

# Une base de données cinéradiographiques du français.

Alain Arnal<sup>1</sup>, Pierre Badin<sup>1</sup>, Gilbert Brock<sup>2</sup>, Pierre-Yves Connan<sup>2</sup>, Evelyne Florig<sup>2</sup>, Noël Perez<sup>1</sup>, Pascal Perrier<sup>1</sup>, Pela Simon<sup>2</sup>, Rudolph Sock<sup>2,1</sup>, Laurent Varin<sup>1</sup>, Béatrice Vaxelaire<sup>2</sup> & Jean-Pierre Zerling<sup>2</sup>

<sup>1</sup>Institut de la Communication Parlée – UMR CNRS 5009- INPG & Université Stendhal

Avenue Félix Viallet – 38031 Grenoble Cedex 01, France

<sup>2</sup>Institut de Phonétique de Strasbourg– Université Marc Bloch

22 Rue Descartes - 67084 Strasbourg, France.

Tél.: ++33 (0)476 57 48 25 - ++33 (0) 388 41 73 68 - Fax: ++33 (0)476 82 71 91

Mél: [perrier@icp.inpg.fr](mailto:perrier@icp.inpg.fr) – [sock@umb.u-strasbg.fr](mailto:sock@umb.u-strasbg.fr)

## ABSTRACT

This paper presents a preliminary version of a large X-ray database that is currently being elaborated at both the Institut de Phonétique de Strasbourg and the Institut de la Communication Parlée de Grenoble. It currently contains 4 movies that present over 2000 images. These X-ray data focus on different phonetic issues in French: juncture, nasals, and coarticulation in VCV sequences. The database contains 3 kinds of digitized data; the cineradiographic data, acoustic signals and hand-drawn sagittal contours of the vocal tract. All files are phonetically labeled and stored on CD ROMs. Management of the database is developed for Windows NT or Windows 95 with "Microsoft ACCESS", with a version for Macintosh. The data are accessed via requests in SQL language, and a user friendly interface is developed under JAVA, allowing easy formulation of requests, display of selected X-ray images and of vocal tract contours, and also listening the selected video portions.

## 1. INTRODUCTION : CADRE DE CE TRAVAIL

L'étude des mécanismes de contrôle de la production de la parole nécessite le recueil d'un nombre important de données tant physiologiques qu'articulatoires et acoustiques. En ce qui concerne les techniques d'acquisition de données articulatoires, il faut reconnaître qu'elles sont nombreuses et qu'elles ont fait l'objet d'efforts de développement et de diversification très intenses dans la communauté internationale au cours de la dernière décennie. Ces travaux se sont essentiellement orientés autour de la recherche simultanée d'une bonne résolution spatio-temporelle et d'une minimisation des effets négatifs potentiels sur la santé des sujets.

En effet, jusqu'au début des années 70, si on excepte les auto-observations à l'aide de miroirs d'illustres précurseurs comme Hellwag [Hel81], les techniques de bases de l'acquisition de données articulatoires reposaient sur la palatographie et la radiographie. La palatographie présentait l'intérêt d'être simple et peu onéreuse. Mais elle est restée statique jusque dans les années 70, et elle limite le champ d'observation à la zone de contact entre la langue et le palais. La radiographie impliquait le recours à une

matériel médical lourd. Cependant elle offrait une représentation de la coupe sagittale du conduit vocal de la glotte jusqu'aux lèvres [Chi41], [Fan60] et, de plus, elle devint, dès la fin des années cinquante, dynamique avec l'avènement de la cinéradiographie qui permit ainsi pour la première fois l'observation attentive de mouvements des articulateurs de la parole. C'est pourquoi la cinéradiographie a connu un réel succès et a été à la base de travaux de référence sur l'articulation des sons essentiellement en français sous l'impulsion de Georges Straka [Str65] et de ses collaborateurs ([Sim67], [Roc73]), et en anglais à l'initiative de Moll [Mol60] ou Perkell [Per69].

