

AN X-RAY DATABASE FOR FRENCH

Alain Arnal¹, Pierre Badin¹, Gilbert Brock², Pierre-Yves Connan², Evelyne Florig², Noël Perez¹,
Pascal Perrier¹, Pela Simon², Rudolph Sock², Laurent Varin¹,
Béatrice Vaxelaire² & Jean-Pierre Zerling²

1: Institut de la Communication Parlée, INPG & Université Stendhal, Grenoble, France

2: Institut de Phonétique de Strasbourg– Université Marc Bloch

22 Rue Descartes - 67084 Strasbourg, France.

ABSTRACT

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

1 1. INTRODUCTION

A systematic study of the control of speech production implies acquiring significant amount of data at different levels, namely the physiological, articulatory and acoustic domains. As concerns articulatory acquisition techniques, numerous and varied experimental setups have, indeed, been developed and refined within the international speech community in the last decade. A priority in elaborating such techniques has mainly been the wish to obtain good spatial and temporal resolutions, while minimizing potential negative effects on subjects' health.

In fact, until the seventies, apart from the use of well-known self-observing methods with mirrors, by precursors like Hellwag [1], basic techniques for acquisition of articulatory data relied essentially on palatography and radiography. Palatography had the advantage of being simple and inexpensive. However, it was static until the seventies, and, by essence, it restricts the domain of observation to the vocal tract regions where tongue and palate are in contact with each other. On the contrary, radiography required sophisticated medical equipment, but it gave a full representation of the vocal tract, in the sagittal plane, from the glottis to the lips ([2]; [3]). Moreover, at the end of the fifties, radiography became dynamic, with the advent of cineradiography, which allowed, for the first time, a fairly precise observation of the movements of the articulators and other speech structures. This explains why cineradiography was very often preferred to the other experimental tools, and was at the basis of reference works on the articulation of sounds, especially in France with the studies by Georges Straka [4] and colleagues ([5]; [6]). For the study of English sounds, precursor researchers were Moll [7] and Perkell [8].

Nevertheless, exposure of healthy subjects to ionizing rays rapidly posed an ethical problem, as this was not in conformity with the legislation of many countries. Moreover, the time resolution of cineradiography does not exceed 50 frames per second, which is insufficient for a precise analysis of the timing of consonants, and particularly of plosives. For all these reasons, researchers looked for new experimental devices their.

The X-ray Microbeam system minimizes exposure of subjects to ionizing rays by focusing X-ray beams on very specific regions of the vocal tract [9]. Electro-palatography records the palatal contacts at 100 or 200 Hz ([10]; [11]). Ultrasound techniques [12] provide 3D data at a sampling rate of around 100 Hz, in a limited part of the vocal tract. Electro-magnetometer ([13]; [14]) monitors displacements, of 5 to 10 points in the vocal tract, at a sampling rate that can exceed 1 kHz. All of these techniques offer a good spatial resolution, and temporal resolutions superior to that of cineradiography. They have significantly contributed to a better understanding of motor control in speech. However, they all have a common drawback: neither of them allows acquiring simultaneous information from the entire vocal tract.

Nuclear Magnetic Resonance Imaging (M.R.I.) could be particularly useful in overcoming this limitation [15]. Unfortunately, dynamic M.R.I. is still in its infancy (less than 10 frames per second in the best of cases, with a low spatial resolution).

Today, cineradiographic data are still of utmost use. Currently, only such data provide, simultaneously, a correct spatial and temporal resolution of the entire vocal tract in the sagittal plane. They are at the basis of geometric models [16] and are highly useful in the study of the spatio-temporal co-ordination of articulators ([17], [18], [19], [20]). Now, health legislation restricts acquisition possibilities of new X-ray data. However, there are many such databanks in various laboratories. It is our priority to ensure their easy access to the speech community. Munhall and colleagues [21] accomplished a remarkable preservation and distribution of cineradiographic data that were acquired in North America, and essentially at Laval University in Quebec in Claude Rochette group.

In France, the Phonetics Institute of Strasbourg offers an exceptional opportunity. This Institute has accumulated, since the end of the fifties, more than 40 cineradiographic recordings, for a large number of languages. Supported by the *Ingénierie des Langues programm* at the C.N.R.S., we have thus started to organize these data with the following aims: (1) ensure their preservation by storage on a high quality video, and by digitization and storage on CDROM (after their acquisition, they were stored on 35 mm films and audio tapes); (2) facilitate their analysis by including sagittal profiles that were hand-drawn by expert phoneticians, showing vocal tract contours; (3) ensure

easy access and processing of these data by integrating them into a database for distribution.

