

REAL TIME MRI AND ARTICULATORY COORDINATIONS IN VOWELS

Didier Demolin*, Thierry Metens^o and Alain Soquet*

**Laboratoire de Phonologie, Université Libre de Bruxelles*

^oUnité de Résonance Magnétique, Hôpital Erasme, Université Libre de Bruxelles

Email: ddemoli@ulb.ac.be

ABSTRACT

This paper describes the use of a real time MRI technique to study the coordination of articulatory movements during the production of vowels. The technique was also applied to study articulatory compensations during bite block experiments (Demolin et al. 1997). One of the main interest of this technique is to allow a detailed spatial description of the main articulators (lip protrusion and opening, jaw opening and retraction, dorsum position, velum opening, larynx height and tongue root advancement) during the production of a sequence of vowels. The technique is also applicable to study the coarticulation of consonants and vowels, but because it is still not possible to obtain a satisfactory sound coordination and because the frame interval is still a bit slow, this question will not be treated in this paper. Measurement accuracy on real time images is evaluated by comparison with similar measurements on static MR images.

1 INTRODUCTION

Magnetic Resonance (MR) allows imaging of the vocal tract during phonation (Baer et al. 1991). More recently, simultaneous MR acquisition of multi oblique slices in a sole 14 sec acquisition provided an improved generation of area functions, which is an important step in the study of the relation between vocal tract geometry and speech acoustics. Similarly, the acquisition of parallel joints slices of 1 mm² resolution was also possible. Sustained phonation was required during acquisitions (14 sec) (Demolin et al. 1996). Nevertheless, these experiments were restricted to the study of oral or nasal vowels, because of the low temporal resolution. Progress made to increase the temporal resolution are limited by low Signal-to-Noise ratio and by susceptibility artifacts when fast gradient echo techniques are involved. In the present work, we adapted an ultra fast implementation of Turbo Spin Echo (TSE) to achieve a dynamic continuous monitoring of the vocal tract with an actual time resolution of 4 images per second (Demolin et al. 1997). We aimed to show that this technique can be used to study the relative movements of the main articulations involved in speech production i.e. lips, tongue, larynx, lower jaw and velum.

2 MATERIALS AND METHODS

Subjects were lying in supine position in a 1.5 T MR system equipped with fast gradients (CompactPlus, PowerTrak 6000, Philips Medical Systems, Best, The Netherlands). The receiver coil was a quadrature neck coil. The subjects were fixed with solid foam cushions in order to prevent any movement of their head. The MR procedures were performed in accordance with

FDA and European Rules concerning individuals safety, including the specific absorption rate below 4W/kg body weight.

The static views were provided by the acquisition of 13 TSE images, independently positioned and orientated, that were simultaneously acquired during a 12 sec continuous phonation. Each of these images covered a field of view of 250 x 200 mm with a spatial resolution of 1.2 mm² and was positioned on a reference image. This reference image consisted of a medio-sagittal section acquired immediately before and under a similar phonation. The contrast of these images was merely influenced by the proton density of the tissues. A remarkable feature of TSE images is that they are free of susceptibility artifacts, unlike other fast image acquisition techniques like the so called Echo Planar images, (Mansfield 1977) and gradient-echo images (Haase et al 1986).

The real time studies were performed with Ultra fast Turbo Spin Echo (TSE) Zoom sequence (previously named Lolo sequence; see Van Vaals et al. 1994 and Metens et al. 1997). A single sagittal T1-weighted section of 6 mm thickness was continuously acquired during at least 20 sec, at the rate of four or five images per second. The sections were carefully positioned in the mid-sagittal plane on survey images acquired in three orthogonal planes. For the four image per second acquisition, we used TR = 250 ms, with 186 ms acquisition and 64 ms delay between acquisition of consecutive image, to allow some magnetization T1-recovery. For the five image per second acquisition: TR = 200 ms, with 186 ms acquisition and 14 ms delay. All acquisitions were implemented with the following parameters: TE=21 ms, theta = 60°, Echo Spacing = 4.4 ms, Water-fat shift = 0.5 pixels, Turbo spin echo factor = 20 and 62.5% Partial Fourier acquisition. The rectangular field of view in the antero-posterior and caudo-cranial directions was of 125 x 250 mm, with a 32 x 128 matrix, providing a spatial resolution of 3.9 x 1.95 mm. The TSE Zoom sequence is designed such that the initial 60° and the subsequent 180° refocusing pulses excite perpendicular slabs, resulting into an intersecting slice, free of foldover artifacts and without compromising the spatial resolution. The images were reconstructed and displayed in real time. The operator instructed the subject to initiate the speech process by counting every second, five seconds before the acquisition start. Compared to a previous implementation on a slow gradient system (Demolin et al. 1997) the echo time was reduced by 9 ms, resulting in an improved image quality because of reduced susceptibility artifacts. The T2 effects in the image contrast and blurring were reduced as well.

