



HAL
open science

Multisensor platform for speech physiology research in a phonetics laboratory

Jacqueline Vaissière, Kiyoshi Honda, Angelique Amelot, Shinji Maeda, Lise Crevier-Buchman

► To cite this version:

Jacqueline Vaissière, Kiyoshi Honda, Angelique Amelot, Shinji Maeda, Lise Crevier-Buchman. Multisensor platform for speech physiology research in a phonetics laboratory. The Journal of the Phonetic Society of Japan, 2010, 14 (2), pp.65-78. halshs-00676165

HAL Id: halshs-00676165

<https://shs.hal.science/halshs-00676165>

Submitted on 3 Mar 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Vaissière, J., Honda, K., Amelot, A., Maeda, Sh., Crevier-Buchman, L., (in press), “Multisensor platform for speech physiology research in a phonetics laboratory.”, *The Journal of the Phonetic Society of Japan*.

Multisensor platform for speech physiology research in a phonetics laboratory

Jacqueline Vaissière⁽¹⁾, Kiyoshi Honda⁽¹⁾, Angélique Amelot⁽¹⁾, Shinji Maeda⁽²⁾, and Lise Crevier-Buchman^(1,3)

(1) Phonetics and Phonology Laboratory, CNRS / University Sorbonne Nouvelle, 19 Rue des Bernardins 75005 Paris

(2) ENST / CNRS, 46 rue Barrault, 75634 Paris

(3) Hôpital Européen Georges Pompidou, 20 rue Leblanc, 75908 Paris Cedex 15

Keywords: multisensor platform, speech physiology, phonation and articulation, signal detection, motion tracking, medical imaging

1. Introduction

Laboratoire de Phonétique et de Phonologie (LPP)¹ is specialized in education and research in experimental phonetics and in phonology. LPP inherits these missions from the first phonetics laboratory created in Paris for Pierre-Jean Rousselot (1846-1924), who is often considered as the founder of experimental phonetics. Currently, LPP is a multi-disciplinary research team, which includes medical doctors and speech therapists, along with phoneticians and phonologists.

The long-term goal of LPP is to achieve an integrated model of phonetics and phonology. It also works on applications to clinical phonetics and language learning. LPP has recently assembled a research platform for investigating the behavior of each speech organ involved in speech production. The platform is elaborated by a joint effort between engineers, phoneticians and clinicians. It provides tools to investigate various phenomena of coordination and compensation across speech organs that are observed in the production of speech by normal or pathological speakers, foreign language learners, and singers. Non-invasive instrumentation techniques are located at LPP to be open to a broader public of researchers. Mildly invasive techniques are applied at the Georges Pompidou European Hospital (otolaryngology department) for diagnostic purposes and for research by the members of LPP.

Table 1 lists the instrumentation techniques that are being used at LPP to examine the speech organs: larynx, tongue, velum, lips, jaw, and face. Electroglottography (EGG) and photoelectroglottography (PGG) are used to evaluate the degree of contact between the two vocal folds, and the degree of opening of the glottis, respectively. Endoscopy is used for direct observation of the laryngeal components. Ultrasonography (USG) allows us to visualize tongue shapes, and static or dynamic palatography methods to acquire data on tongue-palate contact. Nasal acoustic and vibratory signals are recorded using a nasal microphone and a contact transducer. Nasal and oral airflow rates are measured using an aerodynamic sensing system (EVA2TM). Visual estimation of velar height is conducted by endoscopy. Photonography (PNG) evaluates the opening of the velopharyngeal port. Video and motion capture systems are used to study movements of the lips, jaw and face, in particular to investigate visual cues to labial articulation. Several instruments are used simultaneously whenever possible, depending on the instruments.

Table 1. Target organs, instrumentation techniques, and parameters to be measured.

Organs	Instrumentation techniques	Parameters to be measured
Larynx	Electroglottography (EGG)	Contact between the two vocal folds
	Photoelectric glottography (PGG)	Degree of opening of the glottis
	Rigid and flexible endoscope	Visualization of laryngeal components
Tongue	Ultrasonography (USG)	Visualization of contour of the tongue
	Static and dynamic palatography	Degree of linguo-palatal contact

Velum	Nasal acoustic microphone, and Piezoelectric transducer	Nasal acoustic signal, and intensity of skin surface vibrations
	Aerodynamic sensors, EVA2 ^{TM 1}	Nasal and oral airflow, and air pressure
	Videofiberscopy	Height of the nasal surface of the velum
	Photonasography (PNG)	Opening of the velopharyngeal port
Lips and face	Video	Visualization
	Motion capture system	Marker tracking in 3D dimensions

In this paper, we describe the platform in detail, providing data samples or references from our research team, and highlighting the effectiveness of the multisensor approach. Following common practice among engineers, our original techniques are presented in more detail than other existing techniques. The article is organized as follows: Section 2 deals with parametric signal detection techniques, Section 3 with motion tracking systems, Section 4 with laryngeal endoscopy, and Section 5 with other medical imaging systems. In Section 6, we emphasize the merits of the multisensor approach by giving two examples. First, the study of the articulatory realization of the feature [nasal] by combining fiberoptic and aerodynamic measurements. Second, the laryngeal realization of the segments according to their position in the word: glottographic and aerodynamic evidence of initial strengthening of glottal articulation.

2. Parametric Signal Detection Techniques

2.1. Acoustic and vibrosensory techniques

Acoustic and vibrosensory techniques are employed at LPP for the study of the behavior of the soft palate (velum). Analytical evaluation of nasality from acoustic signals is technically difficult, and its perceptual evaluation depends on many factors including the listener's native language. Various instrumentation techniques are available for the study of nasality, as described below. However, data do not always reflect real events when only one instrument is employed. The combined use of multiple techniques is often helpful for comparison and evaluation of data to get a fair picture of the reality. This is a necessary step to advance our understanding of how nasality functions in languages.

Two tools based on acoustic and vibrosensory techniques are available for laboratory studies and fieldwork. Figure 1 shows a piezoelectric vibration transducer and an acoustic nasal microphone. The piezoelectric transducer attached on the surface of the external nose records the amplitude of wall vibration of the external nose when nasal resonance occurs. The small dynamic microphone, called canalphone, inserted into one of the nostrils records air pressure variation in that nostril. The right panel of Figure 1 shows far-field microphone signal and nasal-wall vibration during the French word 'pomme' [pɔm] ('apple'), uttered by a native of French and a native of Chinese. The Chinese speaker produces the [ɔ] in [pɔm] as almost entirely nasalized, anticipating the final nasal consonant, whereas in native French speech, the anticipation in the vowel is minimal – as predicted by the structural fact that nasality is used distinctively for contrasting oral and nasal vowels. There are a few limitations for those techniques. The piezoelectric transducer also collects signals of voicing via vocal-tract wall vibration during oral closure. Furthermore, the signal intensity depends on low-frequency spectral components (i.e., first formant of vowels). Using those devices, Montagu (2007) compared the extent of anticipatory coarticulation in vowel-to-nasal sequences spoken by native French speakers and by learners of French.

