

Investigating speech production

A review of some techniques

Rachid Ridouane
LPP (CNRS, Paris3)
rachid.ridouane@univ-paris3.fr

Thanks to technological advances, a wide variety of tools has been devised to measure and analyse speech production, with ever-increasing detail and accuracy. These techniques include, among others, X-ray, X-ray microbeam, Ultrasound, Magnetic Resonance Imaging (MRI), and Electromagnetic Articulography (EMA). A brief presentation of each one of these tools is presented below. Several criteria are relevant to evaluate the weaknesses and strengths of these techniques. Ideally, they should permit the recording of the dynamics of all the articulators with an accurate temporal and spatial resolution. They should not alter subjects' natural pronunciation, nor degrade the quality of the audio data recorded. And, importantly, these techniques should not involve any health hazard for subjects. These criteria (and others) are reported in Table 1 (see section 5), and provide a concise evaluation of each technique. At the end of this presentation, a detailed bibliography of the research studies conducted using these tools is provided.

1. X-ray¹

X-ray was used to examine the movements of the vocal tract for the first time in 1920's (see Russel 1928). Considerable research on different languages (Arabic, Bulgarian, French, English, Japanese, Russian, etc.) has been conducted since then, and most models defining the vocal tract shape were developed using X-ray images (see the bibliography below). Several X-ray images, such as the one presented in figure 1, as well as film samples can be viewed on the website.²

X-ray data were very useful to deepen our knowledge about the movements of the vocal tract, especially concerning the shape and the position of the tongue during vowels and the role of the pharynx during speech. However, exposure limitations to ionizing radiation make X-ray movies extremely difficult to obtain. Another limitation of X-ray films includes difficulty in accurately deducing the cross-sectional morphology from mid-sagittal profiles (Branderud et al. 1998, http://www.ling.su.se/staff/peter/Pb_Bli.html). The constrained time for acquisition is another severe limitation on the usefulness of X-ray films. Language consultants, for instance, cannot be filmed for more than 20 seconds, to be within a safe radiation dosage (0.1 mSv). Recent developments in digital X-ray technology, however, have allowed reducing some of these limitations (see Branderud et al. 1998)

¹ A very comprehensive presentation of how this technique works is consultable at <http://www.howstuffworks.com/x-ray.htm>

² See, for example, http://psyc.queensu.ca/~munhallk/05_database.htm, which provides a presentation of the Queen's University/ATR cineradiographic database (A brief presentation of this database as well as the IPS X-Ray Database is provided below). See also <http://www.ling.lu.se/persons/Sidney/coartdem/films.html>, for still pictures and X-ray moving sequences drawn from Bulgarian language.

Another approach to vocal tract measurement in speech was the development of X-ray microbeam (see, for example, Westbury 1994, for a detailed presentation of this technique, <http://www.medsch.wisc.edu/~milenkvc/pdf/ubdbman.pdf>). Less toxic than older X-ray imaging, this technique uses narrow X-ray beams to track gold pellets attached to the tongue, jaw and lips. These measurements have been very useful in determining the articulatory characteristics of anterior tongue movements, but their limitations do not enable the imaging of the tongue root and pharynx. Other limitations of this technique include the fact that it is rather expensive and cannot be portable for outside laboratory recordings. All these limitations, in addition to the development of EMA, have severely restricted its use in speech research.

1.1. Some X-ray databases

1.1.1. ATR videodisk

The X-Ray Film Database for Speech Research is a collaborative project by Dr. K.G. Munhall (Queen's University) and Drs. E. Vatikiotis-Bateson & Y. Tohkura (ATR Human Information Processing Laboratories, Kyoto, Japan). It was conceived to create a database that stores a collection of high-quality copies of the original X-ray films in a durable format. The aim is to make these images available to the speech research community and to develop techniques for automated digital processing of these images (see http://psyc.queensu.ca/~munhallk/05_database.htm). This database contains a series of X-ray movies of sideviews of vocal tracts in motion, together with the resulting sound. Specifically, the project offers 25 films totaling 55 min of X-ray footage compiled on a constant angular velocity (CAV) format videodisk. Twenty-four of these 25 films were made in 1974 under the direction of C. Rochette (Université Laval). K. Stevens and J. Perkell (MIT) contributed the remaining film in 1962. The MIT film, which was shot at 45 i/s, shows the entire vocal tract with the lips visible. The films from Laval Université, which were filmed at 50 i/s, do not show the lower pharynx or larynx; however, the hyoid bone is visible and the lips and velum are clear in most of the 14 films. Contrary to MIT film, no enhancing substance was used in these films. The subjects recorded are 9 native speakers of Canadian French and 5 native speakers of Canadian English reading phonetically contrastive sentences. All of the films were accompanied by separate audio recording. A DAT of these original audio tracks was also produced. In order to provide an acoustic reference for reviewing the X-ray images, the audio recordings were synchronized with X-ray images and recorded on the videodisk. An illustration of the X-ray images found in the MIT film is presented in figure 1.



Figure 1. An X-ray image during the recording of a native speaker of Canadian English producing the sentence “Why did Ken set the soggy net on top of his deck”. For a video sample of this recording: http://psyc.queensu.ca/~munhallk/05_database.htm

This X-ray film database is available to researchers at no cost, with a limitation of one per institution. The DAT recordings of the original audio tracks and videotape copies of the disk are also available. A small fee is required to cover the cost of materials and postage. For more information, contact Dr. Munhall: munhallk@psyc.queensu.ca.

1.1.2. The IPS X-ray database, Strasbourg

The Institut de Phonétique de Strasbourg has gathered since 1950's more than 50 X-ray recordings, including data from a large variety of languages. Researchers from the Institut de Phonétique de Strasbourg, with the collaboration of the Institut de la Communication Parlée de Grenoble, have recently undertaken the task of creating a database that stores this collection of the original x-ray films in high-quality copies (cf. Arnal et al. 2000). The aim is to store these films in a durable format and make them available for speech research community.

This database 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. An illustration of the types of images obtained out of these data is displayed in figure (2). Figure (2a) shows the cineradiographic image of the vocal tract during the production of a nasal consonant, and figure (2b) shows the hand-drawn sagittal contour of the same image.

