ESTIMATION OF VOCAL TRACT AREA FUNCTION FROM MAGNETIC RESONANCE IMAGING: PRELIMINARY RESULTS

Bernd J. Kröger[1], Ralf Winkler[2], Christine Mooshammer[2], and Bernd Pompino-Marschall[2]
[1]*Institute of German Language and Linguistics, Humboldt-University Berlin, Germany*
[2]*Research Centre of General Linguistics (ZAS), Berlin, Germany*
Email: kroeger, winkler, timo, bobby@zas.gwz-berlin.de

ABSTRACT

A method has been developed for three-dimensional reconstruction of the vocal tract shape and for the calculation of area function from magnetic resonance imaging (MRI). MR images were acquired and analyzed for 6 German long vowels uttered by one subject. The resulting vocal tract area functions are the basis for the calculation of formant frequencies. Conformity with acoustically measured formant frequencies can be stated. A principal component analysis applied to the vocalic area functions derived from these articulatory data indicates that its variance can be described by few main factors.

1 INTRODUCTION

Magnet resonance imaging (MRI) has increasingly been used in speech production research in the last years in order to acquire the complete three-dimensional shape information about the vocal tract structures. This paper presents a method for three-dimensional reconstruction of the vocal tract and calculation of area function exclusively based on lateral MR images.

2 METHOD

MR images were acquired and analyzed for 6 German long vowels ([iː], [eː], [ɛː], [aː], [oː], and [uː]) uttered by one subject (JD) using a Philips Gyroscan NT scanner at Radiology of Virchow Klinikum Berlin. A 21 slice series of 3.5 mm thick non-contiguous parallel sagittal sections with inter-image space of same thickness was gathered per sound (5 mm thickness for [uː]). Image acquisition takes 21 seconds per sound. Accordingly only one articulation process is needed per vowel. The analysis procedure comprises three main steps as described below.

In a first step the vocal tract airway centerline is calculated using a speaker-dependent but articulation-independent grid (*initial* grid system, see figure 1a-d). The grid location is based on a square defined by three points, i.e. the midsagittal location of a lower and an upper point of the vertebral column in the cervical region and of the highest point of the palate (black dots in figure 1a). The palatal point defines the length of the square and thus the center of rotation for the rotational part of the grid system (arrow in figure 1a). Two parallel grid line parts are attached, one to the left side and one to the bottom. The former ends at the lips and the latter ends at the larynx. The grid line distance in these parallel parts is 0.27cm (0.35cm for [uː]). Depending on the degree of lip protrusion and on the differences in larynx height our grid system comprises 50 to 60 grid lines per vowel.

The air-tissue boundaries were determined separately for each grid line. A semiautomatic procedure has been established for detecting two points representing the air-tissue boundary for each grid line. A gray scale value was chosen representing the threshold between air (dark) and tissue (light) and the procedure is based on the low pass filtered gray scale values for the succession of pixel locations defined by the appropriate grid line. Subsequently the center of the two air-tissue points has been calculated for each grid line. The profile of these center points has been smoothed manually and defines the airway path centerline of the vocal tract for each vowel (small white dots in fig. 1a-d).

In a second step a set of grid lines was calculated. Each grid line intersects with a center point and is perpendicular to the centerline given above (*final* grid system, see figure 1e-h). In order to obtain a homogenous final grid system, a smoothing of the grid line orientation is necessary. This smoothing is done by low pass filtering of the orientation values (i.e. angles of the grid lines in the midsagittal plane) for the succession of the grid lines. The resulting set of final grid lines defines planes for the estimation of cross sectional vocal tube areas, i.e. planes perpendicular to the midsagittal plane and nearly perpendicular to the vocal tract airway path (vocal tube).

In a third step the vocal tube width was estimated for each sagittal (i.e. lateral) MR image and for each grid line of the final grid line system. The algorithm for the determination of the air-tissue boundaries is identical to that used for the estimation of the air-tissue boundary described above. Figure 2a and 2b shows the resulting cross sectional areas perpendicular to the vocal tube mid line for [aː] and [ɛː] for each 8th grid line starting at the glottis and ending at the lips, (see grid lines marked by asterisks in fig. 1f and 1g). The piriformis sinuses (grid line 8), the contour of the epiglottis (lower air-tissue boundary, grid line 16 for [ɛː]), of the uvula (upper air-tissue boundary, grid line 24), of the roots of the molars (dark gray regions in fig 2c), and of the artificial outward lip areas (dark gray regions in fig 2d) can be identified. The complete set of air-tissue boundary points leads to a three-dimensional reconstruction of the vocal tract airway path (figure 3). These vocal tube structures have been opened in the palatal and lip region according to the cutoff of air-tissue points if roots of the molars or outward lip contours were detected in the cross sectional area contours.