¹ <http://www.sqlab.fr/evaSensFR.htm>

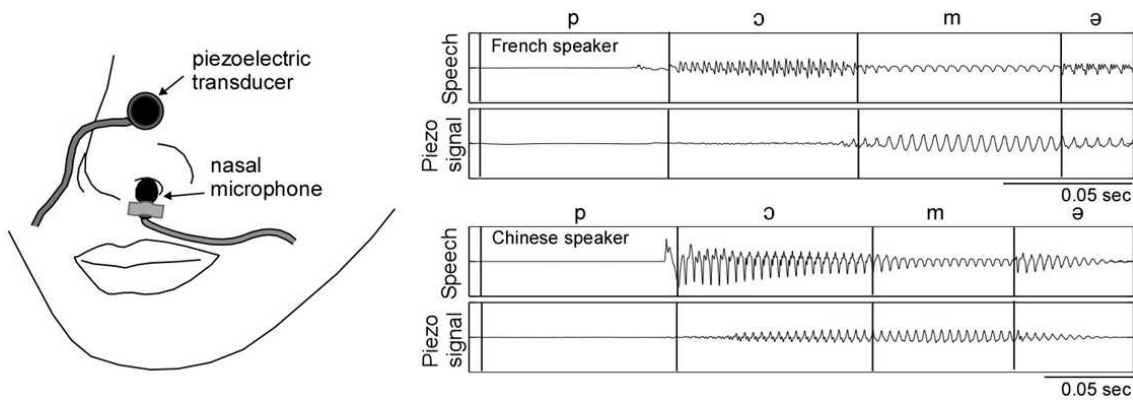


Figure 1. Piezoelectric vibration transducer and nasal microphone (left) with examples of the piezo transducer signals (right). The data correspond to the pronunciation of the French word “pomme” [pɔm] (apple, in English) uttered by a native of French (top) and by a native of Chinese (bottom).

2.2 Optoelectrical techniques

2.2 Optoelectrical techniques

The optoelectrical techniques refer to photoelectro-glottography (PGG) and Photonasography (PNG), which estimate the degree of opening of the glottis and the velopharyngeal port, respectively.

2.2.1 Photoelectro-glottography (PGG) and its recent improvement

Laryngeal adjustment for voiceless consonants has been visually monitored with nasopharyngeal fiberoscopy or estimated using photoelectro-glottography (PGG). PGG is a technique to monitor glottal adduction/abduction movement during speech. This technique employs a set of light source and photo-sensor, either unit being placed inside the body, and glottal aperture is monitored as the intensity of the light passed through the glottis (glottal transillumination). In early research, an intense light source was placed on the neck surface, and a photocell introduced into the pharynx (Fant & Sonesson 1962). Later, a nasopharyngeal fiberoptic was used as a light source to illuminate the larynx from above, and a photosensor (photodiode or phototransistor) is placed on the surface of the front neck below the cricoid cartilage (Ohala 1966). While PGG is a simple and efficient procedure, it has two known limitations: Fiberoptic is a mildly invasive procedure that must be conducted by a laryngologist, and speech utterances to be tested must exclude back vowels to avoid signal artefact from shadowing on transillumination by the tongue base and epiglottis.

A new PGG method is under development at LPP, which offers an alternative means for glottal transillumination by placing both light source and photosensor externally (Honda & Maeda 2008). This external lighting and sensing PGG (ePGG) is non-invasive and applicable to unrestricted speech materials. Figure 2 shows the principle of the ePGG technique with sample data. A high-power light emitting diode (LED) is used as a light source, which is placed on the surface of the neck tissue between the hyoid bone and thyroid cartilage to diffusely illuminate the wall of the hypopharyngeal cavity. The incident light passed through the open glottis casting on the hypoglottal or tracheal wall is detected with a photodiode placed on the surface of the front neck to feed to a high-gain amplifier.

The left panel of Figure 2 illustrates ePGG data for the Japanese word [hakusi] (‘blank paper’), which has three phonemically voiceless consonants [h], [k], and [s]². The glottis opens for [h] with maintained vocal-fold vibration and closes toward the end of [a]. In the next syllables, the glottis shows a single large opening according to the devoicing rule of high vowels in Tokyo dialect: two vowels /i/ and /u/ are devoiced when surrounded by voiceless consonants, e.g., [k], [s], and [t] in this utterance. This word is known as one of the irregular cases where vowel [a] can be devoiced, when the word is followed by a voiced segment, for example. Such a unique variation of concatenated syllable devoicing can be observed with ePGG.

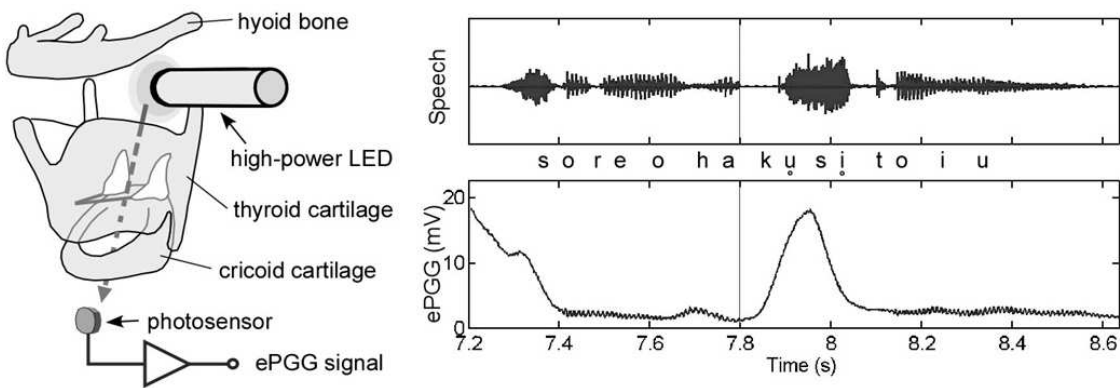


Figure 2. External lighting and sensing photoglottography (ePGG) system (left) with sample data for vowel devoicing in Japanese (right). The ePGG data is shown for a Japanese sentence including the word “hakushi,” where the final two syllables are devoiced.

2.2.2 Photonography (PNG)

Photonography (PNG) evaluates the opening of the velopharyngeal port, and the principle is the same as PGG. It consists of a light source placed above the velopharyngeal port and a photosensor located below (or vice versa). The velopharyngeal aperture is evaluated as the intensity of the light passed through the port (velar transillumination). Two systems are available at LPP (Amelot et al. 2006). First, a transparent flexible tube equipped with two components (a light source and a photosensor) is introduced into one of the nostrils (Ohala 1971). A miniature tungsten light bulb illuminates the cavity above the velum, and a small photosensor captures transillumination via the open port. Second, a fiberscope is used to illuminate the nasal cavity, and a separate tube with a photosensor placed in the pharynx detects velar transillumination. The method with fiberscopy permits video-recording of velar height (detection of the highest point on the nasal side of the velum).

