THREE-DIMENSIONAL ULTRASOUND AND MAGNETIC RESONANCE IMAGING: A NEW DIMENSION IN PHONETIC RESEARCH.

A. K. Foldvik, Dept. of Linguistics, University of Trondheim, Norway

U. Kristiansen, Dept. of Acoustics, University of Trondheim, Norway

J. Kværness, MR-Centre, Medical section, Trondheim, Norway

A. Torp, Dept. of Computer Systems and Telematics, University of Trondheim, Norway

H. Torp, Dept. of Biomedical Engineering, University of Trondheim, Norway

ABSTRACT

This is an overview of the use of magnetic resonance imaging and ultrasound to produce three-dimensional models of the tongue and the vocal tract.

INTRODUCTION

In phonetic descriptions of sounds we are all used to midsagittal views of the vocal tract showing tongue shape and distance from the roof of the mouth. If a sagittal plane e.g. 15mm away from the mid-line for the same sound was chosen, we would get a very different view. Those midsagittal figures are convenient for pedagogical purposes but give no information about the shape and position of the tongue off the midline, e.g. bunching or flattening of the sides of the tongue, and may therefore not necessarily capture important differences between two articulations which mid-sagittally may look quite similar.

Many researchers with an interest in physiological phonetics, who so far have not been too impressed with acoustic vocal tract models, are working on supplying accurate three-dimensional models of the oral and nasal cavities using different techniques.

Even if x-ray techniques used for phonetic research have low radiation levels, techniques with no known health risks like ultrasound and magnetic resonance imaging (MRI) attract interest among researchers in many countries. We will give some examples of advances with these two techniques in phonetic research during recent years.

MRI.

Rokkaku, Hashimoto, Imaizumi, Niimi and Kiritani [1] were among the first to publish phonetic research results based on MRI. The MR images from these first years were fuzzy and were results of dauntingly long acquisition times where subjects would often try to hold an articulatory posture for more than one minute while lying on their backs in the narrow and extremely noisy MR tunnels of that time. The resulting fuzziness in the pictures was an aggregate result of inability to keep the articulators still, too thick picture cuts which might also be laid at acute angles of e.g. the tongue, and acquisition times which ideally should have been even longer to produce sharper images. With recent progress in MR-technology acquisition times of 1s or less for good quality images with a 5mm cut are unproblematic. This means that prolongable sounds like vowels, fricatives, laterals and nasals can be studied by means of MR.

In principle the picture plane, normally 5mm thick, can be placed at any angle, but to increase sharpness it is important that it is laid as close to 90° as possible on the surface of the structure to be imaged.

At first sight an x-ray and MR-picture look similar, but structures which contain little or no hydrogen like teeth and bones do not show up on a MR-picture, while cartilages and particularly soft tissues like the velum and tongue do. One small advantage in favour of MR is that dental fillings which tend to obscure the shape of the tongue in x-ray pictures do not show up on an MR-image. Yang and Kasuya [2] ingeniously solved the problem with no-showing dentition: They put dental impressions of the subjects in water, and took very accurate MR-images of the impressions. Coating mediums of the teeth have been tried with no great success so far.

Recently both mid-sagittal MR films of articulatory movements as well as dynamic three-dimensional vocal tract models by means of MRI have been made [3, 4]. See fig. 1, 2 and 3.

In X-ray technique enormous progress has been made. When we compare the improvements in x-ray techniques during nine decades from e.g. the misty pictures that Grunmach [5] published in 1907 to the razor-sharp xeroradiographic photographs of today [6] with progress in MRI during less than one decade, it does not seem too risky to predict that we will in the not too distant future see improved quality MR images of reduced slice thickness based on shorter acquisition times than today. We also envisage 3D films of articulatory movements using improved computer algorithms for air-tissue boundary detection.