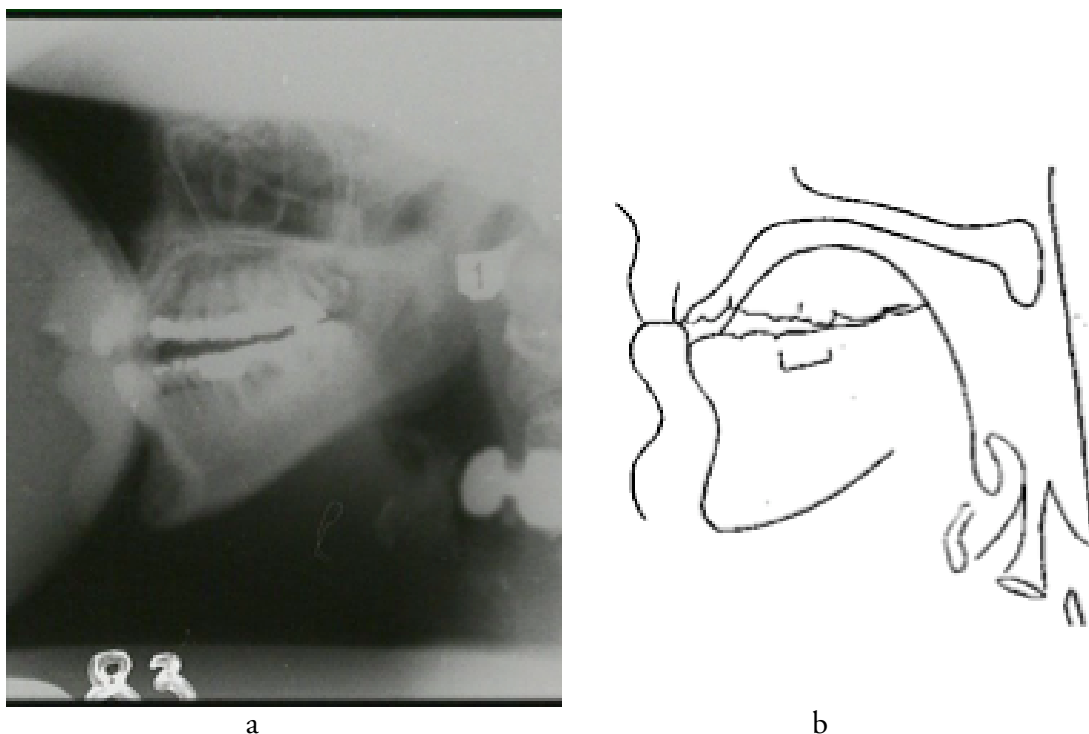


Figure 2. The cineradiographic image of the vocal tract during the production of /m/ in /mi/ (2a). The sagittal contour of the same image is given in (2b). (From Arnal et al. 2000).

All files are phonetically labeled and stored on CD ROMs. The films are available to researchers at no cost. For more information, contact the Institut de Phonétique de Strasbourg, the owner of the Database at <http://misha1.u-strasbg.fr/IPS/>.

2. Ultrasound

Ultrasound is a technique that uses sound waves to collect real-time data, showing tongue surface motion during speech. The application of this technique in speech articulation research was pioneered by Stone & colleagues (Stone et al. 1983, Stone & Davis 1995, see below for additional references). In ultrasound imaging, a transducer emits a beam of ultra high-frequency sound (5-40 MHz) that is directed through the lingual soft tissue. Some of the sound waves are reflected back when they reach the tongue-air boundary on the superior surface of the tongue, and return to the same transducer. A computer is used for taking the signals from the transducer, and converting them to an image. In the constructed image, the surface of the tongue is (more or less) visible as bright line on a black background. As can be seen in Figure 2, ultrasound captures an almost complete view of the tongue.

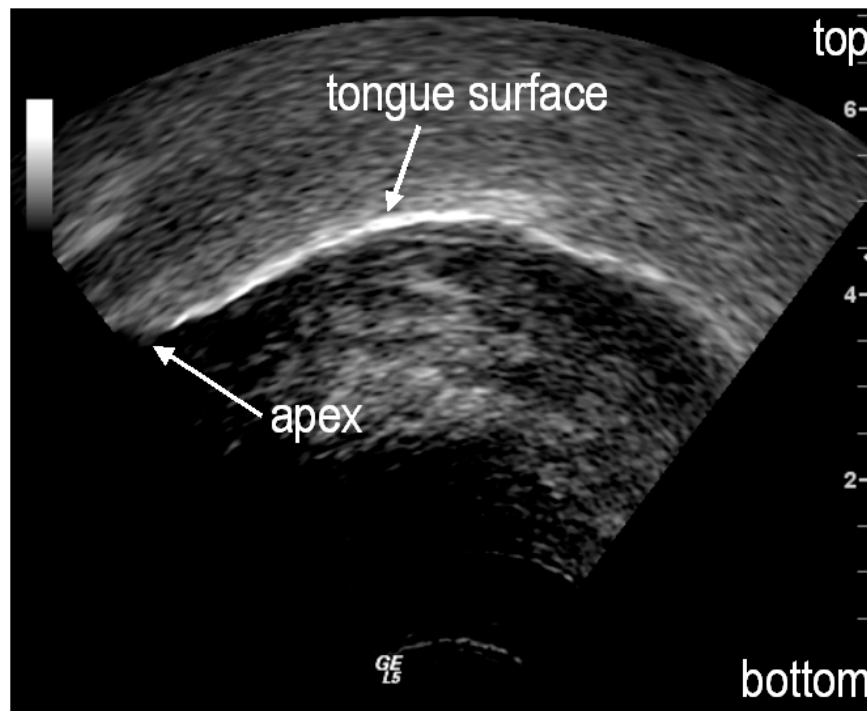


Figure 2. An ultrasound image of the tongue during the production of /a/ by a French male native speaker. (From Aron et al. 2006, <http://aspi.loria.fr/Save/aron.pdf>)

A central concern of ultrasound research in speech is related to the lack of absolute spatial reference in the signal. To determine the exact nature of the constriction, the location of the tongue within the vocal tract must be measured. One way to do this is to immobilize the head and the probe using a specially designed system (for e.g. the Head And Transducer Support HATS. For more information about the HATS system, see Stone & Davis 1995). Participants are seated in the HATS system, which is adjusted to fit the speaker's head comfortably. The transducer is placed under the speaker's chin and adjusted until the crispest image of the tongue is obtained. Tongue shapes are measured using EdgeTrak, an automatic system for the extraction and tracking of tongue contours (Akgul et al. 2000, Li et al. 2003). Few points on the tongue image are chosen, and then EdgeTrak uses an active contour model to determine the location of the tongue edge in the image (see Whalen et al. 2005 on possible limitations of this system). Though this contour tracking system is accurate for speech research, it still has some weaknesses that should be mentioned. For instance, the ultrasound images are quite noisy and there are some unrelated high contrast edges in the images which make the

gradient information insufficient to extract edges of interest. Moreover, the tongue surface might be interrupted in places.

