- Alpes -Cote d'Azur (PACA), nous a amené à développer le système suivant: Il est possible d'utiliser un maximum de cinq caméras vidéo. Toutes ces caméras sont synchronisées par une horloge générale. Un dispositif particulier permet l'incrustation sur chaque image du signal acoustique et des informations de synchronisation. le système est complété par une télécommande générale de toutes les fonctions d'acquisition, d'un dispositif d'affichage du corpus à prononcer, et d'un appareillage mécanique de positionnement du sujet et des différents capteurs d'information.

2. LA PRISE D'IMAGES

Il est possible d'utiliser des caméras de toutes provenances, à condition qu'elles disposent d'une entrée de synchronisation (genlock). Nous n'utilisons cependant, que des caméras à dispositif de transfert de charges (CCD), particulièrement bien adaptées à la décomposition du mouvement. En effet, les CCD ont, parmi d'autres avantages, une rémanence beaucoup plus faible que les vidicons[3] ce qui permet de ne pas éblouir longtemps le capteur (avantage très important en endoscopie), et surtout la possibilité d'utiliser un obturateur (shutter), qui permet de définir plus précisément l'image dans le temps, pour l'étude des mouvements rapides (lèvres) [4] ; Pour l'étude des mouvements des lèvres de face et de profil, nous utilisons des caméras PANASONIC WVF 250 tri CCD, équipées d'un objectif S12X75 BRM avec un temps d'exposition de 1 ms qui nécessite un éclairage puissant. La caméra pour l'étude des mouvements du voile est une PANASONIC M10 couleur mono CCD, caractérisée par un très faible éclairage minimum (10 lux) qui permet de pallier le manque chronique de lumière des fibroscopes. Ce dernier est un bronchoscope OLYMPUS BF20D, choisi pour sa bonne définition d'image (bien supérieure à celle des nasolaryngoscopes) et son double canal d'éclairage. Il est associé à une source de



SCHEMA DE PRINCIPE DU DISPOSITIF

lumière au xénon OLYMPUS CLV10 utilisée en mode continu (non pulsé). Il est possible d'utiliser des caméras tri CCD pour cette application mais elles sont beaucoup plus encombrantes et plus lourdes. Enfin, la caméra monochrome de vidéoradioscopie est caractérisée par une grande définition (700 points par lignes). Elle fait partie d'une installation d'angiographie GENERAL ELECTRIC (CGR) utilisée pour des interventions de micro-chirurgie intra-vasculaire de longue durée. Elle dispose d'une grande luminosité grâce à un amplificateur de brillance très sensible qui permet de faibles doses de radiations X. Son champ a un diamètre de 34 cm, on peut ainsi y cadrer le conduit vocal dans sa totalité.

3. L'HORLOGE DE SYNCHRONISATION

Toutes les caméras et les dispositifs de synchronisation sont cadencés par une horloge générale. Sa fréquence est générée soit par une référence interne à quartz, soit à partir du signal vidéo d'une caméra. Cette possibilité est obligatoire dans le cas de l'utilisation de la vidéoradioscopie dont l'installation ne dispose pas d'une entrée de synchronisation extérieure (genlock). Compte tenu de sa spécificité (12 sorties de synchronisation) cette horloge a été fabriquée localement.

4. LE SYSTEME DE SYNCHRONISATION

C'est le dispositif central de toute l'expérience. Nous l'avons réalisé spécialement pour faire des enregistrements vidéo parfaitement identifiables en toutes circonstances. Pour cela nous faisons apparaître sur chaque image:

- Le numéro de la phrase dans le corpus (clap).
- Le numéro de l'image (trame paire ou impaire).
- Le signal acoustique correspondant à la durée exacte de la trame de télévision (20 ms) de haut en bas sur la partie gauche de l'écran.
- Les références du locuteur.
- La date de l'enregistrement.