Figure 3 illustrates PNG with fiberscopy and sample data indicating velopharyngeal port opening during a French nasal vowel /ã/ in the sequence /dadãda/ spoken by a native French speaker. The transillumination signal begins to rise at the nasal vowel onset. There is no anticipatory velar lowering during the consonant /d/ preceding the nasal vowel (region 1 on the figure). (Nasalization of the release of /d/ would lead to the identification of the sequence /danãda/ instead of /dadãda/.) Then the port starts to open slightly a few milliseconds after the vowel onset (region 2). It is maximally open in the second half of the nasal vowel (region 3). The port opening ends some time after the onset of the following consonant /d/. There is carry-over (region 4) and the first part of the voiced consonant /d/ is nasalized, but this nasalization is generally not perceived. PNG allows a rather straightforward interpretation in terms of opening/closing of the velopharyngeal port, although it is a mildly invasive procedure to be conducted by a medical person and it is less applicable in field work.

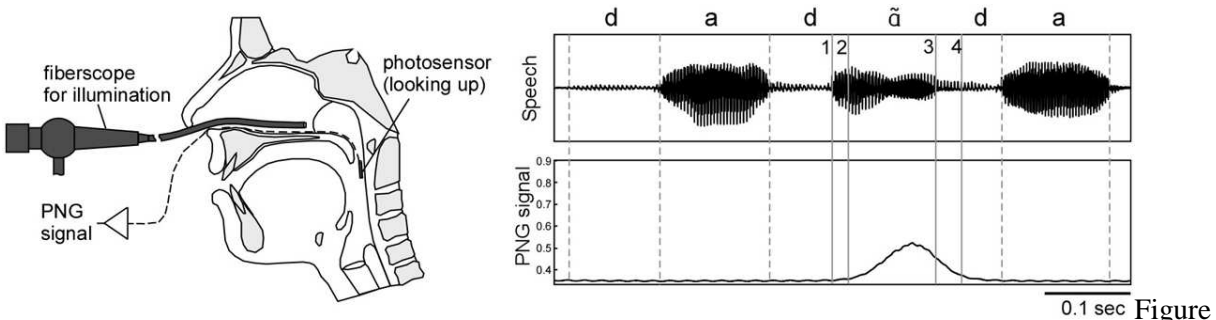


Figure 3. Photonography system (left) with data samples (right). This version of PNG employs a fiberscope for illumination and a phototransistor. The data shows the sequence /dadãda/.

2.3 Electrical techniques

2.3.1. Electroglottography (EGG)

The electroglottography (EGG) senses vibratory behavior of the vocal folds. Two electrodes are placed on both sides of the larynx at the level of the vocal folds. The amplitude of the high-

frequency signal circulating between the two electrodes is proportional to the area of contact of the vocal folds during phonation. This device was applied to study lexical tones, contributing to a debate in Vietnamese tonology by showing that final stops are not glottalized in Vietnamese (Michaud 2004) and refining hypotheses on tonogenesis that had been originally formulated on the basis of expert listening alone (Mazaudon and Michaud 2008).

2.3.2. Electropalatography (EPG)

Electropalatography is a custom instrument designed for monitoring tongue articulation (e.g., Gibbon & Nicolaidis 1999). The speaker is fitted with an artificial palate, covered with a number of electrodes and connected to a PC-based recording system. When the tongue contacts with the electrodes located at the level of the hard palate or teeth, a signal is sent to the computer in real time. We currently use the Reading EPG system commercialized by Articulate Instruments³. EPG allows observation of linguopalatal contact in time and space, and it has often been used for examining coarticulation. Fougeron (2001) used EPG to investigate articulatory properties of initial segments in several prosodic constituents in French; the results confirm that there is a larger contact for consonants in initial position than in other positions. Based on the temporal analysis of articulatory closures using EPG, Ridouane (2007) showed that Berber voiceless geminates in utterance-initial positions are produced with a longer closure duration than singletons, even though these durational differences are not perceptible.

2.3.3 Pneumotachography with shell-type protection mask

Airflow is important to speech production, particularly for the generation of voiced sounds and turbulent voiceless sounds. The airflow rate (liters/s) is often measured using a rigid face mask with a vent filled by a layer of stainless-steel fine mesh that gives small airflow resistance. This type of device is called a wire-screen pneumotachograph. The major inconvenience is that the rigid mask creates an acoustic cavity in front of the mouth and nostril openings. Its acoustic effect as a lowpass filter is so severe that the recorded audio signals cannot be used for speech analysis. Rothenberg (1973) designed a circumferentially vented pneumotachograph mask in order to reduce the spectral distortion due to the mask. LPP is equipped with the EVA2TM station⁴, for measuring oral and nasal airflow and pressure.

Our new pneumotachography developed at LPP uses a mask made of synthetic fibers, instead of conventional rigid types. This mask is acoustically almost transparent, and thus the radiated speech sound through the mask is almost free from acoustic distortion. In this preliminary stage of the development, we use shell-type protection masks commonly available in the market, which are designed to have relatively small airflow resistance in order to ensure the user's comfort. The mask supplies a small resistance, as is necessary to measure airflow without perturbing sound propagation. The mask itself is disposable and can be used in a hospital for clinical evaluation as well as in a laboratory with no risk of infection.

Figure 4 shows the airflow mask and sample data. The left panel illustrates how the mask is fixed on the subject's face. A microphone is placed in front of the mask for sound recording. The center panel shows an example of speech and airflow data for the Japanese word [asa] ('morning'). The airflow peaks indicated by the arrows at vowel-fricative and fricative-vowel boundary regions (/as/ and /sa/) occur at a point in time when the cross-sectional area of the glottis and that of the oral constriction become equal. The right panel compares the spectrum of vowel [a] recorded through the conventional rigid mask (top) and through our shell-type protection mask (bottom). It is evident that only the weakened first and second formants are visible in the top part because of filtering, whereas the vowel formants are complete in the bottom. The airflow measurements using our shell-type protection mask give comparable results with those from the rigid mask.

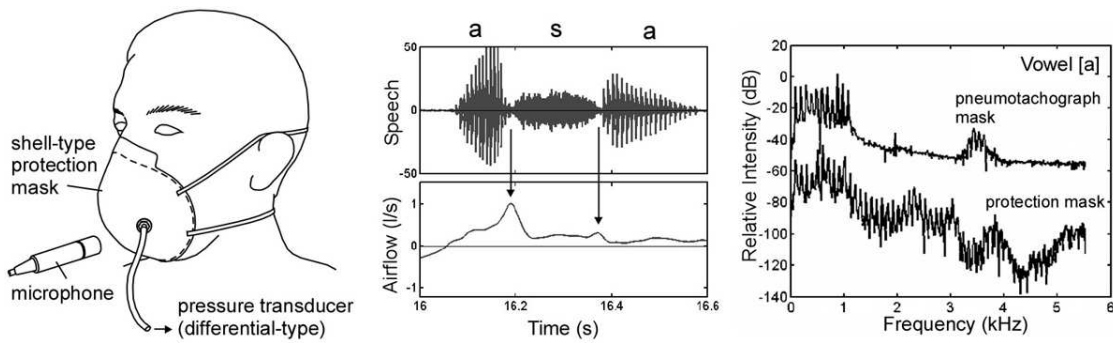


Figure 4. Shell-type protection mask for airflow measurement (left), sample data (mid), and spectral comparison of data from a conventional rigid pneumotachograph mask and from the protection mask (right).

