Abstract: Research works in the field of speech production studies have been carried out at the Phonetics Institute of Strasbourg (IPS) using cineradiography since the 1960s. This contribution will take as a starting point the article entitled “Histoire du radiocinéma à l’Institut de Phonétique de Strasbourg” [14] in order to summarise the main milestones in the evolution of this X-ray method at the institute. While [14] privileged the historical perspective of cineradiography, we shall focus here on the history of the contributions of these studies to the advancement of fundamental knowledge in the field of speech production, from the 1960s to 2000.

1 Introduction

The purpose of this paper is to provide a synthesis of works – mostly PhD dissertation theses – derived from the numerous X-ray movies produced in Strasbourg over 40 years, from 1960 to 2000. This contribution will take as its starting point the article entitled “Histoire du radiocinéma à l’Institut de Phonétique de Strasbourg” [14] in order to summarise the main milestones in the evolution of this X-ray method at the Phonetics Institute of Strasbourg.

We shall focus on the advancement of knowledge in speech production, based on information provided by the X-ray movies, all extracted from the IPS cineradiographic database, which allow studying the movement of the speech articulators in relation to the acoustic signal. These movies offer an overall correct spatiotemporal resolution of the vocal tract in the sagittal plane. They are the basis for the elaboration and evolution of several speech production models.

The progression of the presentation of the different investigations in this paper will follow the technological evolution of the cineradiographic method, i.e. globally according to the date of the movies. The figures chosen to illustrate the main results obtained in a given study often correspond to the measurement grids devised by the authors themselves.

2 1960: The beginnings

The importance of cineradiography as a method of investigation in the field of phonetics was well established by the 1960s. This technique was regarded as an adequate means of capturing movement of articulatory organs in the medial sagittal plane, by providing a direct image of these organs.
This research method was begun in 1959 in Strasbourg on the initiative of Georges Straka, founder of the phonetics institute, as a result of his interest in the physiology of speech production. Thanks to the collaboration of Professor Christian M. Gros, Director of the Central Radiology Department of the Faculty of Medicine, who made the facilities available to the Phonetics Institute of Strasbourg, and his collaborator, Dr. Pierre Bloch, members of the IPS were able to produce a series of phonetic films as early as January 1960 on the articulation of utterances in French and other languages.

Several research centres for experimental phonetics (Berlin 1955, Prague 1955, Turku 1957, Bonn 1958) were interested in this approach; one study particularly worth highlighting during this period is an article by A. Sovijarvi [7] on the articulatory variations of the vowel [a] in Finnish, based on data from a film shot at 24 frames per second (fps) in 1957, showing articulatory movements during the pronunciation of a series of sentences in Finnish, using a synchronised sound recording.

A new impulse to phonetic research using cineradiography was given by the fourth International Congress of Phonetic Sciences in Helsinki in 1961. This is evidenced by the papers of Joanne D. Subtelny and J. Daniel Subtelny, H. M. Truby, A. Sovijarvi and P. H. Damsté [7]. Also in 1961, Péla Simon [12] reported the state of her research, based on X-ray movies, into the quantitative relationships between the organic closures and movements of French consonants in the spoken chain and on several articulatory variations.

For the first films, the field of investigation was based on an image intensifier with a diameter of 13 cm, which increased the brightness of the screen by a thousand times. In 1960, this amplification coefficient was increased to 3000. Films were shot on 16 mm Kodak Tri-X, using an Arriflex electric camera equipped with a special wide-aperture lens. Frame-rates of 50 fps and even 64 fps were achieved without the use of contrast media on the surface of articulatory organs and with a low dose of X-rays. Having shot some films at 64 fps, a frame-rate that had never previously been achieved in speech cineradiography, Péla Simon and colleagues then reduced the speed to 50 fps, which they felt was sufficient.

In order to examine the films, the use of a special projector, equipped with a freeze frame facility, made it possible to observe each image individually and to follow the movement of the articulators, frame-by-frame, at regular intervals of 20 milliseconds, and to note the speed of their movements and the duration of their relative stability. During filming, the subject stood 70 cm from the X-ray source, with the head in the medial sagittal plane, 20 cm from the screen (see Figure 1).

Figure 1. A recording session at the Central Radiology Department of the Faculty of Medicine in Strasbourg (1960)
However, a major drawback was the restricted field of view 13 cm in diameter. Only part of the supraglottal organs could be captured and therefore a complete image – from the larynx to the hard palate and from the outer side of the lips to the pharyngeal wall – could not be obtained, as would have been desirable.