Les phrases sont considérées comme des séquences cinématographiques comprises entre deux claps. Le système de synchronisation peut compter jusqu'à 99 claps. Chaque clap remet à zéro le compteur d'images. C'est le locuteur qui déclenche le clap, qui fait apparaître la phrase à prononcer sur l'écran d'un PC. Il peut également déclencher l'acquisition de signaux issus de différents capteurs (débits, pressions, déplacements, etc...) sur un autre PC disposant du logiciel PHYSIOLOGIA [5]. Le compteur d'images est synchronisé par l'horloge générale. Chaque image est constituée par une trame paire ou impaire du signal vidéo, sa durée est de 20 ms pour le standard européen. Il est possible d'utiliser ce dispositif à la fréquence de 60 trames par seconde si cela s'avère nécessaire. Le système peut compter jusqu'à 999 images entre deux claps. Les impulsions d'images correspondent au retour de

trame (synchronisation verticale du signal vidéo). Elles synchronisent le défilement d'un scope X Y sur lequel est visualisé le signal acoustique. Ce dernier est filmé au moyen d'une caméra monochrome PANASONIC WBL200 équipée d'un objectif LA12B2 synchronisée par l'horloge générale. Les impulsions de claps et d'images sont enregistrées sur une des deux pistes des magnétoscopes. Elles sont reconnaissables par leur différence d'amplitude. Elles peuvent moduler une fréquence de 2 kHz pour les claps et de 7 kHz pour les images pour réaliser des marqueurs sur des analyses en fréquence de type sonagramme [6]. Enfin, les impulsions de claps peuvent être mixées au signal acoustique pour parfaire la synchronisation image-son. Les informations d'identification des locuteurs et la date de l'enregistrement, sont affichées sur un indicateur alphanumérique électroluminescent, ainsi que les numéros des phrases et des images. Cet indicateur est filmé par une seconde caméra monochrome dans les mêmes conditions que précédemment. Toutes ces informations d'identification et de synchronisation sont ensuite incrustées dans chaque image anatomique au moyen d'un mélangeur vidéo.

5. LE MELANGEUR VIDEO

Nous mélangeons à chaque signal vidéo couleur PAL issu des caméras de prises de vues anatomiques, des signaux vidéo monochromes issus des caméras de synchronisation. Tous ces signaux étant en phase au moyen de l'horloge générale, nous pouvons nous contenter de simples mélangeurs sans dégrader les couleurs originales. Nous évitons ainsi d'utiliser des régies de mixage vidéo complexes et coûteuses. Le mélangeur vidéo est constitué par 5 canaux à 3 entrées. Une entrée commune à tous les canaux, contient les informations sur l'enregistrement et les numéros de claps et d'images. Une entrée, sur chaque canal, est réservée à l'image du signal acoustique, qui peut être supprimée pour augmenter la dimension utile de l'écran. La dernière entrée est réservée aux images anatomiques. Les réglages de mélange ne jouent que sur la luminance de chaque entrée, elle peut être ajustée sur une échelle de + ou - 20 dB.

6. LA PRISE DE SON ET SA SYNCHRONISATION

La prise de son est effectuée au moyen d'un microphone à électret AKG C410, associé à un préamplificateur mélangeur. Le signal de parole est enregistré sur une des deux pistes des magnétoscopes; l'impulsion des claps de début des phrases est superposée au signal de parole (pour avoir un bon repérage temporel de ce dernier s'il doit être acquis sur ordinateur en vue d'un traitement acoustique). Le signal de parole peut être également mixé avec les impulsions d'image comme nous l'avons précé-

demment, il est alors enregistré sur un magnétophone bipiste avec le signal original. L'ensemble de l'expérience a un fonctionnement silencieux, excepté la source de lumière OLYMPUS que l'on a du insonoriser pour avoir une bonne dynamique de la prise de son.

7. L'ENREGISTREMENT DES IMAGES

Les images anatomiques contenant les informations d'identification et de synchronisation sont enregistrées sur des magnétoscopes au standard U-MATIC SONY BVU 900 avec le signal acoustique sur la piste 1 et les impulsions de claps et d'images sur la piste 2. Si l'on désire une image synthétique de quatre images anatomiques, il est possible d'en réaliser le montage au moyen d'une régie de découpage FORT A MV 24 GL par exemple et de l'enregistrer sur un cinquième magnétophone. Un dispositif de télécommande générale, permet de déclencher simultanément toutes les fonctions des magnétoscopes et magnétophones.