3. Motion Tracking Systems

Electromagnetic motion tracking systems, such as Electromagnetic Articulography (EMA), allow for the tracking of articulatory movements of the organs within the oral cavity. A number of small sensor coils are glued midsagittally inside the mouth on the tongue and jaw, and externally on the lips or on the face. Transmitters that are fixed on a helmet produce alternating magnetic fields at different frequencies. The distances of each sensor from the transmitters are calculated from the strength of detected signals.

LPP is not equipped with EMA. For studying the movements of the lips and the face, we use the optical marker tracking system (Qualysis motion capture system). The system allows us to explore the visual cues while speaking. A number of passive (light-reflection) markers are positioned on the speaker's face and lips. Four video cameras with infrared (IR) light source are placed around the speaker and capture those markers at four different angles. The system calculates the 3D-coordinates of the markers at a rate of 100 fps. Figure 5 shows IR marker placements with sample data. It has been mainly used at LPP to study lip movements during speech, as a function of the position of the phonemes (Fougeron et al. 2010) and their deviation during emotional speech (Georgeton 2009).

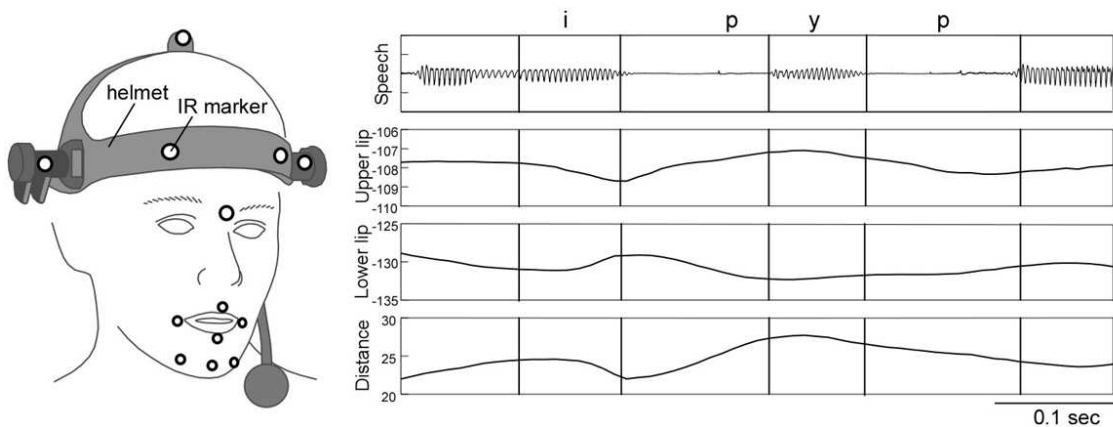


Figure 5. Optical motion tracking system showing markers (left) and sample data for lip movement during [ipyp] (right). The markers on the nose and on the helmet are used as references. The trajectories of the markers on the upper and lower lips are shown together with the Euclidean distance between the two. 4. Laryngeal Endoscopy and High-speed Imaging

4.1 Endoscopy and fiberscopy

For observing the larynx, direct visualization with endoscopes is used at LPP. Two types of endoscopes are used. First, a laryngeal rigid endoscope is an optical system with a shaft diameter of 8 to 10 mm and a prism that reflects the images. It allows excellent visualization of the vocal folds, while it is limited for observing front vowels (such as /i/ or /e/). Second, a flexible endoscope is a thin flexible bundle of 20-40 cm long with 3 to 4 mm in diameter with an optical light guide. It has the

advantage of permitting natural speech articulation; the velopharyngeal and laryngeal movements can be observed in continuous speech. The limiting factor is that it constitutes an impediment to the movements of the tongue and epiglottis in vowels and consonants with pharyngeal constriction. Several studies have been conducted in LPP using this technique. Ridouane (2008) examined laryngeal configuration during voiceless and vowel-less words in Tashlhiyt Berber (e.g., /tkkstst/ ‘you took it off’), showing that these words only contain voiceless obstruents with no intervening voiced elements.

4.2 High-speed endoscopy

Dysperiodicities associated with many voice pathologies makes videostroboscopy inappropriate, since stroboscopy requires periodic fundamental frequency to be synchronized with stroboscope light pulses. One way to circumvent this problem is to combine high-speed imaging techniques and endoscopy. With high-speed imaging, a single vibratory cycle is described with a sequence of several images. High-speed endoscopy provides direct images of the vocal folds in vibration to explore voice quality in normal and dysphonic voices. Moreover, it allows visualization of the supraglottic region during phonation with wave-like mucosal movements. The imaging rate can vary from 500 to 4000 fps. High-speed imaging data is analyzed with the software provided by the manufacturer⁵. Endoscopic visualization of the glottis can be coupled with EGG or ePGG to analyse the timing and degree of glottal opening during voiced and voiceless consonants in sequence.

One of our attempts was to compare the laryngeal behaviour in whispered speech at the levels of the glottis and the supraglottic region using high-speed imaging (Crevier-Buchman et al. 2009). Figure 6 illustrates the endoscope placement, acoustic signal, and the corresponding videokymographic¹⁰ data for the sequences /ivi/ and /ifi/ produced in whispering. In the kymograms (i.e., successive linear image sampling at a selected region), the glottis is more open for the voiceless consonant /f/ than for the voiced one /v/. The difference in width between /f/ and /v/ stresses that voicing features can be preserved in whispered utterances. High-speed imaging reveals laryngeal articulatory gestures and glottal configurations in whispered speech, which could tell us more about how speakers compensate for the lack of phonetic voicing information.

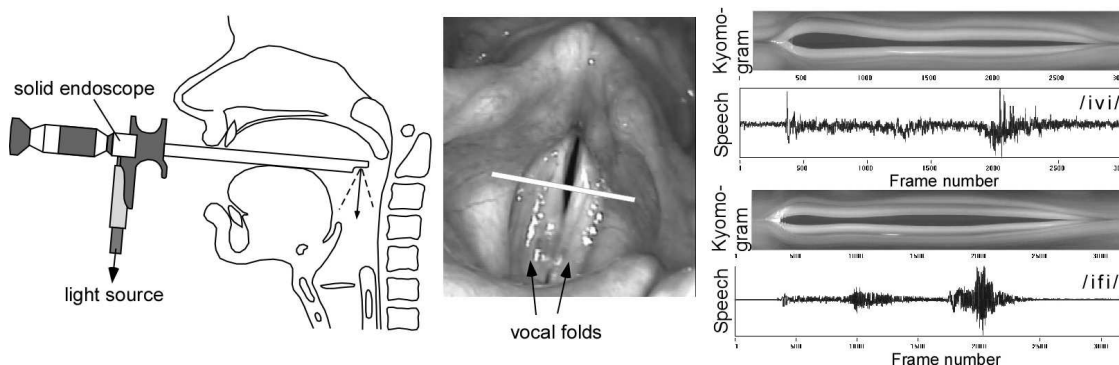


Figure 6. Solid endoscope (left), vocal-fold image (mid), and kymographic data during production of /ivi/ and /ifi/ in whispering. The vocal-fold image indicates the line for kymographic analysis of high-speed endoscopic images, which were sampled at 4000 fps.