Recently, a different system has been developed in Haskins Laboratory that takes advantage of ultrasound without requiring immobilization: the Haskins Optically Corrected Ultrasound System (HOCUS). This system incorporates both ultrasound imaging of the tongue and optical tracking of the probe relative to the head. The optical system (Optotrak) tracks the location of external structures, on the head and on an ultrasound sound transceiver, in 3-dimensional space using infrared emitting diodes (IREDs). The head, probe, and jaw are allowed to move, but their motion is tracked and can therefore be used to correct the tongue measurement to a head-based coordinate frame. The probe may either be held to a fixed orientation during running speech or moved to different orientations during sustained phonation. The use of one or the other method enables to obtain cross-sectional data (probe fixed) or multiple cross-sections for three-dimensional reconstruction (probe moved). Another strength of this system is that optical tracking can also be obtained for different visible structures (such as the lips and jaw) and thus provides complete measurements of the vocal tract during fairly unconstrained speech. The hard palate is obtained, as is commonly used in ultrasound recordings, by asking the participant to take a mouthful of water and force it up into contact with the hard palate. This ensures visualization of the palate because sound waves are no longer impeded by the surface of the tongue. The extracted boundary of the palate can then be inserted into every frame of the speech trials since the head is tracked during the experiment.

Data collection using ultrasound is suitable to the imaging of the tongue and offers many advantages, compared to other tools. It is non-toxic and does not involve any known health hazard; it uses high frequency sound which poses no danger to the subjects. It also offers an accurate temporal and spatial resolution. It is relatively inexpensive, can be portable, and subjects recorded are more comfortable. Compared to MRI (see below), it has other benefits (mainly shorter acquisition time and upright position), though it is unable to systematically image the tongue tip and gives less detailed tongue surface data. Notice, however, that the lack of the tongue-tip image can still be complimented using other systems; the tongue-tip image can, for example, be complimented by its position detection using an electro-magnetic position tracker (see Aron et al. (2006) <http://aspi.loria.fr/Save/aron.pdf>).

3. Magnetic Resonance Imaging (MRI)

Magnetic Resonance Imaging (MRI) is the only tool that can provide detailed 3D data of the entire vocal tract and tongue without any known harmful effects on the subject. The images have good signal to noise ratio, are amenable to computerized 3-D modelling, and provide excellent structural differentiation. In addition, the tract (airway) area and volume can be directly calculated. An illustration of MRI images of the vocal tract is provided in Figure 3. These figures provide MRI images of the vocal tract during the production of the two sibilant fricatives /s/ (3a) and /ʃ/ (3b) by an English speaker.

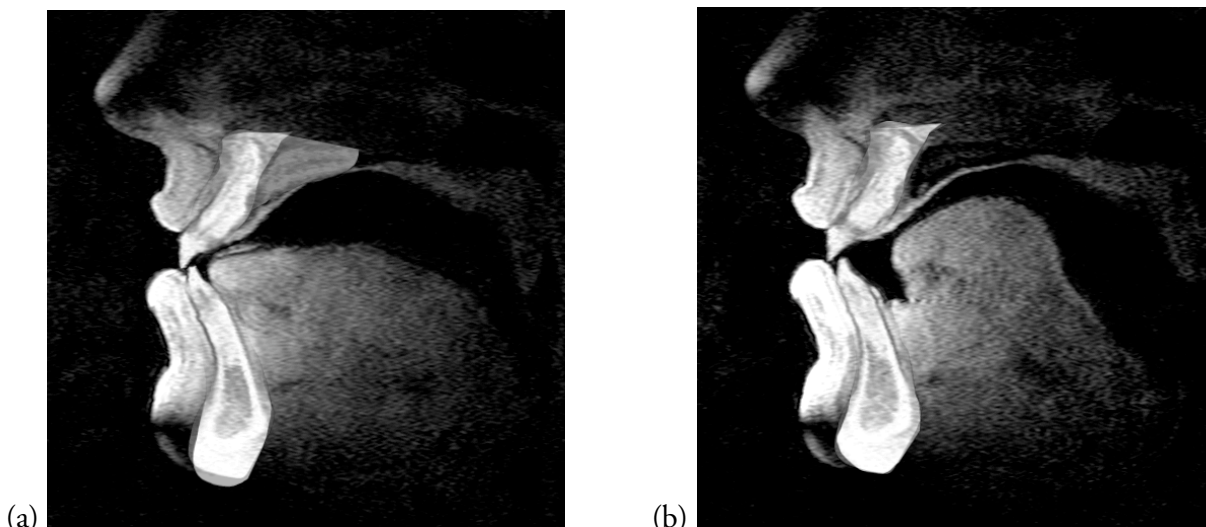


Figure 3. Examples of articulation for an English subject of the sibilants /s/ (3a) and /ʃ/ (3b).
(From Toda & Honda, 2003)

Because of its extremely slow acquisition speed, however, the subject often has to sustain the articulation artificially for 30 seconds or more (for e.g. 43 seconds for the entire set of 54 images in Engwall 2002). This has restricted MRI use to the study of sustained speech sounds, corresponding to 'static' tract shapes. Thanks to technical advances, it is nowadays possible to collect full 3D data in 5 seconds and of the midsagittal plane at 9 images per second (see Engwall 2006). However, the rate of 9 images/s is still not fast enough to observe articulatory movements. In addition, the quality of images is rather low. Another technique employs the stroboscopic principle (mainly ATR group: see Masaki et al. 1999). Because of the stroboscopic principle, a subject has to repeat a short utterance, typically less than 1 second, many times, typically 200 times, to obtain good image quality. Notice, however, that the image quality is highly dependent on the competence of individual speakers.

Another drawback of MRI is that the subjects have to be positioned in supine position lying on their back, due to the construction of the MRI scanner and antenna. The gravitational effects of this positioning might introduce some effects on the articulation. Engwall (2006) presented an evaluation of the effects of the supine position and the artificial sustaining based on MRI itself. His results showed that the MRI images give adequate information on the three-dimensional shape of the vocal tract and articulators, but, as his evaluation suggests, some caution should be taken. First, the artificial sustaining causes the articulations to be both hyperarticulated compared to a normally sustained articulation and more difficult to hold for the subject. Second, sustained productions are hyperarticulated compared to real-time production, in particular concerning coarticulation. Finally, the position, supine and facing upwards, does affect the position and shape of the tongue, often decreasing the passage in the pharynx, especially when the articulation has to be sustained artificially. A consequence of this evaluation is that MRI acquisition time should be kept as short as possible, as differences can be observed not only between real time and static articulations, but also depending on how long the articulation is sustained. The static MRI images thus have to be complemented with other measurements (e.g. EMA, EPG, or X-ray) to correctly replicate not only articulatory movements but also positions of running speech.

4. Electromagnetic Articulography (EMA)

Electromagnetic Articulography (EMA) is a suitable means for tracking movements within the vocal tract during speech production. Different EMA systems are available, including the Carstens Articulograph, the Botronic Movetrack system (Branderud, 1985), and the MIT system (Perkell et al. 1992). Carstens AG100 (<http://www.articulograph.de>) is by far the most used system among speech researchers. It is comprised of (1) a plastic helmet which subjects wear during data recordings (three transmitter coils are mounted equidistant from one another on the helmet) (2) small receiver coils placed inside the mouth or on the face and (3) an electronic connected to the computer. Figure 4 below shows the experimental set-up necessary for the recording of a subject using this technique.

