MRI (MAGNETIC RESONANCE IMAGING) FOR FILMING ARTICULATORY MOVEMENTS.

A.K. Foldvik*, O. Husby*, J. Kværness**, I. C. Nordli*, and P.A. Rinck**

*Department of Linguistics, University of Trondheim, Norway
** MR-Center, Medical Section, University of Trondheim, Norway

ABSTRACT

The range of applications of magnetic resonance imaging is rapidly increasing. Among the promising new techniques is cine-imaging of articulation which offers non-risk insights into speech production.

1. INTRODUCTION

Since speech typically consists of constantly and quickly changing configurations of the vocal tract, imaging of the vocal tract is of interest to speech research.

Cinegraphic X-ray techniques, X-ray computerized tomography, and ultrasound techniques are all in use. However, methodological problems such as side effects of radiation in connection with X-ray and inability to penetrate bone and air in ultrasound studies have caused problems in connection with these techniques.

The advent of magnetic resonance imaging (MRI) opened a new avenue for speech research and it was thought that this method would overcome the disadvantages of earlier techniques [1,5]. MRI shows not only the tongue but all articulators clearly and creates highly accurate three-dimensional images. Its disadvantages lie in the fairly long acquisition times, even when subsecond imaging techniques are used.

2. INSTRUMENTATION

We have used both medium and high field MR systems operating at 0.5 and 1.5 Tesla (Philips Gyroscan S5 and S15; Philips Medical Systems, Eindhoven, The Netherlands) with both head and circular surface coils.

Transverse and coronal spin-echo (SE) sequences were acquired as scout views for double oblique angulation of the desired 5 to 8 mm sagittal scans through the articulators. The sagittal image was used as scout for double oblique transverse and coronal images at different angles through the articulatory tract which provided basis for volume calculations.

For the actual data acquisition either gradient echo, Fast Field Echo, (FFE) or subsecond imaging (SS) sequences were used. In the FFE sequences echo times (TE) ranged from 6 ms to 15 ms, total acquisition times from 3.5 s to 12 s. Acquisition matrix size was 128 times 256, reconstruction matrix size 256 times 256. The acquisition time in SS sequences was 450 ms for a 90 times 128 matrix. Image acquisition could be repeated every 1.5 seconds.

The desired time resolution for speech production imaging is in the order of 10 to 20 ms to cover the movements of the articulators satisfactorily. Subsecond imaging does not provide such speeds. Therefore we returned to FFE sequences with their better signal-to-noise ratio and higher spatial resolution.

3. SUBJECTS AND METHODS

Due to the special and specific conditions under which articulation had to be produced, only phonetically trained subjects were chosen for these studies. A number of sounds were used for interobserver control, others for intraobserver control, and some were carefully selected as markers of certain Norwegian dialects.

Images were acquired both in supine and prone position to compare a possible impact of gravity on the articulators.

The pronunciation was performed in two stages: First a warm-up stage where a predetermined word containing the wanted sound was repeated several times, followed by a five second count-down. Then the speaker "froze" the predetermined word in the middle of the desired sound with a glottal stop, the image data was acquired, and the pronunciation was continued. This procedure allowed the production of static images of certain sounds.

Since speech is a dynamic procedure we wanted to get cine representations of the articulation of complete words.

Since we had no device to trigger the acquisition from spoken sounds, we synchronized an audible beep tone with an artificial ECG that triggered the informant and instrument respectively. The beep was transmitted to the informant through a plastic tube headset as used by airlines. The ECG was picked up by the MR system's telemetry receiver (Hewlett-Packard) and the acquisition was run similarly to a routine cardiac examination.

This meant repeating the word the same number of times as the phase encoding gradients and the spatial resolution. Typically we used 90 to 160 gradient steps resulting in an acquisition time between 1 and 4 minutes.

4. RESULTS

Findings from our MRI studies of Norwegian pronunciation [2,3,4] show that sounds which traditionally have been described as palatals are laminal alveolars and laminal postalveolars, retroflex sounds are in fact apical alveolar sounds, velar stops and nasals are postpalatals even in front of open back vowels, and Norwegian front vowels [i: e: æ:] traditionally described as close, half-close and half-open respectively, have virtually the same position of the front part of the tongue, but with a marked difference in the degree of tongue root fronting (for [i:]) and tongue root retraction (for [æ:]).

5. DISCUSSION

MRI offers the possibility of recording the dynamics of speech and also has possibilities as regards vocal tract shape and volume mapping. The imaging method itself has to be improved. The inherent problem is related to time resolution. Subsecond imaging does not solve the problem since ideally acquisition time will never be short enough and the signal-to-noise ratio is not good enough.

The triggering can be improved in several ways; e.g. by the use of a microphone to pick up the speech and trigger the MR system, or by a sensor to pick up movements from the lips, chin or larynx.

To reduce acquisition time and dependence on cooperation with the informant, retrospective gating seems to be a promising method. This method could also lead to a further improvement in time resolution.

REFERENCES
[1] BAER, T., GORE, J.C., BOYCE, S., NYE, P.W. (1987), "Application of MRI to the Analysis of Speech Production", *Magnetic Resonance Imaging*, *5*, 1-7.
[2] FOLDVIK, A.K., HUSBY, O., Kværness, J. (1988), "Magnetic Resonance Imaging.", *Proceedings Speech '88*. 7th FASE Symposium, Edinburgh, 423-428.
[3] FOLDVIK, A.K., HUSBY, O., Kværness, J. (1989), "An Investigation of Norwegian Vowel Articulation by Means of Magnetic Resonance Imaging (MRI)." *Studies in Languages, 14,* 436-441. Joensuu.
[4] FOLDVIK, A.K., HUSBY, O., Kværness, J., Nordli, I.C., Rinck, P.A. (1990), "Investigation of Articulation by Magnetic Resonance Imaging.", *Society of Magnetic Resonance in Medicine. Proceedings.* Ninth Annual Meeting, 115.New York.
[5] VON WEIN, B., DROBNITZKY, M., KLAJMAN, S. (1990), "Magnetresonanztomographie und Sonographie bei der Lautgebung", *Fortschr. Röntgenstr. 153,* 408-412.

We intend to show our MRI films of articulation of Norwegian in connection with the presentation of this paper.