Figure 1. (below) Midsagittal planes for vowel [iː], [ɛː], [aː], and [uː] (top to bottom). Left side: Initial grid system for the extraction of vocal tube centerline. Right side: Final grid system for estimation of cross sectional vocal tract areas. The lines n*8 are marked for [ɛː] and [aː] (see figure 2).

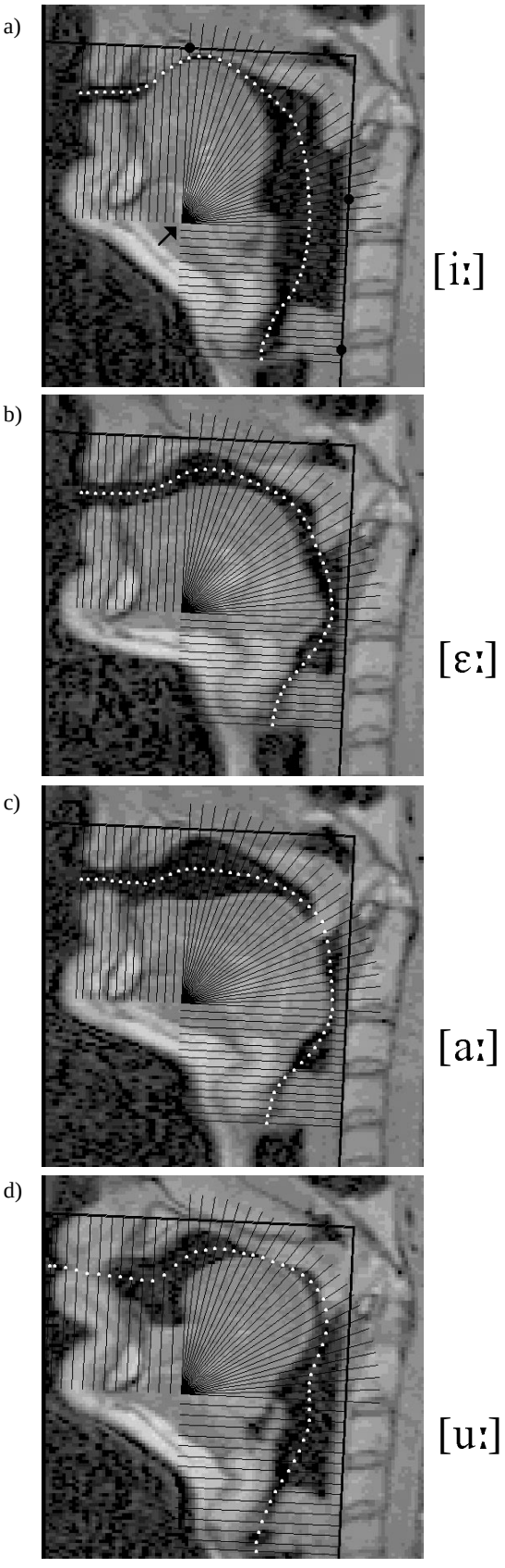

Figure 1

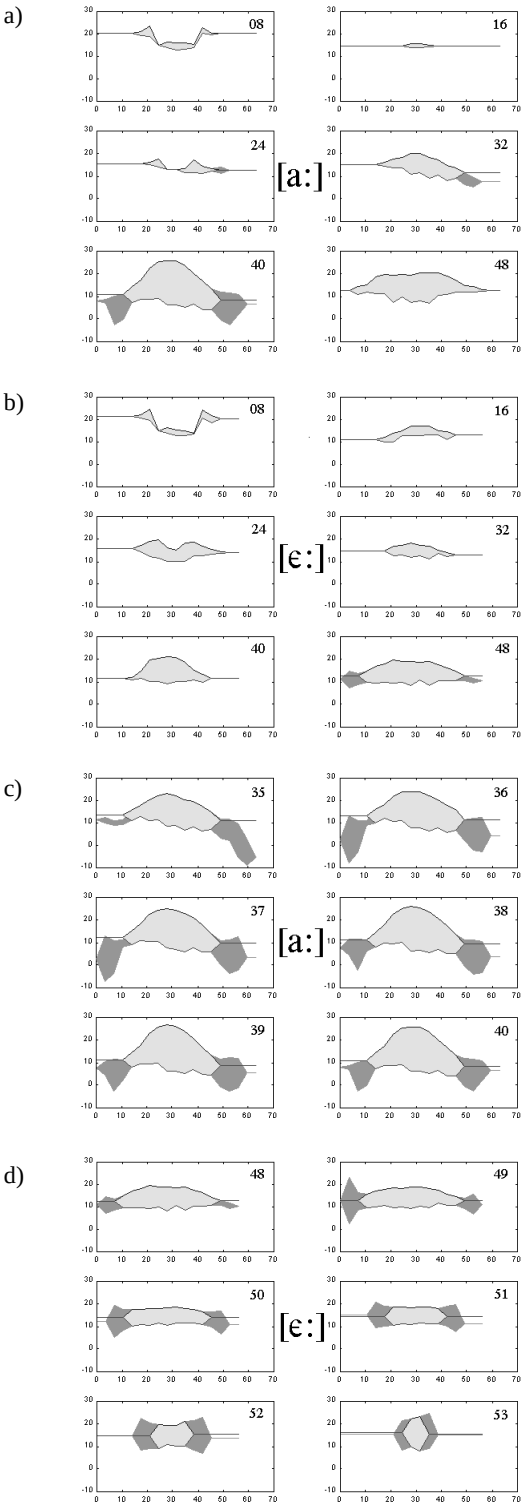

Figure 2. (left side) Cross sectional area contours perpendicular to the vocal tract airway midline for [aː] and [ɛː]. The dark gray areas indicate the location of teeth or the artificial volumes resulting from the outward lip contours. The number in the upper right corner indicates the number of the grid line within the final grid line system (figure 1e-h). Abscissa and ordinate: distance in mm.

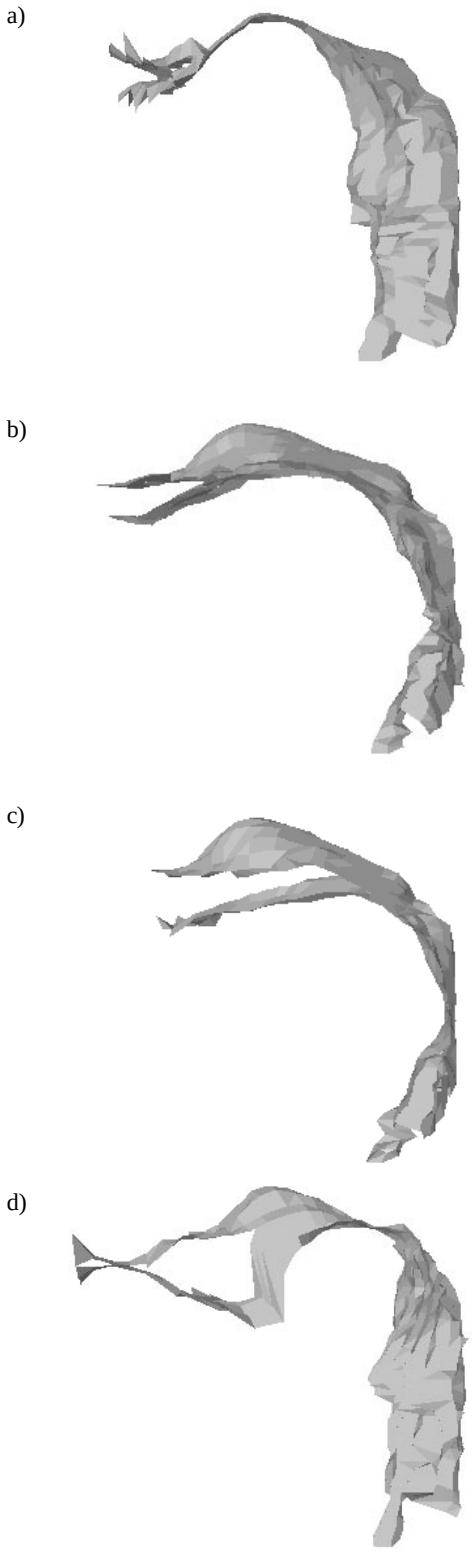

Figure 3. Three-dimensional contours of the airway paths for [iː], [ɛː], [aː], and [uː] (top to bottom). VRML-data of these contours are available on the CD-ROM.

Since no digital cast of the teeth was available in this experiment, teeth volume was approximated by inspecting the cross sectional area contours. The location of the roots of the molars can be identified (dark gray areas in figure 2c, i.e. grid line 35 to 40 for [a:]) and the appertaining areas have been subtracted for the calculation of the acoustically relevant area function. Similar artificial side areas indicate the outward contours of the lips in the cross sectional contours of the areas for the most fronted grid lines (dark gray areas in figure 2d, i.e. grid line 48 to 53 for [ɛ:]). The mouth termination of the vocal tract tube was constructed by taking into account the frontal plane tangent to the lips and the points where the inner lip contours intersect (i.e. points of intersections of the inner lip contours near the location of the corners of the mouth). The former bas been estimated from the midsagittal contour of the lips (figure 1), the later from the cross sectional area contours of the most fronted grid lines. Different models exist for the calculation of the mouth termination (Lindblom and Sundberg 1971, Mermelstein 1973). In this approach the half-distance point between frontal plane and intersections of the inner lip contours defines the termination plane of the vocal tract tube.