Repetitions of /ieaou/ have been recorded for a male and a female speaker at respectively 5 and 4 images per second.

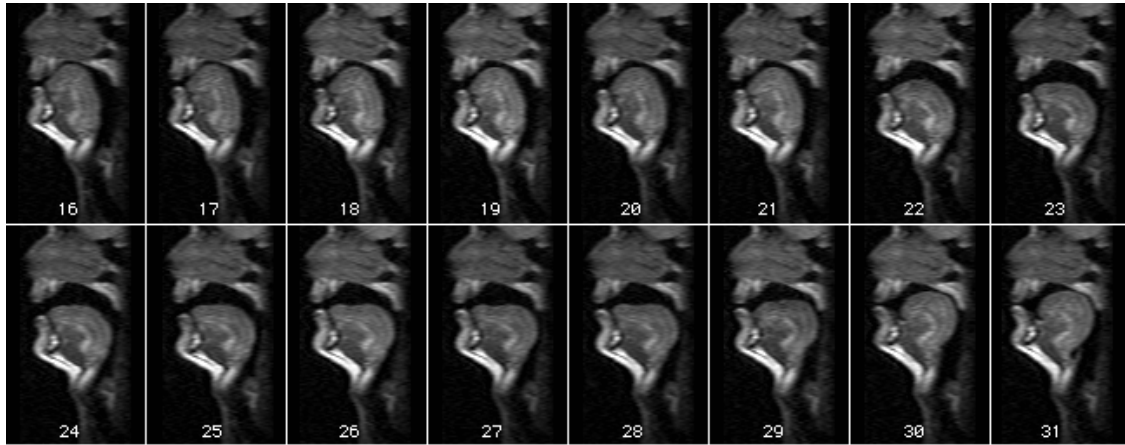


Figure 1: Real time acquisition of one repetition of the sequence /ieaou/ pronounced by the male speaker. The 16 images represent a total of 3.2 s with one image every 200 ms.

3 IMAGE ANALYSIS

Figure 1 presents the images of one repetition of the sequence /ieaou/ pronounced by the male speaker. The duration of this sequence is 3.2 s which gives a total of 16 images. It can be seen that the different articulators involved are well imaged (lips, tongue, hard palate, velum and larynx). The back of the pharynx is less contrasted. This can be explained by the fact that it is located close to the edge of the coil where its detection sensitivity becomes poor. The articulatory configurations of the different vowels can be observed respectively on image 16 for /i/, 19 for /e/, 24 for /a/, 27 for /o/ and 31 for /u/.

In order to measure the deformation of the vocal tract, a measurement grid has been adjusted to both speakers. The grid is designed to allow the measurement of the lip opening (1), larynx position (14 for the male speaker and 13 for the female speaker), and a set of midsagittal distances distributed along the vocal tract (12 for the male speaker and 11 for the female speaker). The grid is placed so that each grid line is in first approximation orthogonal to the fixed articulators (back of the pharynx and hard palate) and to the flux line (see figure 2).

Once the grid is positionned, each image can be sampled along the grid lines. If this process is repeated sequentially for each image of the acquisition, the evolution in time along this grid line can be displayed and the movement of different articulator measured.