Péla Simon and colleagues first limited themselves to a centring on the oral region (see Figure 2, left), in order to analyse the articulation in the narrow sense of the word, i.e. from the incisors to the pharyngeal wall, and from the upper surface of the hard palate to the lower jaw. Then, they proceeded to another centring (see Figure 2, right) to observe the functioning of the velum by moving the field of observation backwards. An audio tape recording was taken simultaneously.

Six films were shot using this procedure. Four were devoted to French, one to Portuguese, one to Serbo-Croatian. The Portuguese and Serbo-Croatian films represent series of short sentences. For French, two subjects, one male and one female, uttered a series of sentences in which the articulations appear in various positions and surroundings. This corpus was followed by a text spoken at normal and fast speech rates.

Regarding general phonetic phenomena, these first films highlighted the fact that utterances consisted essentially of organic movements – their preponderance over static positions is striking. The films also show that closures exist and have characteristics specific to the sounds where they occur. These closures could be modified significantly between two successive movements, both in terms of quality and quantity (see [12]).

Moreover, these films permitted verification of certain results acquired by other methods, more specifically by static radiography, such as: 1) the impact of articulatory energy on the aperture and on the angle of the maxillaries (under the effect of reinforcement, vowels tend to open, consonants tend to close); 2) the duration of vowels (closed vowels are more closed when they are long, whereas for open oral and nasal vowels, the contrary occurs: they are more open when they are long; 3) the maximum narrowing of the vocal tract is in the pharynx, between the posterior region of the tongue dorsum and the pharyngeal wall, for posterior vowels (this certainly justifies their qualification as pharyngeal vowels); 4) the precision of certain places of vowel articulation, for example [u] being more advanced than [o]; 5) the realisation of the schwa is rather central; 6) the relationship between oral and nasalised vowels, the latter being more open than their oral counterparts; 7) the pre-dorsal nature of the consonant [s]; 8) the progressive lowering of the tongue body behind the apical-alveolar contact for [l]. For a discussion of these results, see [12].

![Figure 2. Centring on the oral region (left) and on the velar region (right)](image)

---

1 The quality of these documents, scanned from the originals, is not always of high quality. However, they do reflect the overall evolution of research and publications.
3 1964 & thereafter: increased field of view

In 1964, the Central Radiology Clinic of the Hospital of Strasbourg made a major advance by acquiring an image intensifier with a 23 cm diameter, compared with the previous intensifier whose diameter was 13 cm. With this size, it was hence possible to have all of the supraglottal organs on the same image, from the lips to the pharyngeal wall, from the larynx to the root of the tongue, up to the nasal cavities. Many researchers came to the IPS for training in this speciality and to produce and analyse films for research into diverse languages. These films were utilised for phonological and phonetic analyses, for specific research into the dynamic organisation of speech and articulatory strategies, for differences in speech rate and intra-speaker variations, or for the study of pathological productions. By the 1st of February 1967, the Institute had acquired more than sixty films.

We shall now briefly present some results obtained from a selection of these films in the chronological order of their acquisition.

In 1965, Kieffer [8] analysed the productions of a male subject at the end of speech therapy, who presented a persistent nasal obstruction with deviation of the nasal septum and a very short velum, a mixed rhinolalia, to observe the articulatory particularities and the shapes of the vocal tract due to this pathology. She observed a number of peculiarities on the sagittal profiles. They concern not only the velum and the velopharyngeal closure, but all of the articulatory organs. The production of an exaggerated contraction effort to close the nasopharyngeal passage extends more particularly to the tongue, which tends to rise and move backwards. This leads to a shortening of the oral cavity, behind the occlusal contacts of the alveo-dentals.

The main thrust of the research reported in [13] was to provide a solution to the problem of the existence or absence of articulatory closures as a function of the nature of French consonants. One important finding was that within a continuous movement, the organs attain a position for each articulation that can be considered characteristic of a given consonant, its occlusion or constriction. However, the organs which are not directly involved in the formation of the occlusion/constriction generally continue to move, while the articulatory organ involved maintains its position for a given lapse of time. The duration of this “closure” (occlusion or constriction) is determined by a whole series of factors, such as the position of the consonant in the syllable or word, and the voiceless or voiced nature of the consonant.

In 1969, Petursson [11] proposed a new articulatory system for Icelandic vowels, which he grouped into 4 series according to place of articulation: 1) a pre-palatal series, very anterior, grouping the vowels [i] short and long; 2) a palatal series, under the hard palate [i e y ö] short and long; 3) a velar series [u o] short and long; 4) a pharyngeal “series” for vowel [a], the most posterior of the system. He also defined 4 degrees according to aperture (distance from the back of the tongue to the nearest point of the hard palate).

Investigating French vowels, [5] revealed that there was a vowel hold phase just as there was a consonantal one. This phase of stability was not absolute, nor did it concern all the organs. Its duration in relation to onset and offset movements for the required articulatory position was essentially a function of the length of the vowel.