2 PRESERVATION PROCEDURE AND DATA DIGITIZATION

2.1 Preserving the movies

The first step consists in elaborating a data transfer procedure, from a 35 mm cinematographic standard, to one that is more resistant and can be more easily duplicated. The chosen standard is the professional Betacam SP, which is a reference in the domain of the preservation of cinematographic documents, and also ensures an optimal image accuracy. Two types of movies that differ in speed of image acquisition, 64 or 50 images/second, are available, while the European video standards restore images at 25 images per second, in the form of interlaced frames at 50 frames/sec. The technique adopted, still with the requirement to preserve the quality of original recordings, associates a full video image (2 frames) with each original cineradiographic image (the spatial definition is, thus, preserved) and keeps all images (the temporal definition is, thus, preserved). This procedure obviously provokes a slowing down of the video by a factor of 2.56 (for the movies at 64 frames/second) or of 2 (movies at 50 frames/second). This, of course, does not have any effect on the data, but it should, however, be taken into account for any possible temporal or spectral analysis that may be directly based on the video files.

Originally, the images and the sound were recorded on two different media. Recordings of pips of "image synchronization" on one of the audio recording tracks, allowed conserving traces of their original synchronization. This method, together with the expertise of phoneticians who detected temporal correspondences between images and sounds (labial contact for labial plosives, tongue-palate constriction for palatal fricatives) allowed a high quality post-synchronization. The sound is consequently slowed down by the same factor as that of the video.

2.2 Acquiring sagittal tracings

Determining sagittal contours of the vocal tract in the sagittal plane (hard palate, soft palate, velum, pharynx, larynx, epiglottis and tongue) is not an easy task. A few works have been developed to elaborate automatic contour extraction techniques ([22]; [23]). However, the task remains very difficult and, quite often, calls for assistance or correction by an expert. Consequently, as a specificity of this project, it was decided to associate a certain number of sagittal tracings of the vocal tract with the raw X-ray data. These tracings represent an additional value of the data (see Figure 1). They can, thus, be analyzed by all the speech scientists, even by the ones having no experience with X-ray data, or can serve as a starting point for drawing new tracings.

2.3 Digitizing the data

2.3.1 Digitizing the tracings.

The hand-drawn tracings were in parallel digitized by a scanner, and by a software that separately stores the contours of the hard palate, the soft palate, the pharynx, the tongue, the teeth and the lips. It is thus possible to automatically recover this information for subsequent data processing.

2.3.2 Digitizing the video movies.

The movies were digitized according to the MJPEG standard. These data were then transformed into QuickTime video files and into JPEG files (static images). The QuickTime standard allows visualizing articulatory movements. However, it corresponds to a differential video coding, and consequently does not offer the required spatial precision for subsequent data processing. Thus, it was decided to also furnish the images, one-by-one, coded in JPEG standard, without any loss in information.

3 INTEGRATION INTO A DATABASE

To ensure a large distribution of these data and to facilitate their exploitation, the conception of the database had to satisfy a number of requirements, which are as follows: (1) the final product should be a relational database using a database manager of a reasonable price, that is operational on standard machines (PC or Macintosh types) and affordable by any speech laboratory; (2) search procedures should be based on requests that specify the phonetic characteristics of the sound (vowel/consonant, voiced/voiceless, open/closed...), or the phonetic transcription of the sound, in isolation or in a given context (less than 5 phonemes); (3) the elaboration of the requests should be simple for the users, and should not, in particular, necessitate any knowledge about the database programming.

For PC machines, Microsoft ACCESS was used, with a user friendly interface developed under JAVA. This interface allows formulating requests through questionnaires that can be filled out in a very ergonomic way, without requiring knowledge on databases. A user's manual was also written in HTML language, which allows a quick search of the desired information.

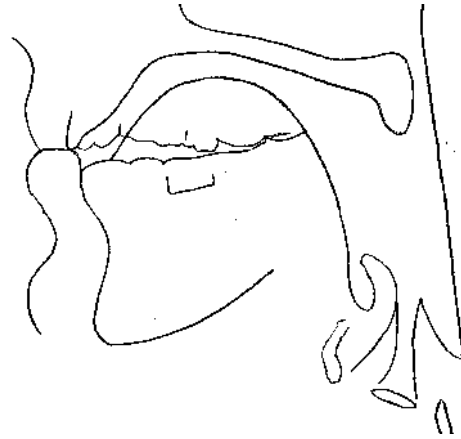
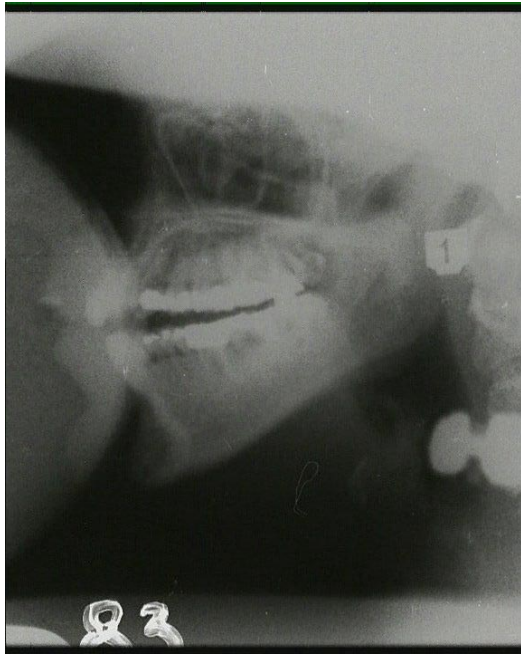
This interface allows integrating, in a very simple way, additional data into the database, together with labeling them phonetically. We used large phonetic labeling and determined the time intervals (from image **I** to image **I+1**), where the desired phonemes are located. Transitions are not labeled, unless when explicitly related to a time variable phoneme (in the case of liquids for example).