3 RESULTS

The derived vocalic area functions (figure 4) serve as input for calculation of the formant frequencies. The transfer function of the vocal tract was calculated using the frequency domain transmission line approach of Sondhi and Schroeter (1987). Formant frequency values were calculated by peak picking and parabolic interpolation of the first three formants. In addition natural speech signals were recorded in a separate session. The subject was set in a supine position in a sound treated chamber in order to simulate the MRI-measurement conditions as closely as possible. Each vowel was uttered one time (duration: 20s each) and formant frequencies were estimated using an average spectrum analysis procedure at six time instants for each vowel. Calculated formant frequency values and acoustically measured formant frequency values match approximately (figure 5). The mean difference is around 8.9% for F1, 10.4% for F2, and 6.1% for F3 for these six vowels.

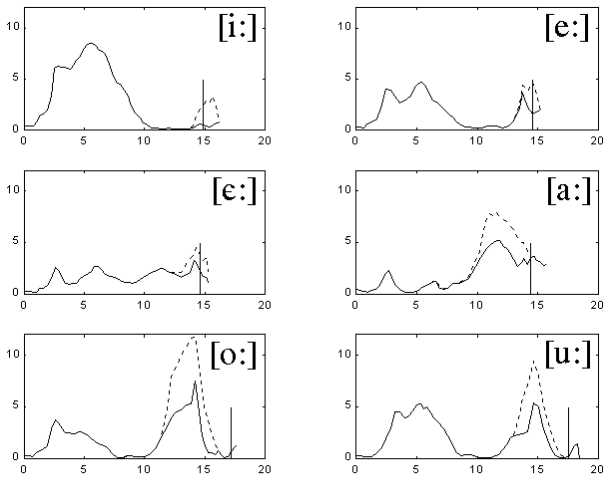

Figure 4

Figure 4. (above) Area functions of the six vowels including (dashed line) and excluding (solid line) teeth and artificial volumes resulting from the outer lip contours. The vertical line indicates the termination of the vocal tract (see text). Abscissa: length in cm. Ordinate: area in cm^2.

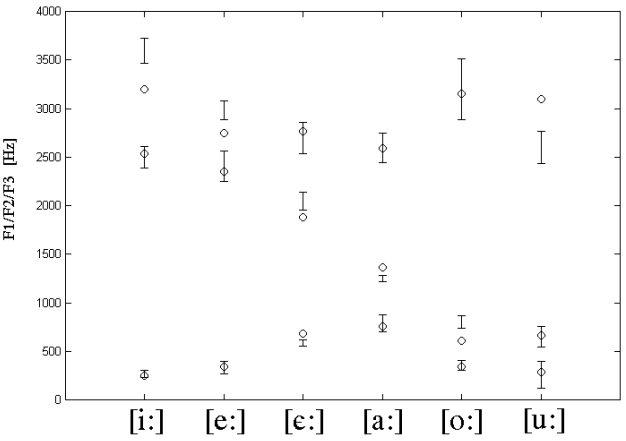

Figure 5. Calculated formant frequencies (circles) and 99% confidence intervals of the acoustically measured formant frequencies (bars) for the first three formants.

Furthermore a principal component analysis was performed for these area functions indicating that its variance can be described by few main modes. The cumulative percentage of variance reaches about 71% for one, 89% for two, and 96% for three modes. The shapes of these first three modes are qualitatively similar to those derived by Story et al. (1998).

4 CONCLUSION

A method for three-dimensional reconstruction of the vocal tract shape and calculation of area function exclusively based on lateral MR images has been developed. Formant frequency values calculated from area function match approximately those acoustically measured. In further studies data of a digital cast of the teeth should be included. A long-term goal is to develop a three-dimensional articulatory model as part of a comprehensive model of speech production.

5 LITERATURE

Lindblom, B., Sundberg, J. (1971): "Acoustical consequences of lip, tongue, jaw, and larynx movements", *Journal of the Acoustical Society of America* **50**, 1166-1179.

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

Sondhi, M.M., Schroeter, J. (1987): "A hybrid time-frequency domain articulatory speech synthesizer", *IEEE Transactions on Acoustics, Speech, and Signal Processing* **ASSP-35**, 955-967.

Story, B.H., Titze, I.R., Hoffman, E.A. (1998): "Vocal tract area functions from magnetic resonance imaging", *Journal of the Acoustical Society of America* **100**, 537-554.