In one of his works on articulatory movements and their acoustic efficiencies in the production of Bulgarian vowels, Tilkov [16] considered the entire supraglottal cavity in his analyses, based on the location of maximum vocal tract narrowing (acoustically significant). He concluded that the following three parameters were the best means to classify vowels into articulatory types: location of the maximum constriction, diameter of the maximum constriction, and contribution of the lips.

Finally, [10], in 1973, pointed out the special character of pre-nasals, the quantitative relationship between the oral and nasal phases of the prenasalised Akan (Fanti) consonants. From
their comparison with nasals, it was clear that they are in complete opposition. The pre-nasalised plosives were neither a simultaneous combination of two segmental units, because the place of articulation of the nasal phase was irrelevant, nor a sequence of a nasal consonant + an oral consonant. Pre-nasals were longer and therefore more tense [+ tense] than nasals. Their occlusion width or contact extent increased and the rise of the velum was slow and progressive towards the oral consonant.

4 1976 - Image/sound synchronisation and automatic image numbering: a turning point

It was in 1976 that cineradiographic installations were used in the department of haemodynamics and cardiovascular operations at the Strasbourg-Schiltigheim Medical-Surgical Centre. This equipment, adapted for the study of speech production by [6], was operated under medical supervision (see Figure 3). It included: 1) a 35 mm camera, electronically regulated, thus avoiding fluctuations in speed; 2) the synchroniser, which allowed the film and the corresponding oscillogram to be synchronised, frame-by-frame.

4.1 Electronic data processing

In order to visualise the different movements of the vocal organs, an X-ray source emitted rays on the corresponding organic region. The subject was seated in a suitable chair with a headband, which prevented the speaker from moving his or her head during recordings. The subject was placed 50 cm from the source and 20 cm from the image intensifier. X-rays were not emitted continuously but in pulses, synchronised with the opening of the camera shutter.

The camera operated in two stages: 1) a duration of 2 ms corresponding to the opening of the shutter and shooting (this 2 ms pulse activated the X-ray source, which only operated during this lapse of time); 2) a further 18 ms, allowing the film to be advanced – shutter closed – during which there was no X-ray emission. As a result, the continuous radiation represented only 10% of total shooting time. Thus, for films with a duration of 2 minutes, the equivalent continuous radiation was 12 seconds.

Figure 3. Recording session in the Radiology Department of Schiltigheim Hospital (from [17]).
The camera pulses were processed by the synchroniser and recorded on the second track of a tape recorder, the first track recording speech from a microphone, which was placed near the speaker's mouth. Two documents of a different nature result from the image/sound synchronisation: the mingogram (or the acoustic signal, when the latter is digitised) and the film. The image/sound synchronisation was reflected on the film by a digital marker on each image, which corresponded to the pulses on the bottom line of the mingogram. Each pulse on this line of the mingogram was equivalent to a frame of a duration of 20 ms, numbered on the film. On the mingogram, in addition to the common image/sound marker appeared the impulses corresponding to the images of the film (see Figure 4, A-D).

![Figure 4. Image/sound synchronisation (adapted from Brock [6]).](image)

### 4.2 Hand drawn tracings

In order to obtain tracings and measurements, the films thus obtained (50 frames per second, i.e. 1 image every 20 ms) were projected at life size using a vertical stand specially designed for the purpose at the IPS (see Figure 5).

The acoustic signal, acquired synchronously, allowed examining, if necessary, the relationship between the articulatory-geometric level and the acoustic-temporal level. Because of the significant background noise level, acoustic analyses were mainly performed in the temporal domain, as the spectrum did not always present clearly defined formant structures.

![Figure 5. Tracings drawn using vertical projection of the film (from [17]).](image)
4.3 Measurements

Measurements of the profiles were taken using a grid developed for each speaker. The method consisted in choosing a reference image in the film for each speaker, from which was established a measurement grid. The reference image chosen corresponded to a resting position, located between two sentences of the corpus, with the velum lowered. For each speaker, fixed points (the hard palate, the teeth and the upper incisors) as could be observed in a profile view, were also determined.

The experimenter then used a sheet of transparent graph paper, on which the profile and its fixed points were traced. On an orthonormal reference point, one looked for a radius which, from the end of the incisor to point 0 of the axis, was equal to 5 cm; this was drawn on each profile tracing. This grid was then applied to each projected image and measurements following the trajectory of each parameter were thus obtained. An error margin of ±1 mm, which may be due to the setting of the grid on the profile view, was usually expected (see Figure 6).

Figure 6. Profiles of the vocal tract showing the measured parameters. Male subject on the left, female on the right (from [17]).