Cependant l'exposition de sujets sains à des rayons ionisants a rapidement posé un problème déontologique, et s'est heurté à la législation de nombreux pays. Par ailleurs, la résolution temporelle de la cinéradiographie ne dépasse pas 50 images/seconde, et ceci est insuffisant pour une étude fine de l'organisation temporelle des consonnes, en particulier des consonnes plosives. Cet ensemble de raisons a incité les chercheurs à s'intéresser à de nouvelles voies d'investigation expérimentale.

Le système Xray Microbeam minimise l'exposition des sujets aux rayons ionisants en concentrant des faisceaux sur des points très localisés du conduit vocal [Wes94]. L'électropalatographie permet des mesures des contacts palataux à 100 ou 200 Hz [Har84] [Fou98]. Les techniques à ultrasons [Sto88] permettent de recueillir des données 3D à des fréquences d'échantillonnage de l'ordre de 100Hz, dans une zone limitée du conduit vocal. L'électromagnétomètre [Sch89] [Per92] mesure les déplacements de 5 ou 10 points du conduit vocal à une fréquence d'échantillonnage pouvant dépasser 1kHz. Toutes ces techniques offrent une bonne résolution spatiale, et des résolutions temporelles supérieures à la cinéradiographie. Elles ont ainsi permis d'approfondir l'étude du contrôle de la parole. Néanmoins elles souffrent d'un même défaut : aucune d'elles ne permet de recueillir en même temps des informations sur tout le conduit vocal.

L'Imagerie par Résonance Magnétique Nucléaire (IRM) [Bae91] est une technique particulièrement intéressante pour pallier ce défaut. Malheureusement l'IRM dynamique n'en est qu'à ses balbutiements (moins de 10 images par seconde dans le meilleur des cas, et avec une faible résolution spatiale.

Aujourd'hui les données cinéradiographiques sont encore d'une grande utilité. Ce sont les seules à offrir aujourd'hui en même temps une résolution spatio-temporelle correcte sur l'ensemble du conduit vocal dans le plan sagittal. Elles sont à la base de l'élaboration de modèles géométriques [Mae89] et sont d'une grande utilité pour l'étude de la coordination spatio-temporelle des articulateurs de la parole [Woo97] [Vax99] [Vil99]. La législation restreint les possibilités d'acquérir de nouvelles données de ce type. Or de nombreuses banques de données existent dans différents laboratoires. Il nous semble de la plus haute importance d'en faciliter l'accès à l'ensemble de notre communauté. Munhall et ses collègues [Mun95] ont fait un magnifique travail de sauvegarde et de distribution de données cinéradiographiques réalisés en Amérique du Nord, essentiellement à l'Université Laval de Québec, sous l'impulsion de Claude Rochette.

En France, nous bénéficions d'une situation tout à fait exceptionnelle. L'Institut de Phonétique de Strasbourg a en effet accumulé, depuis la fin des années 50, plus de 50 enregistrements cinéradiographiques, et ceci sur un ensemble très large de langues. Soutenu par le *programme Ingénierie de Langues* du CNRS, nous avons donc entrepris un travail de mise en forme de ces données avec les objectifs suivants : (1) assurer la sauvegarde des données (actuellement sur films 35mm et bandes audios) par stockage sur un support vidéo de haute qualité, et par numérisation et stockage sur CDROM ; (2) apporter une valeur ajoutée par le biais de tracés sagittaux réalisés par des experts phonéticiens et montrant les limites du conduit vocal ; (3) faciliter l'accès et le traitement de ces données par leur intégration dans une base de données, et la distribution de cette base.

## 2. PROCÉDURE DE SAUVEGARDE ET NUMÉRISATION DES DONNÉES.

### 2.1 Sauvegarde des films

La première étape du travail a consisté à élaborer une procédure de transfert des données, du standard cinématographique 35 mm, vers un standard plus résistant au temps, et plus facilement reproductible. Nous avons choisi le standard professionnel BetacamSP qui est une référence dans le milieu de la conservation de documents cinématographiques, et qui assure une préservation optimale de la précision des images. Nous avons deux types de films disponibles, qui différaient par la vitesse d'acquisition des images, qui était soit de 64, soit de 50 images/seconde. La vidéo restitue les images à 25 images/seconde, sous forme de trames entrelacées à 50 trames/sec. La technique que nous avons choisie, toujours dans le souci de préserver la qualité des enregistrements originaux, associe une image vidéo complète à chaque image originale (préservation de la définition spatiale) et conserve toutes les images (préservation de la définition temporelle). Cette procédure a entraîné évidemment un

ralentissement de la vidéo par un facteur 2.56 (films à 64 images/seconde) ou 2 (films à 50 images/seconde). Ceci n'a évidemment aucune incidence sur les données, mais devra être pris en compte dans toute analyse temporelle ou fréquentielle ultérieure, qui reposerait sur le fichier vidéo.