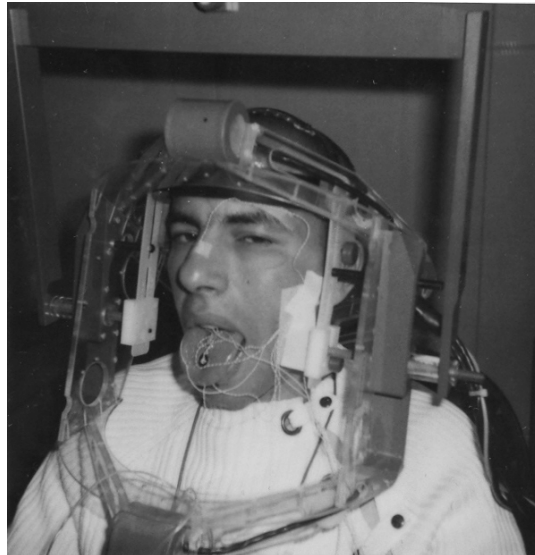


Figure 4. A photo of a subject (the author) during an EMA recording session, showing the helmet and some of the receiver coils inside the mouth and on the face.

EMA tracks midsagittal fleshpoints movements by measuring induced current from receiver sensors moving in a magnetic field (Perkell et al., 1992; Hoole, 1993). The magnetic field, generated at different frequencies by transmitter coils, induces an alternating signal in the receiver coils. The voltage of this signal is inversely related to the distance between the transmitter and the receiver coil. A computer algorithm provides the location of the receiver coil as it moves in x-y space over time.

Six receiver coils are commonly used for the measurements: two are placed on the upper and lower lip, three coils on the tongue (approximately 8 mm, 20 mm and 52 mm from the tip of the tongue, depending on speaker) and one on the base of low front incisors (to measure jaw movement). In addition, two receiver sensors, one on the base of the upper front incisor and one on bridge of the nose, are used as reference points for head-movement correction. Rotation and translation of the EMA sensor data is performed to ensure that the two reference coils are coincident across all frames for a given speaker. This removes any component of head movement from the data. A further rotation is performed to align the occlusal plane (also called 'bite plane') with the x-axis and a translation sets the origin at the position of the upper incisor reference coil.

The most important advantage of this system is related to the rapid tracking rate: EMA system samples articulatory data at 200 Hz. Another advantage is related to the ability of tracking multiple articulators simultaneously. These two aspects make it possible to measure with increasing accuracy interarticulation among different articulators. Notice, however, that the accuracy of the data recorded

decreases away from the center of the triangle of the transmitter coils. Hoole (1993), for instance, reported an error of 0.67 mm +/- 0.42 for positions more than 6 cm away from the center (in the midsagittal plane) and 0.2 mm +/- 0.13 for positions up to 6 cm. Another important aspect concerning the reliability of EMA data is related to rotational misalignments. Articulatory data can only be collected at midline and are thus subject to error as the articulators rotate left-to-right (see Hoole 1993).

The kinematic data recorded by the AG-100 system can be analyzed using the Carstens programs, including³: the *EMALYZE* program: a Windows program which analyzes and evaluates the movement and acoustic data acquired in the investigation, the *MultiCV* program which allows to create a copy of the Articulograph AG100 data in different formats or to rotate the movement curves (this is necessary for example if one plans to analyze EMA data with a program other than *EMALYZE*), and the *POINTS* program for measuring the accuracy of the Articulograph AG100.

5. Comparison among systems

As already stated, each technique has its own set of weaknesses and strengths. Ideally, techniques used to measure the movements of the vocal tract should:

- (i) permit the recording of the dynamics of all the articulators with an accurate temporal and spatial resolution,
- (ii) not perturb subjects' natural articulation and comfort,
- (iii) not involve any health hazard for subjects,
- (iv) not degrade the quality of the speech signal,
- (v) be portable for outside laboratory recordings, and
- (vi) be inexpensive.

These criteria (and others) are reported in Table 1 below in order to evaluate the weaknesses and strengths of each one of the five tools presented above.

Table 1. Comparison of the 5 speech measurement systems (V.T. = Vocal Tract, SS = Sound Signal, Mvt = Movement)

	EMA	MRI	Ultrasound	X-ray	X-ray microb.
<i>Whole V.T.</i>	No	Yes	No	Yes	No
<i>Tongue imaging</i>	Pellets	Full-length	Full-length	Full-length	Pellets
<i>Tongue root imaging</i>	No	Yes	Yes	Yes	No
<i>Velum imaging</i>	No	Yes	No	Yes	No
<i>Time resolution</i>	200 Hz	----	30-200 Hz	50 Hz	40-160 Hz
<i>3D</i>	Yes	Yes	No	No	No
<i>Health hazard</i>	No	No	No	Yes	?
<i>Natural art.</i>	Affected	Yes ⁴	Yes	Yes	Affected
<i>Quality of SS</i>	Good	Good	Degraded	Good	Good
<i>Head Mvt.</i>	Restricted ⁵	Restricted	Restricted ⁶	Free	Free
<i>Portable</i>	No	No	Yes	No	No
<i>Expensive</i>	Yes	Yes	No	Yes	Yes

³ A detail description of these and additional programs is spelled out in <http://www.linguistics.ucla.edu/faciliti/facilities/physiology/Emamual.html>.

⁴ The supine position during MRI recording may also affect the articulation (see above).

⁵ Head movement is free using a 3-dimensional magnetometry and restricted using a 2D system.

⁶ As was mentioned above, head movement is free using the Haskins Optically Corrected Ultrasound System (HOCUS).