8. LES EQUIPEMENTS ANNEXES ET LES PROCEDURES DE CALIBRATION

Tous ces équipements sont solidaires d'un ensemble mécanique dont le centre est occupé par le fauteuil sur lequel prend place le locuteur. La tête de ce dernier est maintenue par un céphalostat efficace et peu contraignant. Le fauteuil est contenu dans un parallélépipède dont on peut très précisément ajuster le positionnement. Il peut supporter les caméras, l'endoscope, les dispositifs de calibration, les filtres pour rayons X, le microphone d'enregistrement ainsi que divers capteurs. Cet ensemble mécanique est adapté aux dimensions des installations d'angiographie, il est démontable pour être facilement transporté.

Les calibrations sont effectuées de deux manières différentes soit en positionnant des repères réalisés par maquillage, soit à partir de références dimensionnelles mécaniques. La première procédure s'applique aux images des lèvres de face et de profil. Deux vidéofilms sont réalisés simultanément, l'un en gros plan sur les lèvres, l'autre de toute la face. Dans les deux cas, les lèvres sont maquillées en bleu afin d'assurer une meilleure précision à la mesure d'aperture [7]. Des repères de calibration peuvent être peints sur le visage du locuteur. La méthode qui consiste à placer des repères dans le conduit vocal pour l'étude de sa coupe sagittale s'apparente à cette procédure. Pour cela, des billes de nickel-chrome sont placées par collage sur la langue pour repérer plus facilement ses mouvements [8]. La deuxième procédure consiste à placer une règle de référence dans le plan des lèvres de face et de profil et à la filmer avant chaque séquence de prise d'images.

CONCLUSION

Ce dispositif d'étude des mouvements des organes articulatoires, est l'aboutissement d'une lente évolution, et sa mise au point, le résultat de nombreuses manipulations. De telles expériences, qui nécessitent le contrôle de très nombreux paramètres ne sont jamais simples à mettre en oeuvre. Nous commençons à bien les maîtriser. Les principes fondamentaux de ce système sont maintenant bien établis, et les améliorations actuelles ne portent que sur des points de détail. Il nous reste à dépouiller et à traiter l'énorme quantité d'informations ainsi récoltées. Ceci ne peut se faire qu'au moyen de dispositifs automatiques sur lesquels nous travaillons et qui feront l'objet d'un prochain exposé.

REMERCIEMENTS:

Nous remercions M. CHERIGUENNE, technicien au CNRS, pour l'aide constante qu'il nous a apportée dans la réalisation des matériels.

BIBLIOGRAPHIE

- [1] MAREY, E.J. (1891), "La chronophotographie." Revue générale des Sciences pures et appliquées, vol 16/11 1891, 659-719.
- [2] TESTON, B. et AUTESSERRE, D. (1986), "Description d'un dispositif d'enregistrement simultané des mouvements des organes articulatoires." Actes des 16^e J.E.P. de la Société Française d'Acoustique, AIX, Mai 1986, 65-68.
- [3] 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.
- [4] MIQUEL, J.C. (1985), "L'observation en vidéo rapide." Chap. 3, Lavoisier, Paris. 133-212.
- [5] TESTON, B. et GALINDO, B. (1990), "Design and development of a work station for speech production analysis." VERBA 90, ALCATEL FACE, Jan 90, ROME, 400-408.
- [6] SIMON, P., BOTHOREL, A., WIOLAND, F., et BROCK, G. (1979), "Méthode de synchronisation image-son pour l'étude radiologique des faits de parole." Actes du 9^e Congrès International des Sciences Phonétiques, Copenhague, Vol 1, 213.
- [7] LALLOUACHE, M.T. (1990), "Un poste «visage-parole»: Acquisition et traitement de contours labiaux." Actes des 18^e J.E.P. de la Société Française d'Acoustique, Montréal, Mai 1990, 282-285.
- [8] KIRITANI, S., ITOH, K., and FUJIMURA, O. (1975), "Tongue pellet tracking by a computer controlled X-ray microbeam system." J.A.S.A., vol 57, 1516-1520.