5. Other Medical Imaging Systems

5.1 X-ray cinematography

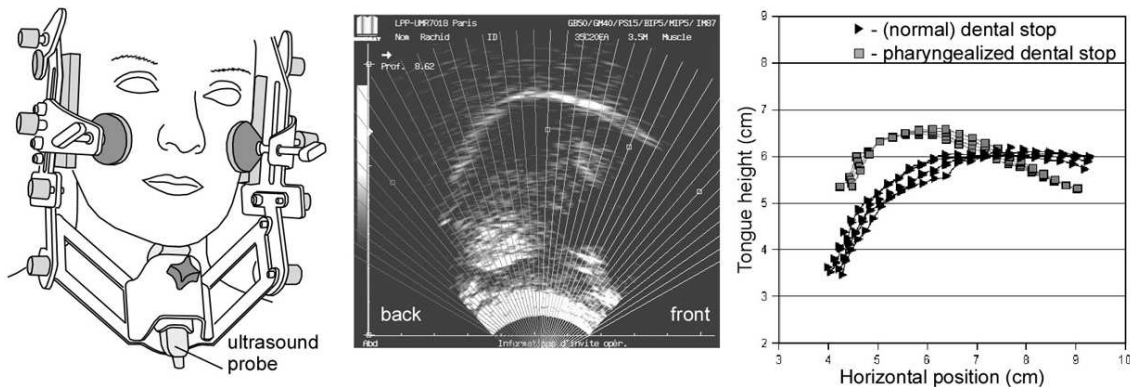
X-ray photography was used to examine shapes of the vocal tract for the first time in the 1920s. A number of studies on different languages (Arabic, Bulgarian, French, English, Japanese, Russian, etc.) have been conducted since then, and vocal-tract models defining the vocal tract shape have been developed using X-ray data. The X-ray images show clear outlines of rigid structures and less obvious boundaries of the soft tissue. The outline of the tongue can be enhanced by applying liquid contrast media on the surface. The data from the lateral X-ray cinematography with labial filming during running speech was used to develop Maeda’s anthropologic articulatory model (Maeda 1979). Many of the X-ray data for speech research in France are provided by the Phonetics Laboratory of Strasbourg University.

The X-ray microbeam system was a custom motion tracking instrument developed for phonetic studies, permitting accurate data acquisition, very low radiation dosage, and automatic marker tracking

at a maximum rate of 160 samples per second. Effort was made to construct English and Japanese speech production databases at the University of Wisconsin. Magnetic position-tracking systems were developed to emulate the X-ray microbeam system to be available in laboratories, and its three-dimensional (3D) version shows improvement in accuracy and rate of data acquisition.

5.2 Ultrasonography (USG)

The ultrasonography (USG, also referred to as echography) is a diagnostic imaging technique to obtain cross-sectional motion images of the soft tissues, which has been applied to monitoring tongue movements during speech. An ultrasonography system consists of a hand-held ultrasound probe (combined transmitter-receiver unit) and image processor. The probe is placed on the skin below the tongue to image the tongue surface in the sagittal or coronal plane. In radial B-mode (2D) scan, fan-shaped images are obtained at a frame rate of up to 100 fps. In the midsagittal scan, as seen in the mid panel of Figure 7, the image field is limited due to the surrounding bony structures, and the tongue tip and tongue base are out of view. Recently the systems have become compact enough to be used in phonetic laboratories or in field studies, while there remain a few technical problems for analytical work. Misorientation of the probe results in unnatural midsagittal images because the scan plane crosses the midline groove of the tongue obliquely. The hand-held probes change in position due to jaw opening in speech, and thus the image coordinate system alters relative to the head. The outline of the hard palate must be obtained using a position tracking system to examine tongue position relative to the palate. In order to solve those problems, many researchers have made probe-supporting devices, probe-position monitoring systems, or combined both. We employ at LPP a compact scanner and a headset for head-probe stabilization developed by Articulate Instrument (see Figure 7, left panel). When the video acquisition is made via the Advance Articulate Assistant software⁶, EPG data can be acquired simultaneously to monitor tongue contact with the palate. Using the combined technique, Fougeron & Ridouane (2008) provide evidence that schwa-like elements visible in the acoustic record in Tashlhiyt Berber during consonantal sequences like /t-gnu/ (she sewed) are transitional vocoids, and there is no specific lingual articulatory gesture associated with them. Ridouane (2009) studied pharyngealization and pharyngeals in Tashlhiyt Berber. The right panel of Figure 7 shows that during the pharyngealized stops, the tongue base is raised and retracted, while the anterior tongue is lowered.



7. Headset for stabilizing a probe for ultrasonography (left), sample image of the tongue surface (mid), and tongue surface contours of pharyngealized /d/ (grey squares) and non pharyngealized /d/ (filled triangles) in the context of /a/.

5.3 Magnetic Resonance Imaging (MRI)

Magnetic resonance imaging (MRI) is a medical imaging technique that excels at 2D and 3D imaging of soft tissues. Image brightness reflects the density of the proton (hydrogen nucleus) that is populated in water and lipid. Excitation and relaxation of the protons in a high homogeneous magnetic field in response to radio-frequency (RF) pulses is detected for image reconstruction, while varying gradient magnetic fields determines the position of the scan planes. Possible problems of MRI include speaker's supine body posture during scans, lack of dental images necessitating additional air-teeth boundary detection, and slow scan time. MRI has been used for static imaging (Toda et al. 2003 for fricatives in several languages), while motion imaging setups with synchronized (stroboscopic) techniques have been applied to visualize articulatory and glottal movements (Kim et al. 2005, for the three-way phonation of Korean consonants). On the other hand, real-time MRI has become available,

and the frame rate in real-time imaging depends on system performance and image size. Currently, an approximate maximum rate of 5 fps in 1.5 T and 10 fps in 3 T is attainable.

6. Examples of studies based on multisensor platform approach

This section provides two examples illustrating the effectiveness of the multisensor platform. We will first present the case of the realization of the feature [nasal] on data simultaneously acquired for the velar height with fiberscopy and the degree of velopharyngeal opening estimated by nasal airflow measurement. The second example shows simultaneous recordings of ePPG and aerodynamic data that provide new insights into laryngeal configuration during speech.

6.1 Multisensor analysis of nasalization

Nasality can be described by different aspects, the implication of the muscles, velar movements, velopharyngeal port opening, aerodynamic and acoustic output, and perceptual correlates. Depending on the instrumentation techniques, the conclusions on anticipatory and carry-over nasalization could differ. First, electromyographic (EMG), fiberscopic, and EMA studies on the realization of the nasal consonants tend to favor the conclusion that there is more anticipatory nasality than carry-over⁷. Second, other studies, mainly based on aerodynamic measurements, lead to the inverse conclusion: Anticipatory nasal coarticulation is not as pervasive as carry-over nasalization (Basset et al. 2001, Amelot 2004, Delvaux et al. 2007).