À l'origine, les images et le son ont été enregistrés sur deux supports différents. L'enregistrement de tops de "synchro image" sur une des pistes de l'enregistrement audio a permis de conserver des traces de leur synchronisation originale. Ceci, associé à l'expertise des phonéticiens repérant des rendez-vous temporels entre l'image et le son (contacts labiaux pour une plosive labiale, rapprochement langue/palais pour une constrictive palatale) a permis une post-synchronisation de qualité. Le son est donc ralenti par le même facteur que la vidéo.

### 2.2. Réalisation de tracés sagittaux

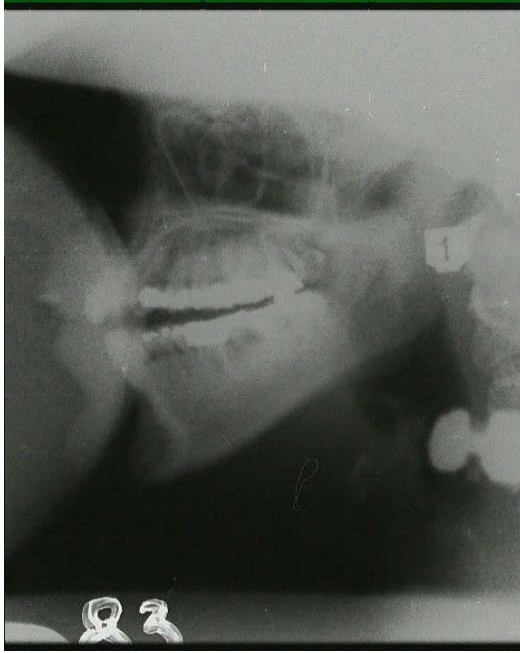
La détermination des contours sagittaux du conduit vocal dans le plan sagittal (palais dur, palais mou, vélum, pharynx, larynx, épiglotte, langue) n'est pas chose aisée. Un certain nombre de travaux ont été développés pour élaborer des techniques d'extraction automatiques de contours [Tie94] [Lap96]. Mais la tâche reste très difficile, et nécessite le plus souvent l'assistance ou la correction d'un expert. C'est pourquoi, et ceci constitue une des particularités de notre projet, nous avons décidé d'associer aux images radiographiques brutes, un certain nombre de tracés des contours sagittaux du conduit vocal. Ces tracés offrent une plus value aux données (cf. Figure 1). Ils peuvent être exploités par un utilisateur non expert, ou servir de point de départ à de nouveaux tracés.

Toutes les images ne sont pas associées à des tracés. En effet, dans certains corpus, les séquences pertinentes sont placées au sein de phrases porteuses, dont nous n'avons pas tenu compte pour la réalisation des tracés.

### 2.3. Numérisation des données

**Numérisation des tracés.** Les tracés manuels ont été numérisés d'une part par scanner, et d'autre part via une "numérisation intelligente", qui stocke séparément les contours du palais dur, du palais mou, du pharynx, de la langue, des dents et des lèvres. Ceci permet une récupération automatique de cette information pour des traitements ultérieurs.

**Numérisation des films vidéo.** Les films ont été numérisés à la volée via une carte de numérisation qui génère des images numériques au standard MJPEG. Ces données ont été ensuite transformées en fichiers vidéo QuickTime, et en fichiers JPEG (images statiques). Le standard QuickTime permet une visualisation des mouvements articulatoires. Mais il correspond à un codage vidéo différentiel, et n'offre donc pas la précision spatiale requise pour un traitement ultérieur des données. C'est pourquoi nous fournissons en parallèle les images une à une codées au standard JPEG, sans perte d'information.