Figure 2 shows the results obtained for both speakers; the beginning and end of one repetition of the sequence /ieaou/ is labelled respectively by the cursors *a* and *b*. It can be observed that:

- The movement along each grid line appears clearly.
- The problem of low contrast between the back of the pharynx and the vocal tract depends on the size of the speaker. Given a width of the imaging window of 125 mm, it was easier to cover entirely the vocal tract of the female than the one of the male speaker.
- In the hard palate region, the distances are quite stable during each vowel and move rapidly into the configuration of the next vowel. On the contrary, in the soft palate region and in the pharynx, the deformation appears to be more gradual, except for the transition from /e/ to /a/ which

involves larger modification of the overall shape of the vocal tract.

- The larynx goes downward during the production of the sequence /ieaou/ for the male speaker. The position for the five vowels are given in table 1.

Table 1: Larynx position in mm relative to the grid line 14 for the male speaker (lower values meaning lower position).

Vowel	/i/	/e/	/a/	/o/	/u/
Larynx height [mm]	33.1	31.6	22.6	19.7	17.6

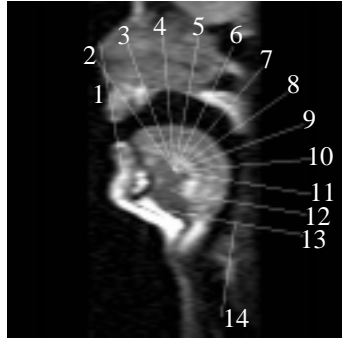
Figure 3 shows a comparison of the sagittal distances for the vowels /i, e, a, o/ for the male speaker measured on a real time acquisition and on a static acquisition (Demolin et al. 1996). The measurement grid has been placed in a similar way on both the real time and the static midsagittal images. It can be seen from the results that sagittal distances are comparable for the two sets of measurements. The differences are always inferior to 5 mm with an average of 2.0 mm. This result is acceptable regarding the fact that (1) the two acquisitions were made at 3 years interval, (2) that the task of the subject were different (the vowels were either produce in a sequence or sustained for 12 seconds), and (3) the pixel size for the real time sequence is 3.9 x 1.95 mm.

4 DISCUSSION

Subsecond MRI technique allows to explore movement of articulators involved during normal speech production in real time. In order to validate the technique, we have compared the midsagittal distances measured on a measurement grid on both static MR images and real time MR images of the same speaker pronouncing the same vowels. The results of this comparison show that the real-time technique gives accurate and reliable information on the position of the articulators involved in speech production.

The simultaneous acquisition of the speech sound remains a problem given the high intensity of the noise during acquisition. We currently use the sound at the input of the intercom; this does not provides a quality sufficient for