5 Contributing to advancement of basic knowledge in speech production

What follows is a condensed chronological presentation of some of the various investigations that relied on the then novel acquisition system.

In 1977, Bonnot [2] carried out an experimental cineradiographic research on the nature of emphatic consonants in classical Arabic. For these consonants, he reported two places of articulation: an anterior place of articulation and a remarkably stable posterior place of articulation, which consists of a constriction between the root of the tongue and the posterior wall of the pharynx (mesopharynx).

Investigating the vertical displacement of the hyoid bone in relation to frequency variation, [3], in 1978, presented results on the movements of the hyoid bone/larynx set. The movements were shown to be involved in the F0 control mechanism, as they do act on the vocal folds (in the longitudinal direction and on their thickness).

Maeda [9], in 1979, carried out a statistical analysis of tongue positions in a preliminary study of French vowels. One finding was that statistical processing could account for the multidimensional characteristics of articulatory organisation. Based on such analyses, he succeeded in quantifying, to a certain extent, the behaviour of coordinated non-linear structures, and could hence contribute to the shaping of a model integrating variability. Such models operate with the
help of parameters which are not articulatory, but determined by factorial analysis or principal component analysis.

In the same year, Zerling [20] looked at articulations and coarticulation in occlusive-vowel groups in French. He chose the principal component method, based on 60 sagittal views and on the exploitation of 380 images. He showed that it was possible to find the principal muscular activities of the vocal tract using an articulatory model based on the variations of three components.

Looking at plosives in Turkish spoken in Istanbul, [15] found, in 1981, that the duration of voiceless plosives was longer than that of voiced ones, because the burst-friction of the voiceless plosives was longer. They could have aspirated variants without aspiration being a relevant feature. Oral voiced plosives in absolute initial position were devoiced, as well as in contact with a voiceless consonant, regardless of their position in the syllable or the sentence.

Studying juncture in French, its implication for the articulatory and acoustic levels, and its impact on linguistic functions, in 1985, [19] revealed the fact that distinctive features contained in two contrastive transitions between two sounds could be characterised in the same way as accentual and intonational features.

In 1986, the famous Blue Book appeared [4], which is a collection of vocal tract diagrams, drawn from X-ray movies. This reference book is entitled “Cineradiography of French vowels and consonants. Collection of synchronized documents for four subjects: lateral views of the vocal tract, frontal views of the labial orifice, acoustic data” [our translation].

In [18], sagittal profile measurements obtained at normal speaking rate showed that the lip-lip and tongue-palate (apex and tongue dorsum) contact extents for double abutted consonants were equal to or eventually longer than those for their singleton counterparts. This finding was, of course, valid for a term-by-term and intra-speaker comparison only, for all the linguistic categories examined, namely bilabial, apical and velar contacts, and for the two speakers investigated. This difference in contact area can be seen in Figure 7 showing the sagittal profiles for: /aka/ vs. /akka/ and /ada/ vs. /adda/.

Figure 7. Profile views showing the production of sequences /aka/ and /akka/ (left) and the sequences /ada/ and /adda/ (right), at fast speech rates. Longer contact extents can be observed for double abutted consonants (dotted lines), compared to their singleton counterparts (solid lines). Male speaker. From [18].

In fast speech, a trend emerged suggesting that the contact extent was longer for double abutted consonants; thus indicating the relevance of this parameter in differentiating the two phonetic categories, i.e. even when the linguistic system was subject to prosodic perturbation, by increasing speech rate. It is important to note, however, that the increase in speaking rate led to an increase in contact extents for both the single and double abutted categories. It was posited that this parameter did not only correspond to the obstruent phase of the plosive, but could reflect – in cases where the linguistic system was subjected to a perturbation – other factors such as the speed and impact of the lingual muscle tissues on the palate.

The study also revealed longer acoustic consonantal closures for the double abutted consonants, at both speaking rates.
6 Conclusion

In the years following 2000, significant advances were made with developments in digital X-ray technology [1]. As a result of such breakthroughs, there was a significant reduction of ionising radiation and diffusion of radiation during data acquisition. Radiation level during the experiment became less than the naturally occurring background radiation levels, which typically range from about 1.5 to 3.5 millisievert per year. Nevertheless, ethical permission for the acquisition of new X-ray movies in Strasbourg was withdrawn from 2006.

Dynamic MRI (Magnetic Resonance Imaging) techniques, which continue to be developed in the area of speech production, have replaced cineradiography. However, existing cineradiographic data are still of great value. They offer an overall correct spatiotemporal resolution of the vocal tract in the sagittal plane. They have been the basis for the elaboration and evolution of several speech production models [1].

7 References

Note: TIPS = *Travaux de l’Institut de Phonétique de Strasbourg*