References *(additional references are given below)*

- Akgul, Y., Kambhamettu, C., & Stone, M. (2000). A Task-Specific Contour Tracker for Ultrasound. IEEE Workshop on Mathematical Methods in Biomedical Image Analysis, Hilton Head Island, South Carolina, 135-142. <http://citeseer.ist.psu.edu/613227.html>
- Aron, M., Kerrien, E., Berger, M.O., & Laprie, Y. (2006) Coupling electromagnetic sensors and ultrasound images for tongue tracking: acquisition set up and preliminary results. In *Proceedings of the International Seminar on Speech Production*. <http://aspi.loria.fr/Save/aron.pdf>
- Arnal, A., Badin P., Brock G., Connan, P.Y., Florig, E., Perez, N., Perrier, P., Simon, P., Sock, R., Varin, L., Vaxelaire, B., Zerling, J.P., (2000). An X-ray database for French. In *Proceedings of the 5th Seminar on Speech Production*.
- Badin, P., Bailly, G., Revéret, L., Baciú, M., Segebarth, C., & Savariaux, C. (2002). Three-dimensional articulatory modelling of tongue, lips and face, based on MRI and video images. *Journal of Phonetics*, 30(3), 533-553.
- Branderud, P. (1985). Movetrack – a movement tracking system. In *Proceedings of the French-Swedish Symposium on Speech*, 113-122, Grenoble.
- Branderud, P., Lundberg, H-J., Lander, J., Djamshidpey, H., Wäneland, I., Krull, D., & Lindblom, B. (1998). X-ray analyses of speech: methodological aspects. *FONETIK 98*, Paper presented at the annual Swedish phonetics conference, Dept of Linguistics, Stockholm University.
- Engwall, O. (2002). *Tongue Talking – Studies in Intraoral Visual Speech Synthesis*. Ph.D. thesis KTH, Stockholm, Sweden.
- Engwall, O (2006). A revisit to the application of MRI to the analyses of speech production – testing our assumptions. *Proceedings of the 6th International Seminar on Speech Production, Sydney*. http://www.speech.kth.se/ctt/publications/papers03/ISSP6_MRI.pdf.
- Hoole, P. (1993). Methodological considerations in the use of electromagnetic articulography in phonetic research. *Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München* 31, 43-64.
- Li, M., Kambhamettu, C., & Stone, M. (2003). EdgeTrak, a program for band-edge extraction and its applications. Sixth IASTED International Conference on Computers, Graphics and Imaging, August 13-15, Honolulu, HI. <http://www.speech.umaryland.edu/Publications/Min%20Li%20band%20extraction.pdf>
- Masaki, S., Tiede, M.K., Honda, K., Shimada, Y., Fujimoto, I., Nakamura, Y., & Ninomia, N. (1999). MRI-based speech production study using a synchronized sampling method. *Journal of Acoustical Society of Japan*, 20, 375-379.
- 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.
- Perkell, J.S. (1969) *Physiology of speech production: results and implications of a quantitative cineradiographic study*. MIT Press, Cambridge, Mass,

Rochette, C. (1973). *Les groupes de consonnes en français*. Québec, Canada : Les Presses de l'Université Laval.

Perkell, J.S., Cohen, M.H., Svirsky, M.A., Matthies, M.L., Garabieta, I., & Jackson, M.T.T. (1992). Electromagnetic midsagittal articulometer systems for transducing speech articulatory movements. *Journal of the Acoustical Society of America* 92, 3078-3096.

Russell, G.O. (1928). *The vowel, its psychological mechanism, as shown by x-ray*. Columbus, OH: Ohio State University Press.

Stevens, K.N., & Öhman, S.E.G. (1963). Cineradiographic studies of speech. *KTH STL-QPSR*, 2, 9-11.

Stone M., Sonies, B.C., Shawker, T.H., Weiss, G., & Nadel, L. (1983). Analysis of real-time ultrasound images of tongue configuration using a grid-digitizing system. *Journal of Phonetics*, 11, 207-218.

Stone, M. & Davis, E. (1995). A head and transducer support system for making ultrasound images of tongue/jaw movement. *Journal of the Acoustical Society of America*, 98, 3107-3112.

Tiede, M., Masaki, S., & Vatikiotis-Bateson, E. (2000). Contrasts in speech articulation observed in sitting and supine condition. Proceedings of the 5th Speech Production Seminar: Models and data 25-28.

Toda, M, & Honda, K. (2003). An MRI-based cross-linguistic study of sibilant fricatives. In proceedings of the 6th International Seminar on Speech Production, Sydney.

Whalen, D.H., Iskarous, K., Tiede, M.T., Ostry, D., Lehnert-LeHoullier, H., & Hailey, D. (2005). The Haskins Optically-Corrected Ultrasound System (HOCUS). *Journal of Speech, Language, and Hearing Research*, 48, 543-553.

Speech production studies

Some Selected References

1. Studies using X-ray

Abbs, J.H., & Nadler, R.D. (1987). *User's manual for the University of Wisconsin X-ray microbeam*. Madison: University of Wisconsin—Madison, Waisman Research Center.

Al-Ani, S.H. (1970). *Arabic Phonology: An Acoustical and Physiological Investigation*. Motoun: The Hague.

Arnal, A., Badin P., Brock G., Connan, P.Y., Florig, E., Perez, N., Perrier, P., Simon, P., Sock, R., Varin, L., Vaxelaire, B., Zerling, J.P., (2000a). An X-ray database for French. In *Proceedings of the 5th Seminar on Speech Production*.

Arnal, A., Badin P., Brock G., Connan, P.Y., Florig, E., Perez, N., Perrier, P., Simon, P., Sock, R., Varin, L., Vaxelaire, B., Zerling, J.P., (2000b). Une base de données cinéradiographiques du français. In *Actes des 23e Journées d'Etude sur la Parole*, pp. 425-428.

Ball, M. (1984). X-ray techniques. In Code, C. & Ball, M. (eds.), *Experimental Clinical Phonetics. Investigatory Techniques in Speech Pathology and Therapeutics*, pp.107-128.

Bell-Berti, F. & Raphael, L.J. (eds.), (1995). *Producing Speech: Contemporary Issues*. For Katherine Safford Harris. AIP Press, New York.

Berger, M.-O., Mozelle, G. & Laprie, Y. (1995). Cooperation of active contours and optical flow for tongue tracking in X-ray motion pictures. In *Proceedings of the 9th Scandinavian Conference on Image Analysis*. [abstract].

Berger, M.-O. & Laprie, Y. (1996). Tracking articulators in X-ray images with minimal user interaction: example of the tongue extraction. In *Proceedings of IEEE International Conference on Image Processing - ICIP'96*. [abstract].

Branderud, P., Lundberg, H-J., Lander, J., Djamshidpey, H., Wäneland, I., Krull, D., & Lindblom, B. (1998). X-ray analyses of speech: methodological aspects. *FONETIK 98*, Paper presented at the annual Swedish phonetics conference, Dept of Linguistics, Stockholm University.

Browman, C., & Goldstein, L. (1992). "Targetless" schwa: an articulatory analysis. In Gerard Docherty & D. Robert Ladd, (eds.), *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*. Cambridge: Cambridge University Press.

Carmody, F. (1941). An x-ray study of pharyngeal articulation. *University of California Publications in Modern Philology*, 21(5), 377-384.

Cerda, R. (1968). Une méthode pour la mesure physiologique d'après les films radiologiques. *Zeitschrift für Phonetik XXI*, 6, pp. 518-520.

Chiba, T. & Kajiyama, M. (1958). *The vowel: Its nature and structure*. The Phonetic Society of Japan, Tokyo.

Dart, S.N. (1987). A bibliography of X-ray studies of speech. *UCLA Working Papers in Phonetics*, 66, pp. 1-97.

Fant, G. (1970). *Acoustic theory of speech production: with calculations based on X-ray studies of Russian articulations*. The Hague, Paris.