In fact, there is often a poor correlation across measurements of different types. First, there are no direct relationships between the degree of suppression of levator palatini muscle activity and velar height as observed by fiberscopy (Ushijima & Hirose 1974). While there is no carry-over suppression of LP activity from a preceding syllable-final nasal consonant as described above (such as in the case of a /seN'ee/ sequence in Japanese), fiberscopic studies actually show a low velar height for the vowel segments following the moraic N. In this case, carry-over effects on velar height after /N/ are considered to be due to some inherent mechanical response characteristic of the velum. Second, there is an expected temporal delay between velar movements and velar opening (Lübker & Moll 1965), but the delay is not constant. Third, the relationships between velar height and nasal airflow are nonlinear. The presence of a steady airflow implies that the velopharyngeal port is open but its absence does not signify that the velum is in a high position.

Figure 8 shows the time course of velar height as estimated from fiberscopic films and the nasal airflow during the sequence /tatēta/. The reference for estimating the elevation of the velum from the fiberscopic films corresponds to a frame where the velum is at the lowest position (Amelot 2010). Manual measurements of the distance between the reference and the observed velar height are made at 10 ms intervals and plotted with interpolation between the points (Figure 8: bottom). Nasal airflow was measured using the EVA2 with a pressure transducer attached to a tube inserted in one of the nostrils and secured by rubber plug.

As can be seen from this figure, the velum starts to lower well before the onset of the nasal vowel, i.e. velar lowering is anticipated; it is maximal in the vicinity of the midpoint of the nasal vowel (as expected), whereas the peak of nasal airflow appears well after the nasalized vowel, during the initial part of the following oral consonant. Figure 8 also shows that there is no correspondence between the peak of nasal airflow and that of velar height. For vowels, velopharyngeal port opening does not imply an amount of nasal airflow proportional to the degree of the opening. This is because (i) airflow from the lungs is primarily determined by the oscillating glottal state and (ii) how this flow separates between the nasal and oral tract depends not only on the opening of the velopharyngeal port but also on the state of the oral tract. Lübker & Moll (1965) measured oral and nasal airflow while simultaneously recording the articulatory positions of the tongue and velum in a single normal speaker. They reported that anticipatory nasal airflow increased whereas the size of the velopharyngeal orifice remained constant during the production of a vowel-nasal consonant sequence. Lübker and Moll attributed this finding to an increase in oral cavity constriction. The delayed peak in nasal airflow in Figure 8 can therefore be interpreted as resulting from the increase in airflow resistance in the oral cavity during the closing gesture for the realization of the following obstruent consonant /t/.

Furthermore, bursts of positive nasal airflow can occur due to the pumping action of the closed and rising velum (Lübker & Moll 1965), and bursts of negative nasal airflow due to the sucking of the air by the closed and lowering velum (Benguerel 1974). Airflow measurements have the great

advantage of being non-invasive but the interpretation of the data in terms of nasality demands careful analysis, and the multisensor instrumentation appears to be a great help for the correct interpretation of aerodynamic events in their full complexity.

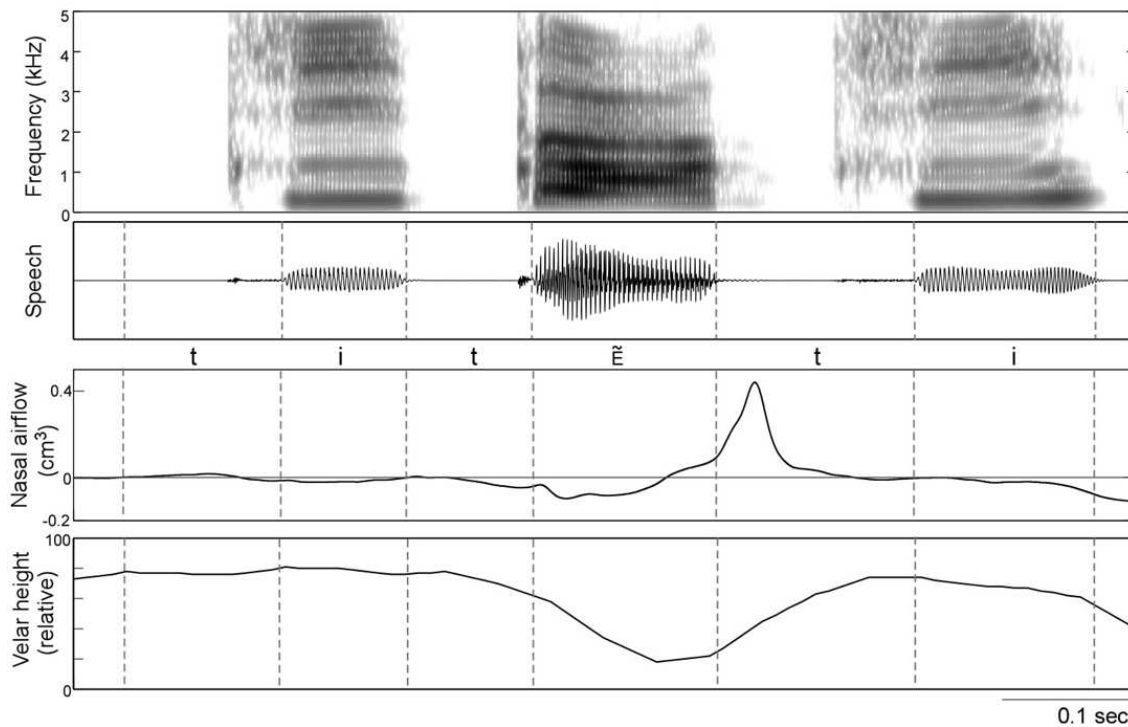


Figure 8. Measurements of velar height with fiberoptic and nasal airflow during production of /tatɛta/ by a native French speaker. From top to bottom, spectrogram, speech signal, nasal airflow, and velar height.

6.2 Glottographic and aerodynamic observation for initial strengthening of glottal articulation

Previous fiberoptic studies on glottal adjustments demonstrated an obvious tendency for the glottis to open wider and longer for voiceless word-initial consonants than for word-medial consonants (Sawashima & Hirose 1981), suggesting evidence on the glottal level of initial strengthening (Fougeron & Keating 1997). Domain-initial strengthening seems to concern all articulators, but few data focused on glottal behavior in continuous speech⁸.

We have conducted a combined experiment to obtain ePGG (see Section 2.2), airflow (see Section 2.4), and intraoral pressure (P_{io}) signals to examine the nature of initial glottal gestures. Here, the P_{io} was measured using a tube from a nostril to the pharynx, which was connected to a pressure transducer placed externally.

We have attempted a preliminary study to examine the nature of word-initial glottal widening by conducting two experiments: 1) standard PGG with videofiberscopy, and 2) combined recording of ePGG, airflow, and intraoral air pressure (P_{io}). The sample data are shown in figure 9.