Male speaker



Female speaker

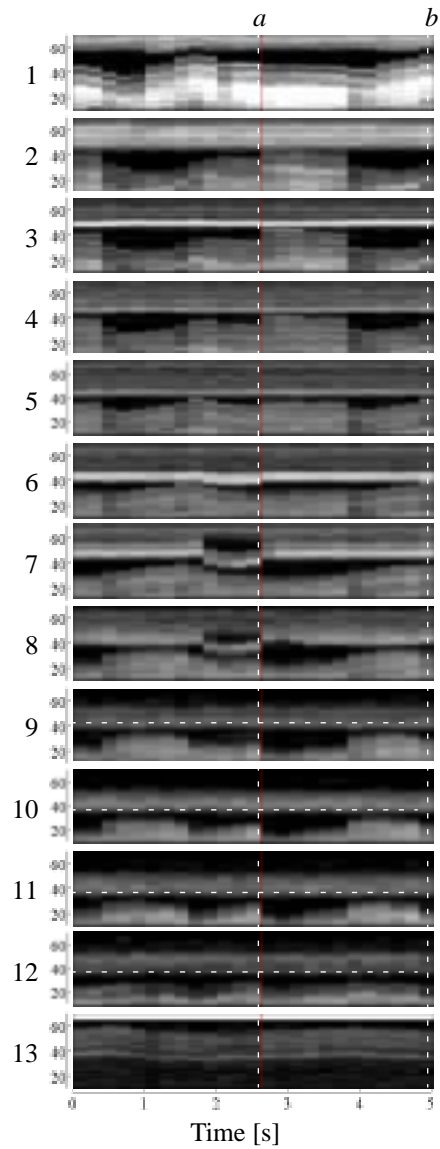
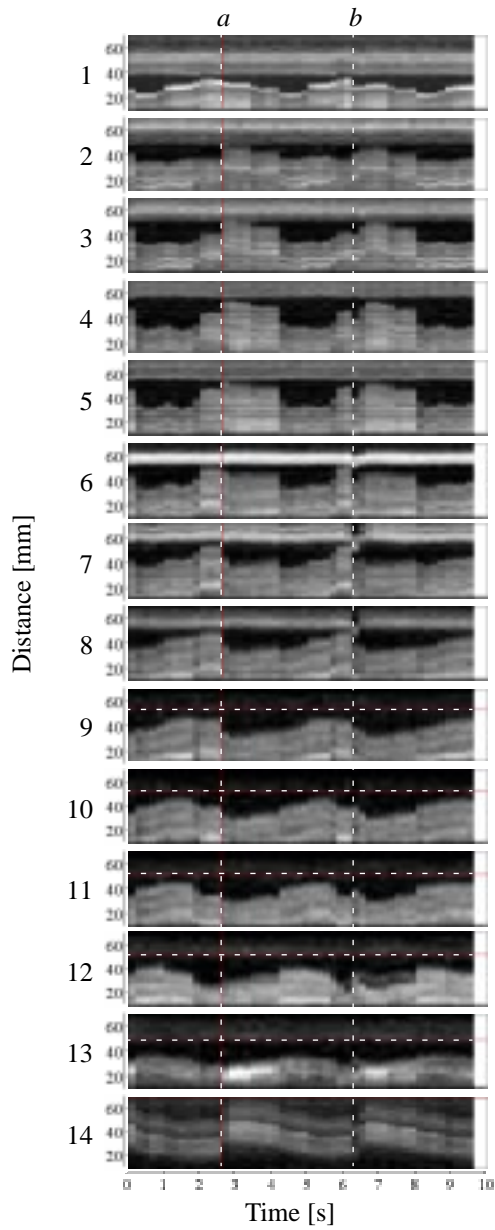
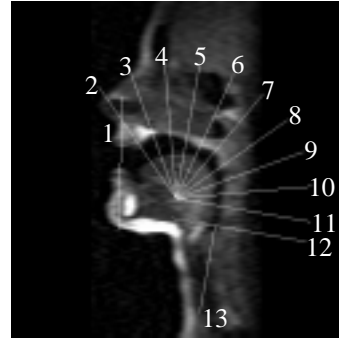


Figure 2: Position of the measurement grid on the male (left) and female (right) speaker and analysis of the movement in time along each gridline.