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.

Ghazali, S. (1977). *Back consonants and backing coarticulation in Arabic*. Ph.D. Dissertation, University of Texas, Austin.

Gick, B. (2002). An X-ray investigation of pharyngeal constriction in American English schwa. *Phonetica*, 59, pp. 38-48.

Gick, B., Min Kang, A., & Whalen, D.H. (2002). MRI and X-ray evidence for commonality in the dorsal articulations of English vowels and liquids. *Journal of Phonetics*, 30, pp. 357-371.

Grégoire, L. (1983). Contribution à l'étude des groupes consonne plus voyelle en français, à l'aide de la radocinématographie et de l'oscillographie. Publication B-120, C.I.R.B.

Harshman, R., Ladefoged, P., & Goldstein, L. (1977). Factor Analysis of Tongue Shapes. *Journal of the Acoustical Society of America*, 62, pp. 693- 707.

Hashimoto, K. & Sasaki, K (1982). On the relationship between the shape and position of the tongue for vowels. *Journal of Phonetics*, 19, pp. 291-299.

Keating, P.A. (1988). Palatals as complex segments: X-ray evidence. *UCLA Working Papers in Phonetics*, 69, pp. 77-91.

Kiritani, S. (1978). Perturbation of the consonant and vowel articulation by a adjacent segments. *Journal of the Acoustical Society of Japan*, 34, pp. 132-139.

Kiritani, S. (1986). X-ray microbeam method for measurement of articulatory dynamics: techniques and results. *Speech Communication*, 5(2), pp. 119-140.

Kiritani, S., Itoh, K., & Fujimura, O. (1975). Tongue-pellet tracking by a computer-controlled x-ray microbeam system. *Journal of the Acoustical Society of America*, 57, pp. 1516-1520.

Lindau, M. & Ladefoged, P. (1989). Methodological studies using an x-ray microbeam system. *UCLA Working Papers in Phonetics*, 72, pp. 83-90.

Lindfelt, B. (1988). L'enchaînement des consonnes occlusives suivies de voyelles antérieures en français. Publication B-165, C.I.R.B.

Maeda, S. (1979). Un modèle articulatoire de la langue avec des composantes linéaires. In *Actes des 10^e Journées d'Etudes sur la parole*, pp. 152-164.

Maeda, S. (1990). Compensatory articulation during speech: evidence from the analysis of vocal tract shapes using an articulatory model. In W.J. Hardcastle and A. Marchal (eds.), *Speech Production and Speech Modelling*, pp. 131-149. Kluwer Academic Publishers, Dordrecht.

Mermelstein, P. (1973). Articulatory model for the study of speech production, *Journal of the Acoustical Society of America*, 53 (4), pp. 1070-1082.

McGowan, R.S. (2002). Tongue position and orientation for front vowels in the X-Ray Microbeam Speech Production DataBase. *Journal of the Acoustical Society of America Journal*, 111(5), pp. 2481-2481.

Moll, K. (1960). Cinefluorographic techniques in speech research. *Journal of Speech and Hearing Research*, 3, pp. 227-241

Munhall, K.G., Vatikiotis-Bateson, E., & Tohkura, Y. (1994). Manual for the X-ray film database. *ATR Technical Report*, TR-H-116.

Munhall, K.G., Vatikiotis-Bateson, E., & Tohkura, Y. (1995a). X-ray Film database for speech research. *Journal of the Acoustical Society of America*, 98, 1222-1224.

Munhall, K.G., Vatikiotis-Bateson, E., & Tohkura, Y. (1995b). *X-ray film database for speech research* [Videodisc]. Kyoto, Japan: ATR Laboratories.

Nadler, R.D., Abbs, J.H., & Fujimura, O. (1987). Speech movements research using the new X-ray microbeam system. In *Proceedings of ICPHS*, Vol. 6, pp. 10-27.

Papcun, G., Hochberg, J., Thomas, T. R., Laroche, F., Zacks, J., & Levy, S. (1992). Inferring articulation and recognizing gestures from acoustics with a neural network trained on x-ray microbeam data. *Journal of the Acoustical Society of America*, 92, pp. 688-700.

Perkell, J.S. (1969). *Physiology of Speech Production: Results and Implications of a Quantitative Cinreradiographic Study*. MIT Press, Cambridge, MA.

Rochette, C. (1973). *Les groupes de consonnes en français*. Les Presses de l'Université Laval, Québec.

Rochette, C. (1977). Radiologie et phonétique. *Vie Médicale au Canada Français*, 6, pp. 55-67.

Rochette, C. & Grégoire, L. (1981). Contribution à l'étude des groupes consonne plus voyelle en français, à l'aide de la radocinématographie. *Phonétique combinatoire I*, C.I.R.B.

Rochette, C. & Lindfelt, B. (1987) L'enchaînement des consonnes occlusives suivies de voyelles en français, à l'aide de la radocinématographie. *Phonétique combinatoire III*, C.I.R.B.

Rochette, C. & Simard, C. (1985). Étude des séquences de type consonne constrictive plus voyelle en français, à l'aide de la radocinématographie. *Phonétique combinatoire II*, C.I.R.B.

Roy, J.P., Sock, R., Vaxelaire, B., & Hirsch, F. (2003). Auditory effects of anticipatory and carryover coarticulation: X-ray and acoustic data. In *Proceedings of the 6th International Seminar on Speech Production*, pp. 243-248.

Russell, G.O. (1928). *The vowel: Its physiological mechanism as shown by x-ray*. Columbus: Ohio State University Press.

Shirai, K. & Honda, M. (1978). Estimation of articulatory parameters from speech sound. *Trans. IECE* 61, pp. 409–416.

Simard, C. (1988). *Étude des séquences de type consonne constrictive plus voyelle en français, à l'aide de la radocinématographie et de l'oscillographie*. Publication B-148, C.I.R.B.

Simon, P. (1967). *Les consonnes françaises. Mouvements et positions articulatoires à la lumière de la radiocinématographie*. Paris: Klincksieck

Stone, M. (1990). A three-dimensional model of tongue movement based on ultrasound and x-ray microbeam data. *Journal of the Acoustical Society of America*, 87, pp. 2207-2217.

Straka, G. (1965). *Album Phonétique*. Presses de l'Université Laval, Québec.

Strenger, F. (1968). Radiographic, palatographic and labiographic methods in phonetics. In B. Malmberg (ed.), *Manual of Phonetics*, pp. 334-364. Amsterdam: North-Holland.

Thompson, M. (1984). *X-ray microbeam manual* (unpublished draft). Stoughton, WI: University of Wisconsin Physical Sciences Laboratory

Tiede, M. & Vatikiotis-Bateson, E. (1994) Exploiting a videodisc-based cineradiographic database for speech research. *Journal of the Acoustical Society of America*, 95, p. 2822.