The left figure shows data from the PGG experiment with a Japanese utterance [sore-o-ke:-ke:-to-iu] ('It is called injudiciousness'), an utterance taken up from Sawashima and Hirose (1981). The word [ke:ke:] is placed at the second phrase in the sentence and produced nearly as two contiguous repetitions of the same syllable. As shown in fiberoptic images and PGG data, the glottis opens wider for the first [k] (k1) than for the second [k] (k2), in accordance with Sawashima and Hirose (1981). There is a general tendency that the longer the voiceless segment, the larger the glottal opening, whereas this example indicates that another factor may be involved in determining the magnitude of glottal opening, since the two [k]s have about the same duration.

The right figure shows a sample of data from the combined recording using a Japanese sentence [i:-yo:ke:-ke:se:-suru] ('one formulates better juvenile forms'), where [yo:ke:] ('juvenile form') and [ke:se: suru] ('to formulate') are separated by a phrase boundary. In this sentence, the first [k] (k1) is in the second syllable of the word [yo:ke:], and the second [k] (k2) is at the initial position of the second phrase [ke:se: suru]. It is shown, in contrast to the left one, that the glottis opens wider for the second [k], and this tendency has been observed in other test utterances with repetitions of the same

syllable. Moreover, aerodynamic parameters show consistent differences: a higher airflow rate after release, and slightly reduced intraoral pressure in [k] at the phrase-initial position, regardless of the order in the syllable sequence. The higher airflow rate after the release may be related to a longer period of frication noise, while the intraoral pressure during the stop closure is lower at the initial position despite the higher fundamental frequency for the following vowel. This observation does not agree with generalized reinforcement of aerodynamic factors at initial position: the higher airflow is determined by both glottal and vocal-tract apertures. Rather, the initial glottal widening may be due to a certain physiological control to signal a phrase boundary, which may for instance be realized as a phrase-final mild glottalization followed by a rapid cessation of vocal-fold vibration for a devoiced stop segment.

The observation described above is a small step towards resolving the issue of initial glottal widening. The entire mechanism needs to be further investigated together with an effort for instrumental development to acquire more direct evidence. There can be little doubt that the multisensor-platform approach helps researchers approach causal mechanisms of phonetic phenomena.

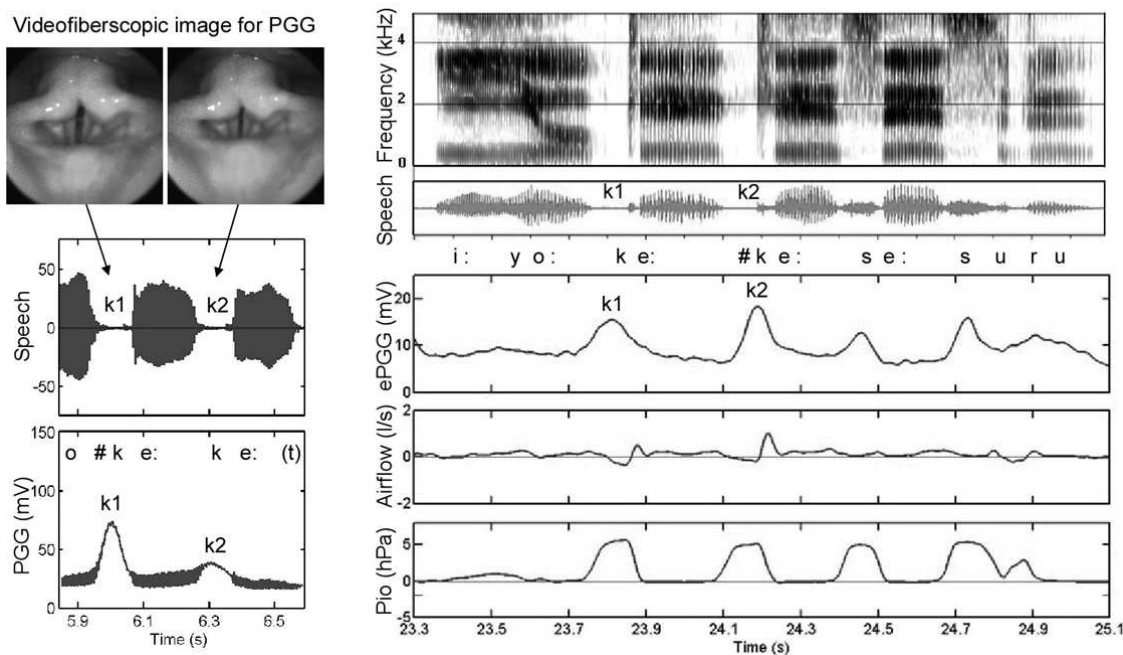


Figure 9.

Initial glottal widening observed by standard PGG in [#ke:ke:] (left) and by combined ePGG, airflow, and intraoral pressure measurement in [yo:ke: #ke:se:] (right). The mark # corresponds to the phrase boundary in the sentences.

7. Summary

The paper described instrumentation techniques, both conventional and new, that are currently used at our laboratory for investigating the behavior of the speech organs involved in speech production. The examples that were provided highlight the innovative value of the multisensory platform for phonetic studies. The first example combined fiberscopic and aerodynamic data for the study of nasality; the second combined photoglottography, pneumotachography and intraoral pressure measurements for the study of glottal articulation. Both illustrate the benefit of using several exploratory techniques in parallel, suggesting that data based on a single technique can be misleading.

Research issues in phonetics and phonology are too numerous to be listed in a few lines: they include patterns of synchronic variation and evolutionary paths (in particular prosodic influences over the realization of segments) and typological patterns of phoneme distributions and phonotactic constraints (in particular recurrent asymmetries within sound systems). Progress in analyses and models requires the study of physiological, articulatory, acoustic, aerodynamic and perceptual parameters. The multisensor-platform approach, which concerns laboratory speech and spontaneous speech alike, can advance our fundamental knowledge on speech. It can contribute to the establishment of panchronic laws of evolution (the panchronic program dates back to Haudricourt 1940; about recent developments, see Mazaudon and Michailovsky 2007), and open into new applications for the future.

Acknowledgements

The authors would like to express many thanks to the members, ex-members, and external colleagues of LPP, in particular to Cécile Fougeron, Alexis Michaud, Rachid Ridouane and Nicolas Audibert. This work is in part supported by the national grant (ESAPVI) from the French National Research Agency (ANR) and by special funds from the French National Centre for Scientific Research (CNRS) to LPP.