When a user makes a request, the software displays the relevant QuickTime sequence and, if they exist, the corresponding hand-drawn tracings. The user can then note the name of the video that matches, as well as the numbers of the selected images retained. He may then copy the adequate JPEG images from the CDROM, together with the files containing the digitized tracings, the sound of the complete video and the QuickTime file for subsequent analyses.

4 CONCLUSION

Four movies of a few minutes, each associated with about 550 X-ray tracings, were thus processed and included into the database. The corpus of these movies focus on the following problems: The effect of juncture in French [24], French plosives [25] and French nasals [26].

The database will be distributed for free by the Phonetics Institute of Strasbourg, who is the owner of the data.



**Figure 1. X-ray of [m] in [mi]. Figure on left.
Sagittal tracings of the same figure. Figure on right.**

5 REFERENCES

- [1] Hellwag C.F. (1781). 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..
- [2] Chiba T. & Kajiyama M. (1941) The Vowel, its Nature and Structure. Tokyo Kaseikan Pub.
- [3] Fant G. (1960). Acoustic Theory of Speech Production. Mouton, La Hague, The Netherlands.
- [4] Straka G. (1965). Album Phonétique. Presses de l'Université Laval, Québec
- [5] Simon P. (1967) Les consonnes Françaises. Mouvements et positions articulatoires à lumière de la radiocinématographie. Paris: Klincksieck
- [6] Rochette C. (1973). Les groupes de consonnes en Français. Presses Université Laval, Québec
- [7] Moll K. (1960) Cinefluorographic techniques in speech research. Journal of Speech and Hearing Research, 3, 227-241.
- [8] Perkell J.S. (1969). Physiology of Speech Production. Massachusetts Institute of Technology: Cambridge, Ma, USA.
- [9] Westbury J.R., Turner G. & Dembowski J. (1994) X-ray microbeam speech production database users' handbook. Waisman Center, Université du Wisconsin.
- [10] Hardcastle W. (1984) New methods of profiling lingual-palatal contact patterns with electropalatography. Working papers Phonetics Lab., 4 (pp. 1-40). Université de Reading
- [11] Fougeron C. (1998). Variations articulatoires en début de constituants prosodiques de différents niveaux en français. Thèse de l'Université Paris III – Sorbonne Nouvelle
- [12] 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.
- [13] Schönle P., Müller C. & Wenig P. (1989). Echtzeitanalyse von orofacialen Bewegungen mit Hilfe der elektromagnetischen Artikulographie. Biomedizinische Technik, 34, 126-130
- [14] Perkell J., Cohen M., Svirsky M, Matthies M., Garabieta I., & Jackson M. (1992). Electro-magnetic midsagittal articulometer (EMMA) systems for transducing speech articulatory movements, Journal of the Acoustical Society of America. 93, 3078-3096
- [15] 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
- [16] 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.

- [17] Wood S.A.J. (1979). A radiographic examination of constriction location for vowels. *Journal of Phonetics*, 7, 25-43
- [18] Wood S.A.J (1997). A cinefluorographic study of the temporal organization of articulator gestures: Examples from Greenlandic. *Speech Communication*, 22, 207-225.
- [19] Vaxelaire B., Sock R., Bonnot J.F. & Keller D. (1999). Anticipatory labial activity in the production of French rounded vowels. *Proceedings of ICPhS 99* (Vol. 1., pp. 53-56).
- [20] Vilain A., Abry C.& Badin P. (1999). Motor equivalence evidenced by articulatory modelling. *Proceedings of Eurospeech99* (Vol.1, pp. 169-172)
- [21] 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.
- [22] Tiede M. & Vatikiotis-Bateson E. (1994). Extracting articulator movement parameters from a videodisc-based cineradiographic database. *Proceedings of ICSLP 94* (pp.45-48)
- [23] Laprie Y. & Berger M.-O. (1996). Extraction of Tongue Contours in X-Ray Images with Minimal User Interaction. *Proceedings of ICSLP'96* (vol 1. pp.268-271).
- [24] Wioland F. (1985) *Faits de jointure en français. Implications aux niveaux artic-uloire et acoustique. Incidences sur le plan des fonctions linguistiques.* Doctorat d'Etat, Institut de Phonétique - Université des Sciences Humaines de Strasbourg.
- [25] Zerling J.-P. (1979) *Articulation et coarticulation dans des groupes occlusive-voyelle en français. Etude cinéradiographique et acous-tique : contribution à la modélisation du con-duit vocal.* Doctorat 3^e Cycle, Institut de Phonétique, Univ. de Nancy II.
- [26] 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 radiociné-matographie.* Doctorat d'Etat, Institut de Phonétique - Université des Sciences Humaines de Strasbourg.

Acknowledgements

This work was supported by the CNRS.