**Figure 1:** Radiographie de [m] dans [mi].(Figure de gauche).

Tracé des sagittaux pour la même image (Figure de gauche)

### 3. INTÉGRATION DANS UNE BASE DE DONNÉES.

Pour permettre une large distribution de ces données, et pour faciliter leur exploitation, nous avons posé, *a priori*, les contraintes suivantes pour la conception du système de gestion de la base de données : (1) travailler avec un SGBD relationnel d'un prix raisonnable et tournant sous des machines standards, accessibles dans tout laboratoire de parole, de type PC ou Macintosh ; (2) permettre la recherche dans la base de données à l'aide de requêtes portant sur les caractéristiques phonétiques des sons (voyelle/consonne, voisé/non voisé, ouvert/fermé...), sur l'écriture phonétique des sons, de manière isolée ou en contexte (n'excédant pas 5 phonèmes), et ceci de manière totalement transparente pour l'utilisateur.

Nous avons donc eu recours, pour les machines de type PC, au SGBD relationnel Microsoft ACCESS, et nous avons développé une interface utilisateur sous Java. Cet interface permet la formulation de requêtes à l'aide de questionnaires que l'on peut remplir de manière très ergonomique, sans que cela ne nécessite de connaissances sur les bases de données. Un manuel utilisateur a aussi été écrit en langage HTML qui permet une recherche rapide de l'information.

Pour l'administrateur de la base de données, cette interface permet aussi d'entrer de manière très simple de nouvelles données dans la base et d'en faire l'étiquetage phonétique. Cet étiquetage phonétique est grossier et détermine les zones (de l'image I à l'image I+11), où l'on peut trouver les différents phonèmes. Les zones de transition ne sont pas

étiquetées, sauf si elles sont explicitement liées à un phonème (dans le cas des liquides par exemple).

Lorsqu'un utilisateur effectuera une requête, le logiciel visualise la séquence QuickTime correspondante, et, s'ils existent, les tracés correspondants. L'utilisateur peut alors repérer le nom de la vidéo correspondante ainsi que les numéros des images retenues. Il pourra ensuite copier depuis le CDROM, les images JPEG correspondantes, ainsi que les fichiers contenant les tracés numérisés, le son de la vidéo complète, et le fichier QuickTime pour analyse ultérieure.

### 4. CONCLUSION.

Quatre films d'une durée de quelques minutes et associés chacun à environ 550 tracés radiographiques, ont été ainsi traités et entrés dans la base de données. Les corpus de ces films sont centrés sur les questions suivantes : *L'effet de jointure en Français* [Wio85], *Les consonnes plosives du Français* [Zer79], et *les nasales du Français* [Fla84].

La base de données sera diffusée gratuitement par l'Institut de Phonétique de Strasbourg, propriétaire des données.

### BIBLIOGRAPHIE

- [Bae91] Baer T., Gore J., Boyce S. & Nye P. (1991) Analysis of vocal tract shape and dimensions using magnetic resonance imaging: vowels. *Journal of the Acoustical Society of America*, 90, 799-828
- [Chi41] Chiba T. & Kajiyama M. (1941) *The Vowel, its Nature and Structure*. Tokyo Kaseikan Pub.