References

- Amelot, A. (2004) "Etude aérodynamique, fibroscopique, acoustique et perceptive des voyelles nasales du français." Thèse de doctorat, Université Paris III - Sorbonne Nouvelle, Paris, France.
- Amelot, A., Honda, K., Maeda, S., Crevier-Buchman, L., Basset, P. (2006) "Evaluation of a photonasograph." *Journal of the Acoustical Society of America*, 120, Issue 5, 3353.
- Amelot A. (2010) "Dispositifs d'imagerie pour l'observation de l'activité vélo-pharyngée." In A. Marchal and C. Cavé (eds.) *L'imagerie pour l'étude sur la parole*. (pp. 65-82). Paris: Hermès/Lavoisier.
- Basset, P., Amelot, A., Vaissière, J., Roubeau, B. (2001) "Nasal airflow in French spontaneous speech." *Journal of the International Phonetic Association*, 31, 87-100.
- Benguerel, A. P. (1974) "Nasal airflow patterns and velar coarticulation in French." In G. Fant (ed.) *Speech Wave Processing and Transmission*. (pp. 105-112). Stockholm: Almqvist & Wiksell.
- Crevier-Buchman, L., Vaissière, J., Henrich, N., Vincent, C., Hans, S., Brasnu, D. (2009) "Laryngeal behavior in whispered voice: A study using high speed imaging." *The Voice Foundation's 38th Annual Symposium*, Philadelphia, United States.
- Delvaux, V., Demolin, D., Harmegnies, B., Soquet, A. (2007) "The aerodynamics of nasalization in French." *Journal of Phonetics*, 36, 578-606.
- Dilley, L., Shattuck-Hufnagel, S., M. Ostendorf, M. (1996) "Glottalization of word-initial vowels as a function of prosodic structure." *Journal of Phonetics*, 24, 423-444.
- Fant, G., Sonesson, B. (1962) "Indirect studies of glottal cycles by synchronous inverse filtering and photo-electrical glottography." *Speech Transmission Laboratory, Quarterly Progress and Status Reports*, 3, 1-3.
- Fougeron, C., Keating, P. A. (1997) "Articulatory strengthening at edges of prosodic domains." *Journal of the Acoustical Society of America*, 101, 3728-3740.
- Fougeron, C. (2001) "Articulatory properties of initial segments in several prosodic constituents in French". *Journal of Phonetics*, 29, 109-135.
- Fougeron, C., Ridouane, R. (2008) "On the nature of schwa-like vocalic elements: Some Berber within clusters". In *Proceedings of The Eighth International Seminar on Speech Production (ISSP'08)*, 441-444.
- Fougeron, C., Georgeton, L., Amelot, A., Vincent, C. (2010) "Vowel articulation in absolute initial position of different prosodic constituents in French." *12th Conference on Laboratory Phonology (LabPhon 12)*, Albuquerque, United States.
- Georgeton, L. (2009) *Etude acoustique, perceptive et articulatoire (mouvements des lèvres) lors d'émotions simulées en français*, Master in Phonetics, Université Paris III - Sorbonne Nouvelle, Paris, France.
- Gibbon, F., Nicolaidis, K. (1999) "Palatography". In W. J. Hardcastle and N. Hewlett (eds.), *Coarticulation: Theory, Data and Techniques*. (pp. 229-244). Cambridge: CUP.
- Haudricourt, A.-G. (1940) "Méthode pour obtenir des lois concrètes en linguistique générale." *Bulletin de la Société de Linguistique de Paris*, 41, Issue 1, 70-74.
- Honda, K., Maeda, S. (2008) "Glottal-opening and airflow pattern during production of voiceless fricatives: A new non-invasive instrumentation." *Journal of the Acoustical Society of America*, 123, Issue 5, 3738.
- Kim, H., Honda, K., Maeda, S. (2005) "Stroboscopic-cine MRI study of the phasing between the tongue and the larynx in the Korean three-way phonation contrast." *Journal of Phonetics*, 33, 1-26.

- Lübker, J. & Moll, K. (1965) "Simultaneous oral-nasal air flow measurements and cinefluorographic observations during speech production," *Cleft Palate Journal*, 2, 257-272.
- Maeda, S. (1979) "An articulatory model of the tongue based on a statistical analysis." *Journal of the Acoustical Society of America*, 65, Issue S1, S22.
- Mazaudon, M. & Michailovsky, B. (2007) "La phonologie panchronique aujourd'hui: quelques repères." In J. Fernandez-Vest (ed), *Combats pour les langues du monde: hommage à Claude Hagège*. (pp. 351-362). Paris: L'Harmattan.
- Mazaudon, M., Michaud, A. (2008) "Tonal contrasts and initial consonants: A case study of Tamang, a 'missing link' in tonogenesis." *Phonetica*, 65, 231-256.
- Michaud, A. (2004) "Final consonants and glottalization: New perspectives from Hanoi Vietnamese." *Phonetica*, 61, 119-146.
- Montagu, J. (2007) *Étude acoustique et perceptive des voyelles nasales et nasalisées du français parisien*. PhD. Thesis in Phonetics. University Sorbonne Nouvelle.
- Ohala, J. (1966) "A new photo-electric glottograph." *UCLA Working Papers in Phonetics*, 4, 40-52.
- Ohala, J. (1971) "Monitoring soft palate movement in speech." *Journal of the Acoustical Society of America*, 50, Issue 1, 140.
- Ridouane, R. (2007) "Gemination in Tashlhiyt Berber: An acoustic and articulatory study." *Journal of the International Phonetic Association*, 37, 119-142.
- Ridouane, R. (2008) "Syllables without vowels: Phonetic and phonological evidence from Tashlhiyt Berber." *Phonology*, 25, 321-359.
- Ridouane, R. (2009) "Pharyngealization and pharyngeal Tashlhiyt in Berber: A descriptive and comparative study." *International Workshop on Pharyngeal and Pharyngealization*, United Kingdom.
- Rothenberg, M. (1973) "A new inverse-filtering technique for deriving the glottal air flow waveform during voicing." *Journal of the Acoustical Society of America*, 53, 1632-1645.
- Toda, M., Honda, K. (2003) "An MRI-based cross-linguistic study of sibilant fricatives." *Proceedings of the 6th International Seminar on Speech Production*, Sydney, 290-295.
- Sawashima, M., Hirose, H. (1981) "Abduction-adduction of the glottis in speech and voice production." In K. N. Stevens and M. Hirano (eds.) *Vocal Fold Physiology*. (pp. 329-346). Tokyo: University of Tokyo Press.
- Ushijima, T., Hirose H. (1974) "Electromyographic study of the velum during speech." *Journal of Phonetics*, 2, 315-326.

Notes

¹ Phonetics and Phonology Laboratory in Paris (<http://lpp.univ-paris3.fr/>). LPP is associated with University Sorbonne Nouvelle (University of Paris III) and Centre National de la Recherche Scientifique (National Center for Scientific Research), France.

² In Japanese, [s] before [i] is strongly palatalized and often described as [ç]. In this paper, [s] is used in the text and figures for convenience.

³ <http://www.articulateinstruments.com/>

⁴ <http://www.sqlab.fr/evaSensFR.htm>

⁵ Weinberger Camera commercialized by Vannier-Photelec and HRES Endocam Camera commercialized by Wolf.

⁶ <http://www.articulateinstruments.com/>

⁷ Many EMG studies demonstrate the suppression of the levator palatini activity (LP) for velar lowering. In a study of Japanese, Ushijima and Hirose (1974) found that there is no carry-over suppression of LP activity, suggesting no velar lowering for vowel segments following syllable-final nasals. Fiberscopic studies (Benguerel et al. 1975, Amelot 2004) lead to the same conclusion for French: more anticipatory nasality than carry-over.

⁸ See however Dilley et al. (1996) on initial glottalization, based on acoustic data.