Turk, A.E. (1993). *Effects of Position-in-Syllable and Stress on Consonant Articulation*. Ph.D. Dissertation, Cornell University.

Vaxelaire, B. (1995a). Single vs. double (abutted) consonants across speech rate. X-ray and acoustic data for French. In *Proceedings of ICPhS*, vol. 1, pp. 384-387.

Vaxelaire, B. (1995b) Geometric and temporal constraints in the production of French consonant sequences. X-ray and acoustic Data for French. In *Proceedings of the 4th European Conference on Speech Communication and Technology*, vol. 2, pp. 1285-1288.

Vaxelaire, B. & Sock, R. (1996). A cineradiographic and acoustic study of velar gestures in French consonant sequences as a function of speech rate. In *Proceedings of the 4th Speech Production Seminar*, pp. 65-68.

Vaxelaire, B., Sock, R., Bonnot, J.F., & Keller, D. (1999). Anticipatory labial activity in the production of French rounded vowels. X-ray and acoustic data. In *Proceedings of ICPhS*, pp. 53-56.

Vilain, A., Abry, C., & Badin, P. (1999). Motor equivalence evidenced by articulatory modelling. In *Proceedings of Eurospeech99*, vol.1, pp. 169-172.

Weismer, G., Yunusova, Y., & Westbury, J.R. (2003). Interarticulator coordination in dysarthria: an X-ray microbeam study. *Journal of Speech, Language, and Hearing Research*, 46, pp. 1247–1261.

Westbury, J.R. (1994). *X-ray microbeam speech production database user's handbook* [Software manual]. Madison: University of Wisconsin—Madison, Waisman Research Center.

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.

Wood, S. (1979). A radiographic examination of constriction location for vowels. *Journal of Phonetics*, 7, pp. 25-43.

Wood, S. (1982). X-ray and model studies of vowel articulation. *Working Paper 23*. Lund, Sweden: Lund University, Department of Linguistics.

Wood, S. (1991). X-ray data on the temporal coordination of speech gestures. *Journal of Phonetics*, 19, pp. 281-292.

Wood, S. (1993). Syllable structure and the timing of speech gestures: an analysis of speech gestures from an X-ray motion film of Bulgarian speech. In R. Aulanko & A-M. Korpijaakko-Huuhka (eds.), *Proceedings of the Third Congress of the International Clinical Phonetics and Linguistics Association*, pp. 191-200.

Wood, S. (1997a). The gestural organization of vowels and consonants: a cineradiographic study of articulator gestures in Greenlandic. In *Proceedings of the 5th European Conference on Speech Communication and Technology September*, pp. 387-388.

Wood, S. (1997b). A cinefluorographic study of the temporal organization of articulator gestures: examples from Greenlandic. *Speech Communication*, 22, pp. 207-225.

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^o Cycle, Institut de Phonétique, Université de Nancy II.

2. Studies using ultrasound

Akgul, Y., Kambhamettu, C., & Stone, M. (1999). Automatic extraction and tracking of the tongue contours. *IEEE Transactions on Medical Imaging*, 18, pp. 1035-1045.

Aron, M., Kerrien, E., Berger, M.O., & Laprie, Y. (2006) Coupling electromagnetic sensors and ultrasound images for tongue tracking: acquisition set up and preliminary results. In *Proceedings of the International Seminar on Speech Production*. <http://aspi.loria.fr/Save/aron.pdf>

Davidson, L., & Stone, M. (2003). Epenthesis versus gestural mistiming in consonant cluster production: an ultrasound study. In *Proceedings of the WCCFL 22*.

Epstein, M.A. (2005). Ultrasound and the IRB. *Clinical Linguistics and Phonetics*, 19 (6-7), pp. 567-572.

Epstein, M.A. & Stone, M. (2005). The tongue stops here: ultrasound imaging of the palate. *Journal of the Acoustical Society of America*, 118 (4), pp. 2128-2131.

Epstein, M.A., Stone, M., Pouplier, M., & Parthasarathy, V. (2004). Obtaining a palatal trace for ultrasound images. *Journal of the Acoustical Society of America*, 115(5). [Abstract].

Kaburagi, T. & Honda, M. (1994). A trajectory formation model of articulatory movements based on the motor tasks of phoneme-specific vocal tract shapes. In *ICSLP-1994*, pp. 579-582.

Keller, E. (1987a). Mesures ultrasoniques des mouvements du dos de la langue en production de la parole: aspects cliniques. *Folia Phoniatica*, 39, pp. 52-62.

Keller, E. (1987b). Ultrasound measurement of tongue dorsum movements in articulatory speech impairments. In J.H. Ryalls (ed.), *Phonetic Approaches to Speech Production in Aphasia and Related Disorders*, pp. 93-112.

Keller, E. & Ostry, D. (1983). Computerized pulsed echo ultrasound measurements of tongue dorsum movements. *Journal of the Acoustical Society of America*, 73, pp. 1309-1315.

Lundberg, A. & Stone, M. (1999). Three-dimensional Tongue Surface Reconstruction: Practical Considerations for Ultrasound Data. *Journal of the Acoustical Society of America*, 106, pp. 2858-2867.

Morrish, K., Stone, M., Shawker, T., & Sonies, B.C. (1985). Distinguishability of tongue shape during vowel production. *Journal of Phonetics*, 13(2), pp. 189-204.

Morrish, K., Stone, M., Sonies, B., Kurtz, D., & Shawker, T. (1984). Characterization of tongue shape. *Ultrasound Imaging*, 6(1), pp. 37-47.

Ong, D. & Stone, M. (1998). Three-dimensional vocal tract shapes in /r/ and /l/: A study of MRI, ultrasound, electropalatography, and acoustics. *Phonoscope*, 1, pp. 1-13.

Parthasarathy, V., Prince, J.L., & Stone, M. (2005). Spatiotemporal visualization of the tongue surface using ultrasound and Kriging (SURFACES). *International Journal of Clinical Linguistics and Phonetics*.