ULTRASOUND

Compared to using MRI, ultrasound equipment is from our experience cheaper and easier to get access to, but there are limits to what you can do with it. Therefore anybody who has tried to record tongue movement and tongue shape for speech sounds by means of ultrasound must be impressed with Stone and Lundberg [7] who have presented three-dimensional reconstructions of the tongue surface based on speech sounds which were sustained for 10s.

One of the limitations to ultrasound equipment is that it does not register the tip of the tongue if the apex is so far forward in the mouth that the ultrasound transducer waves are unable to reach it, but instead register the sublingual air/tissue borderline. Also edge/surface detection is not always unproblematic. Sounds produced with contact between the tongue and the roof of the mouth pose another problem since tongue edge detection then becomes difficult. Therefore we have so far only succeeded

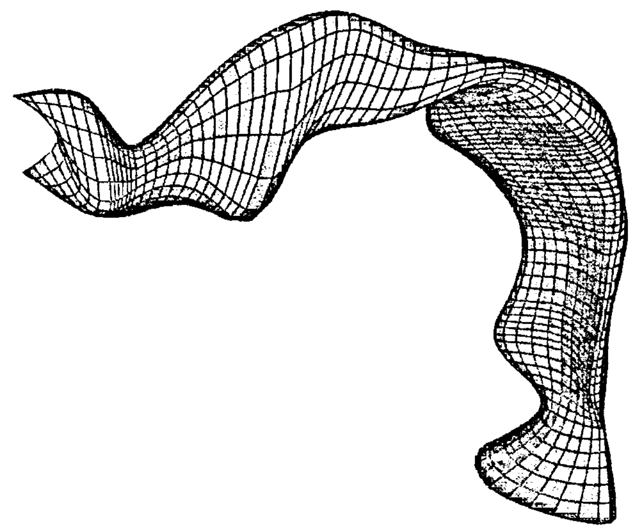

Figure 1. Lateral view of 3D model of the vocal tract tube for [ɑ]. The mouth is on the left. From an MR imaging based on 200 repetitions of the utterance [ɑi]. See [4].

in making ultrasound-based 3D models of the surface of the tongue for sounds where the tongue is in a central or back position. See figure 4.

Obviously ultrasound is a technique which offers tremendous possibilities for phonetic research. And it is only a matter of time before we see the first time-evolving three-dimensional models of the tongue based on ultrasound images with far better quality and shorter acquisition time than we use today.

During this 1/3 plenary session we intend to show examples of dynamic 3D models of the vocal tract.

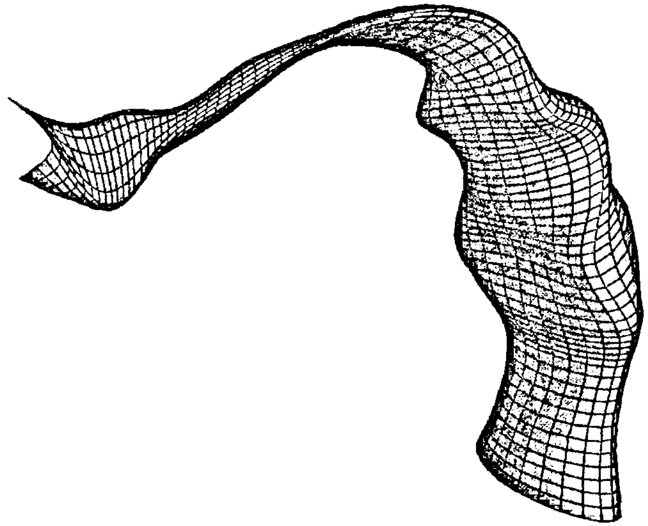

Figure 2. Lateral view of 3D model of the vocal tract tube for [i]. The mouth is on the left. From an MR imaging based on 200 repetitions of the utterance [ɑi]. See [4].