- [Fan60] Fant G. (1960). *Acoustic Theory of Speech Production*. Mouton, La Hague, The Netherlands.
- [Fla84] Flament B. (1984) *Recherche sur la mise en relief en français. Approche théorique et essai de caractérisation phonétique à partir de données de la mingographie et de la radiocinématographie*. Doctorat d'Etat, Institut de Phonétique - Université des Sciences Humaines de Strasbourg.
- [Fou98] Fougeron C. (1998). *Variations articulatoires en début de constituants prosodiques de différents niveaux en français*. Thèse de l'Université Paris III – Sorbonne Nouvelle
- [Har84] Hardcastle W. (1984) New methods of profiling lingual-palatal contact patterns with electropalatography. *Working papers Phonetics Lab.*, 4 (pp. 1-40). Université de Reading
- [Hel81] Hellwag C.F. (1971). *De Formatione loquelæ*. Thèse de Médecine. Université Eberhard-Karl de Tübingen. Réédité dans *Les Cahiers de l'ICP (Bulletin de la Communication Parlée n°1)*. Institut de la Communication Parlée, Université Stendhal : Grenoble, France..
- [Lap95] Laprie Y. & Berger M.-O. (1996). Extraction of Tongue Contours in X-Ray Images with Minimal User Interaction. *Proceedings of ICSP'96* (vol 1. pp.268-271).
- [Mae89] Maeda S. (1989) Compensatory articulation during speech: Evidence from the analysis and synthesis of vocal tract shapes using an articulatory model. In Hardcastle W. & Marchal A. (Eds.) *Speech Production and Modelling* (pp. 131-149). Kluwer: Academic Publishers.
- [Mol60] Moll K. (1960) Cinefluorographic techniques in speech research. *Journal of Speech and Hearing Research*, 3, 227-241.
- [Mun95] Munhall, K.G., Vatikiotis-Bateson, E., & Tohkura, Y. (1995). X-ray Film database for speech research. *Journal of the Acoustical Society of America*. 98, 1222-1224.
- [Per69] Perkell J.S. (1969). *Physiology of Speech Production*. Massachusetts Institute of Technology: Cambridge, Ma, USA.
- [Per92] Perkell J., Cohen M., Svirsky M, Matthies M., Garabieta I., & Jackson M. (1992). Electromagnetic midsagittal articulometer (EMMA) systems for transducing speech articulatory movements, *Journal of the Acoustical Society of America*. 93, 3078-3096
- [Roc73] Rochette C. (1973). *Les groupes de consonnes en Français*. Presses Université Laval, Québec
- [Sch89] Schönle P., Müller C. & Wenig P. (1989). Echtzeitanalyse von orofacialen Bewegungen mit Hilfe der elektromagnetischen Artikulographie. *Biomedizinische Technik*, 34, 126-130
- [Sim67] Simon P. (1967) Les consonnes Françaises. Mouvements et positions articulatoires à lumière de la radiocinématographie. Paris: Klincksieck
- [Sto88] Stone M., Shawker T., Talbot T. & Rich A. (1988). Cross-sectional tongue shape during vowels. *Journal of the Acoustical Society of America*, 83, 1586-1596.
- [Str65] Straka G. (1965). *Album Phonétique*. Presses de l'Université Laval, Québec
- [Tie94] Tiede M. & Vatikiotis-Bateson E. (1994). Extracting articulator movement parameters from a videodisc-based cineradiographic database. *Proceedings of ICSP 94* (pp.45-48)
- [Wes94] Westbury J.R., Turner G. & Dembowski J. (1994) *X-ray microbeam speech production database users' handbook*. Waisman Center, Université du Wisconsin.
- [Vax99] Vaxelaire B., Sock R., Bonnot J.F. & Keller D. (1999). Anticipatory labial activity in the production of French rounded vowels. *Proceedings of ICPH 99* (Vol. 1., pp. 53-56).
- [Vil99] Vilain A., Abry C. & Badin P. (1999). Motor equivalence evidenced by articulatory modelling. *Proceedings of Eurospeech99* (Vol.1, pp. 169-172)
- [Wio85] Wioland F. (1985) *Faits de jointure en français. Implications aux niveaux articulaire et acoustique. Incidences sur le plan des fonctions linguistiques*. Doctorat d'Etat, Institut de Phonétique - Université des Sciences Humaines de Strasbourg.
- [Woo79] Wood S.A.J. (1979). A radiographic examination of constriction location for vowels. *Journal of Phonetics*, 7, 25-43
- [Woo79] Wood S.A.J (1997). A cinefluorographic study of the temporal organization of articulator gestures: Examples from Greenlandic. *Speech Communication*, 22, 207-225.
- [Zer79] Zerling J.-P. (1979) *Articulation et coarticulation dans des groupes occlusive-voyelle en français. Etude cinéradiographique et acoustique : contribution à la modélisation du conduit vocal*. Doctorat 3<sup>e</sup> Cycle, Institut de Phonétique, Université de Nancy II.