- Shawker, T., Stone, M., & Sonies, B.C. (1985). Tongue pellet tracking by ultrasound: Development of a reverberation pellet. *Journal of Phonetics*, 13, pp. 135-146.
- Slud, E., Smith, P., Stone, M., & Goldstein, M. (2002). Principal components representation of the two-dimensional coronal tongue surface. *Phonetica*, 59 (2-3), pp. 108-133.
- Stone, M. (1990). A three-dimensional model of tongue movement based on ultrasound and x-ray microbeam data. *Journal of the Acoustical Society of America*, 87, pp. 2207-2217.
- Stone, M. (1991). Imaging the tongue and vocal tract. *British Journal of Disorders of Communication*, 26, pp. 11-23.
- Stone, M. (2005). A guide to analysing tongue motion from ultrasound images. *Clinical Linguistics and Phonetics*, 19 (6-7); 455-502.
- Stone, M., Epstein, M., & Iskarous, K. (2004). Functional segments in tongue movement. International Journal of Clinical Linguistics and Phonetics. *Clinical Linguistics and Phonetics* 18(16-18), 507-522.
- Stone, M., Epstein, M.A., Kambhamettu, C., & Li, M. (2006). Predicting 3D tongue shapes from midsagittal contours. In J. Harrington & M. Tabain (eds.), *Speech Production: Models, Phonetic Processes, and Techniques*, pp. 315-330. Psychology Press.
- Stone, M., Epstein, M.A., & Sutton, M.W. (2003). Predicting 3D tongue shapes from midsagittal contours. In *Proceedings of the 6th International Seminar on Speech Production*.
- Stone, M. & Lundberg, A. (1996). Three-Dimensional Tongue Surface Shapes of English Consonants and Vowels. *Journal of the Acoustical Society of America*, 99 (6), pp. 3728-3737.
- Stone, M. & Shawker, T. (1986). An ultrasound examination of tongue movement during swallowing. *Dysphagia*, 1, pp. 78-83.
- Sze, C-F. (2000). *Reconstructing 3-D tongue motion from 2-D ultrasound images and speech signals*. Ph.D. Dissertation, University of Maryland.
- Unser, M. & Stone, M. (1992). Automated detection of the tongue surface in sequences of ultrasound images. *Journal of the Acoustical Society of America*, 91, pp. 3001-3007.
- Wrench, A. A., & Scobbie, J. M. (2003). Categorising vocalisation of English / l / using EPG, EMA and ultrasound. In *Proceedings of the 6th International Seminar on Speech Production*, pp. 314-319.
- Yang, CS., & Stone, M. (2002). Dynamic programming method for temporal registration of three-dimensional tongue surface motion from multiple utterances. *Speech Communication*, 38 (1-2), pp. 199-207.

3. Studies using Magnetic Resonance Imaging (MRI)

Bresch, E., Nielsen, J., Nayak, K., & Narayanan, S. (2006). Synchronized and noise-robust audio recordings during real-time MRI scans. *Journal of the Acoustical Society of America*, 2006 [Abstract].

Bresch, E., Nielsen, J., Nayak, K., & Narayanan, S. (2006). Synchronized audio recording and real-time MR imaging of fluent speech. In *Proceedings ISMRM Workshop on Real-Time MRI*. Santa Monica, February 2006.

Engwall, O. (2006). Assessing MRI measurements: Effects of sustenation, gravitation and coarticulation. In Harrington, J. and Tabain, M. (eds.) *Speech production: Models, Phonetic Processes and Techniques*, pp. 301-314. Psychology Press, New York.

Engwall, O. (2006). A revisit to the application of MRI to the analyses of speech production – testing our assumptions. In *Proceedings of the 6th International Seminar on Speech Production*. Sydney, Australia.

Inoue, MS., Ono, T., Honda, E., Kurabayashi, T., & Ohyama, K. (2006). Application of magnetic resonance imaging movie to assess articulatory movement. *Orthodontics and Craniofacial Research* 9(3), pp. 157-162.

Lee, S., Bresch, E., Adams, J., Kazemzadeh, A., & Narayanan, S. (2006). A study of emotional speech articulation using a fast magnetic resonance imaging technique. In *Proceedings of InterSpeech ICSLP*, Pittsburgh, PA, September 2006.

Narayanan, S. Bresch, E., Tobin, S., Byrd, D., Nayak, K., & Nielsen, J. Resonance tuning in soprano singing and vocal tract shaping: Comparison of sung and spoken vowels. *Journal of the Acoustical Society of America*, 119(5), p. 3305. [Abstract].

NessAiver, M., Stone, M., Parthasarathy, V., Kahana, Y., Kots, A., & Paritsky, A. (2006) Recording high quality speech during tagged Cine MRI studies using a fiber optic microphone. *Journal of Magnetic Resonance Imaging* 23, pp. 92-97.

Takemoto, H., Honda, K., Masaki, S., Shimada, Y., & Fujimoto, I. [2006] Measurement of temporal changes in vocal tract area function from 3D cine-MRI data. *Journal of the Acoustical Society of America*, 119, pp. 1037-1049.

Tobin, S, D. Byrd, E. Bresch, & S. Narayanan. Syllable structure effects on velum-oral coordination evaluated with real-time MRI. *Journal of the Acoustical Society of America*, 119 (5), p. 3302. [Abstract].

2005

Kitamura, T., Takemoto, H., Honda, K., Shimada, Y., Fujimoto, I., Syakudo, Y., Masaki, S., Kuroda, K., Oku-uchi, N., & Senda, M. (2005). Difference in vocal tract shape between upright and supine postures: Observation by an open-type MRI scanner. *Acoustical Science and Technology*, 5, pp 465-468.

Narayanan, S. (2005). Imaging for understanding speech communication: Advances and challenges. *Journal of the Acoustical Society of America*, 117, p. 2501 [Abstract].

Serrurier, A. & Badin, P. (2005). Towards a 3D articulatory model of velum based on MRI and CT images. *ZAS Papers in Linguistics*, 40, pp. 195-211.

2004

Engwall, O. (2004a). From real-time MRI to 3D tongue movements. In *Proceedings of ICSLP 2004*, vol. II, pp. 1109-1112.

Engwall, O. (2004b). Speaker adaptation of a three-dimensional tongue model. In *Proceedings of ICSLP 2004*, vol. I, pp. 465-468.

Honda, K., Takemoto, H., Kitamura, T., Fujita, S., & Takano, S. (2004). Exploring human speech production mechanisms by MRI. *IEEE Trans., Inf., & Syst., E87-D*, pp. 1050-1058.

Narayanan, S., Nayak, K., Lee, S., Sethy, A., & Byrd, D. (2004). An approach to real-time magnetic resonance imaging for speech production. *Journal of the Acoustical Society of America*, 115(4), pp. 1771-1776.

Takemoto, H., Kitamura, T., Nishimoto, H. & Honda, K (2004). A method of tooth superimposition on MRI data for accurate measurement of vocal tract shape and dimensions. *Acoustical Science and Technology*, 25(6), pp. 468-474.

Vaissière, J. (2004). From X-ray or MRI data to sounds through articulatory synthesis: towards an integrated view of the speech communication process. In *Proceedings of ICSLP 2004*. http://www.isca-speech.org/archive/interspeech_2004/i04_P4.html

References up to 2003 are available at:

<http://aune.lpl.univ-aix.fr/~ghio/bib-IRManat.htm>

On the basics of MRI: <http://www.cis.rit.edu/htbooks/mri/>

4. Studies using Electromagnetic Articulography (EMA)

An updated list of research studies using EMA is available at:

<http://www.articulograph.de/pub2006-new.pdf>.