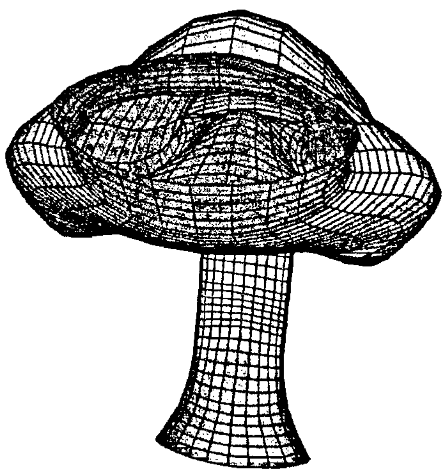

Figure 3. Frontal view of 3D model of the vocal tract tube for [ɑ]. From an MR imaging based on 200 repetitions of the utterance [ɑi]. See [4].

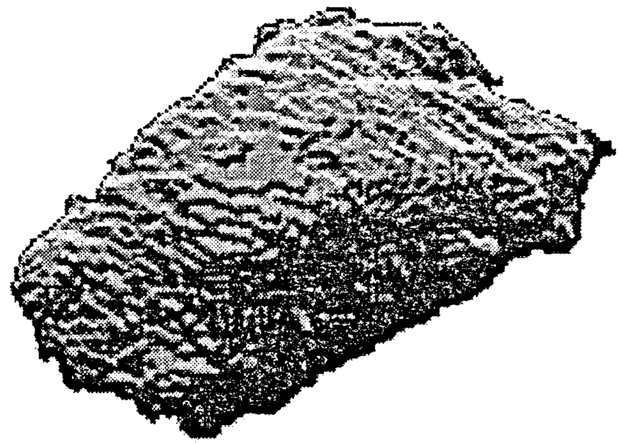

Figure 4. This figure shows a three-dimensional reconstruction of the surface of the tongue for [ɑ:]. Semi-lateral view of the tongue with apex to the left. The sulcus in the back part of the tongue can be seen. The ultrasound images were collected with a Vingmed Sound CFM-800 ultrasound scanner. The ultrasound probe was a mechanically steered annular array with 5 elements, producing a sector scan image with an opening angle of 90 degrees. The ultrasound frequency was 5 MHz. Resolution was 1 mm in the axial direction (along the ultrasound beam), and 3 mm in the transversal direction (normal to the ultrasound beam). The 2D image-frame rate was 35 frames/sec.The ultrasound probe was mounted in a custom made tilting device, using a stepper motor tilting the probe 0.7° per 2D image plane.

REFERENCES

[1] Rokkaku, M., Hashimoto, K., Imaizumi, S., Niimi, S. and Kiritani, S. (1986), "Measurements of the three-dimensional shape of the vocal tract based on the magnetic resonance imaging technique". *Annual Bulletin Research Institute of Logopedics and Phoniatrics*, vol. 20, pp.47-54.

[2] Yang, C.-S. and Kasuya, H. (1994), "Accurate measurement of vocal tract shapes from magnetic resonance images of child, female and male subjects". *Proc. ICSLP 94*, pp. 623-626.

[3] Foldvik, A.K., Husby, O., Kværness, J., Nordli, I.C. and Rinck, P.A. (1990), "MRI (magnetic resonance imaging) film of articulatory movements.", *ICSLP90*, pp. 421-422.

[4] Foldvik, A.K., Kristiansen, U., Kværness, J. and de Bonnaventure, H. (1993), "A time-evolving three-dimensional vocal tract model by means of magnetic resonance imaging (MRI).", *Proceedings Eurospeech '93*, vol. 1, pp. 557 - 558.

[5] Grunmach, E. (1907), "Die Röntgentechnik zur Untersuchung der Mund-, Schlund- und Nasenhöhle bei der Phonation", *Archiv für Laryngologie*, vol. 19:3, pp. 405-407.

[6] Laver, J. (1994), *Principles of Phonetics*. Cambridge University Press.

[7] Stone, M. and Lundberg, A. (1994),"Tongue-palate interactions in consonants vs. vowels." *ICSLP94*, pp. 49-52.