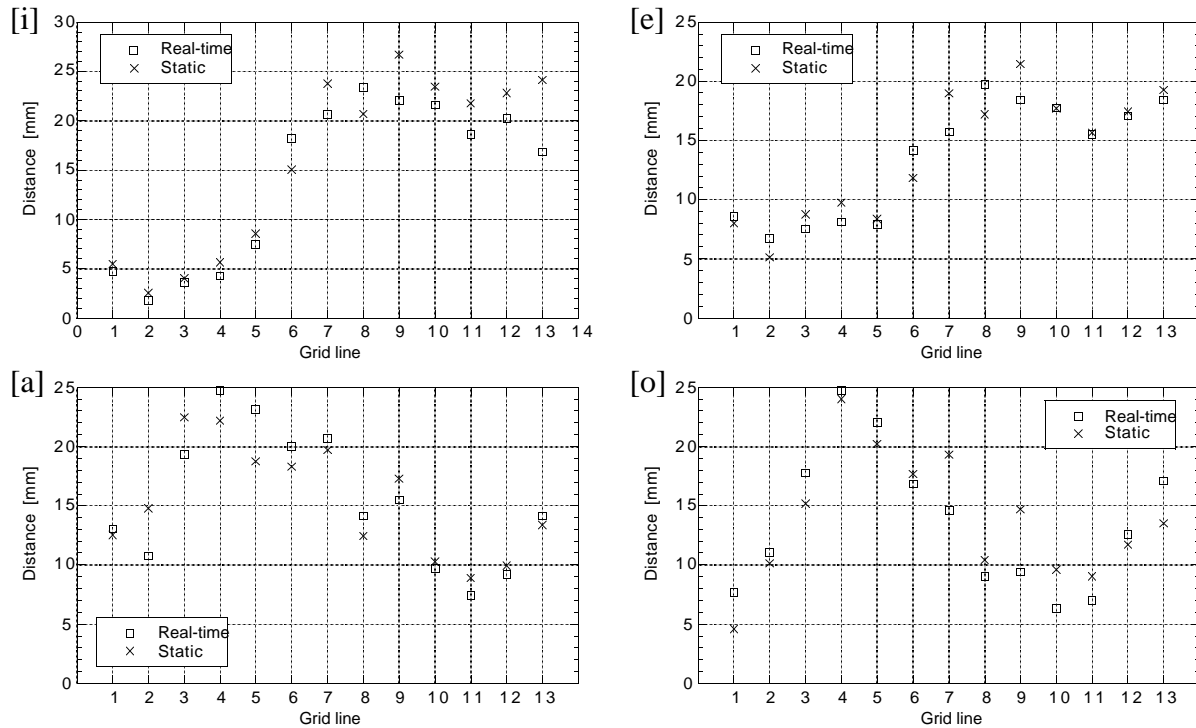


Figure 3: Comparison of the midsagittal distances obtained with the measurement grid on four vowels in the real-time acquisition (squares) and static acquisition (crosses) for the male speaker.

accurate segmentation of the speech signal. Solutions are under investigation to improve signal to noise ratio and to reduce noise level with signal processing.

Real time MRI can be compared with other techniques:

- Cineradiography provides a higher number of image per second and a much sharper image resolution, but is limited to sagittal projection and is dangerous for health.
- Electro-magnetography and microbeam allow the tracking of fleshpoint located usually in front cavity of the vocal tract and in the mid-sagittal plane. The study of movements in the pharynx is for example not possible.
- Dynamic MRI rely on numerous repetition of the same sequence to reconstruct the impression of a movements in time (Foldvik et al. 1993). Moreover, the alignment of a speech signal with the so-obtained images has to handle with care, given the fact that the reconstructed sequence do not correspond to any individual production (Shadle et al. 1999).

Real time MRI allows to study the dynamics of vocal tract deformation in any plan. This permits the collection of new data and gives new perspectives to study co-articulation processes, even if the speed of image acquisition is still abit slow.

5 REFERENCES

- Baer T., J.C. Gore, L.C. Gracco and P.W. Nye. (1991). Analysis of vocal tract shape and dimensions using magnetic resonance imaging : vowels. *J. Acoust. Soc. Am.* 90, 2. 799-828.
- Demolin D., T. Metens and A. Soquet. (1996). Three-dimen-

sional measurements of the vocal tract by MRI. *Proc. ICSLP-96, Philadelphia, USA.* 272-275.

Demolin D., M. George, V. Lecuit, T. Metens, A. Soquet and H. Raeymaekers. (1997). Coarticulation and articulatory compensations studied by dynamic MRI. *Proc. Eurospeech 97, Rhodes, Greece.* 31-34.

Foldvik A. K., U. Kristiansen, J. Kværness. (1993). A time-evolving three-dimensional vocal tract model by means of magnetic resonance imaging (MRI). *Proc. Eurospeech 93.* 557-558.

Haase A. et al. (1986). *J.Magn. Res.* 67. 258.

Mansfield P. (1997). *J.Phys C.* 10. L55.

Metens T., D.Demolin, M George, V Lecuit, A Soquet, H Raeymaekers, C.Matos. (1997). Ultra Fast Subsecond Lolo TSE For Continuous Real Time Imaging Of Articulators Movements Involved In Speech Production. *ISMR 5th Meeting, 12-18 April 1997, Vancouver B.C., Canada.* #1832.

Shadle C.H., M. Mohammad, J.N. Carter and P.J.B. Jackson. (1999). Multi-planar dynamic magnetic resonance imaging: new tools for speech research. *Proc. ICPhS-99, San Francisco, USA.* 623-626.

Van Vaals J., G. Van Yperen, A. Hoogenboom and M. Duivestijn. (1994). Local Look (LOLO) : Zoom fluoroscopy of a moving target. *Proc. of the First SMR Meeting, Dallas* 38.

This research is supported by the convention ARC "Dynamique des système phonologiques" 98-02, n°226.