DETAIL IN VOWEL AREA FUNCTIONS

Christine Ericsdotter

Stockholm University
ericsdotter@gmx.de

ABSTRACT

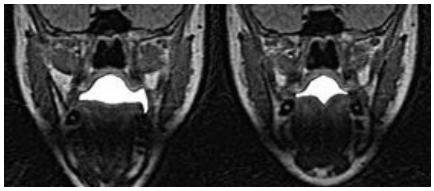
This paper presents some results and a small follow-up investigation from an MRI study of vowels [3], in which classical distance-to-area equations [5] were evaluated for implementation in sagittal view articulatory modelling. It was shown that an articulatorily more detailed application of the conversion rules improved the accuracy of the predicted areas, but that this increased realism failed to improve acoustic performance, if midline derivation and vocal tract termination points were kept the same. These results are discussed in relation to articulatory modelling in linguistic research. Work funded by the NIH (R01DC02014) and Stockholm University (SU617023001).

1. INTRODUCTION

Predictions of cross-sectional areas from mid-sagittal vocal tract profiles, in articulatory models as well as on raw articulatory images, typically encounter problems in the pharynx. In earlier work, these problems had their origin in the difficulties of achieving frontal and axial articulatory data from the region. In more recent work, the problems rather reside in how to account for the multiplicity of shapes that the pharynx takes during speech. Only the posterior wall seems to be fixed, while the anterior and the lateral walls continuously reshape the resonator [1, 3, 11, 12], both actively and as a consequence of larynx height. A given mid-sagittal distance does often not have a unique corresponding area, but multiple possible areas depending on the identity of the articulation (Figure 1**;** Figure 3**,** left). The mid-sagittal dimension at a given place in the pharynx is therefore not sufficient for accurate prediction of the cross-sectional area at that vocal tract place.

The classical method for predicting areas from mid-sagittal dimensions is that presented by Heinz and Stevens [5]. It relates, at each point in the vocal tract, the mid-sagittal distance d to the area A of the cross-section at that point by $A = \alpha d^{\beta}$, where α and β are constants depending on speaker and position along the vocal tract. The predictability of A from d was proposed using a simplified assumption that the tongue has a flat surface, and that the opposite vocal tract wall has a fixed, parabolic shape. Despite the fact that cross-sectional areas along the vocal tract do not always conform to the shapes implied by this equation, the conversion formula has proven useful and economic in articulatory models over the years. Given new knowledge, particularly on the pharynx as discussed above, more elaborated geometrical prediction routines are however increasingly being used (e.g. [1, 13]). Despite recent developments, the present study adopts the classical method, and suggests how, by adding vowel identity dependence to the speaker and vocal tract place dependence, it can still be applied with realistic area estimations for vowels. It also presents a formant comparison between area functions based on more specific and more general application of the classical method.

Figure 1. Example of articulations where the mid-sagittal distance is the same, but the cross sectional areas differ in shape and size (by 2.5 cm^2). Male subject, velar region, vowels œ (left) and ʉ̟ (right).



2. DATA

The data consisted of one mid-sagittal set and one axial/coronal set of MR images from the vocal tract (plane distribution is displayed in Figure 2), combined with simultaneous audio (with a fiberoptic microphone) and video recordings. These data were collected from two Swedish speakers at the Unité de Résonance Magnétique de l'Hôpital Erasme in collaboration with Université Libre de Bruxelles, in June 2000. Extended articulation of eleven Swedish vowel sounds composed the speech materials. An important step in the data post processing was

integration of teeth contours in the MR images. Articulatory measurements were made of mid-sagittal distances and cross-sectional areas from the whole vocal tract, together with estimations on the vocal tract termination points. Acoustic and auditory analyses were made of the sound recordings. A detailed description of the data acquisition and processing can be found in [3].

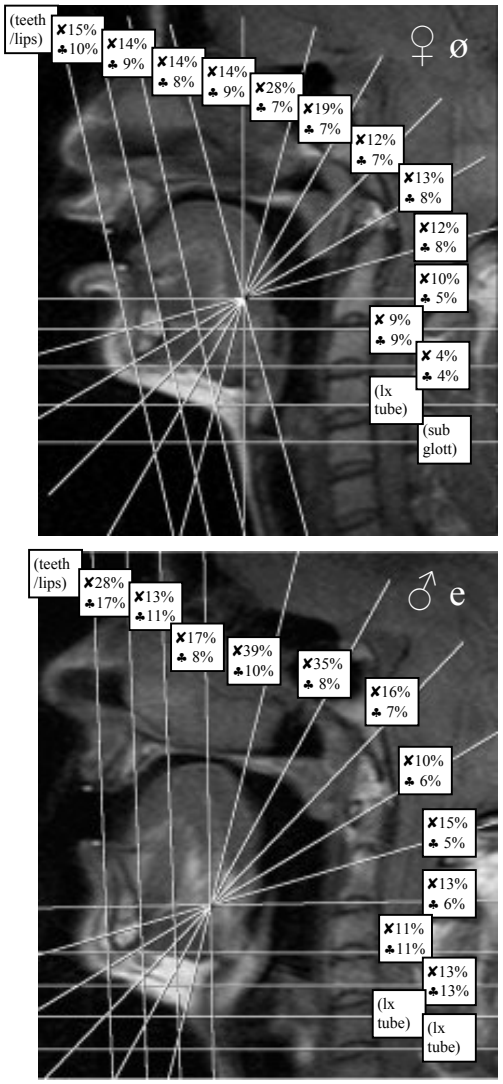
3. DISTANCE-TO-AREA PREDICTIONS

In each subject, for each vocal tract slice, one power equation was derived from measured distances and areas. Because of the fact that there were multiple areas to a given distance, as discussed above and exemplified in Figure 1, these equations produced substantial errors in individual cases. Further refinement of the predictions was therefore sought by derivation of power equations from smaller, phonologically and/or geometrically (anatomically) motivated subsets of data, defined by visual inspection of the actual shapes of the cross sectional areas. This move solved, to a great extent, the problem of multiple areas to a given distance (exemplified in Figure 3), also in the lip area.

The area predictions based on the latter equations (i.e. vowel, place and speaker specific), generally showed a closer match with observations, with average absolute percentage deviations from observations smaller than 10% at most vocal tract places (Figure 2). Comparison between these averages however gives a partly distorted picture, since the vowel dependent subsets were based on a lower, and varying, number of vowel tokens. The most problematic region was the posterior oral cavity/velopharynx, but the oral region was not ideally predicted either. The laryngopharynx was unproblematic in the female subject, who showed negligible larynx movements, but more problematic in the male, who moved his larynx considerably between vowels. Analysis of the errors for each vowel, rather than for each vocal tract place, showed that large percentage errors often derived from small areas having been wrongly predicted, although without losing their identity as "small".

To determine which equation to use for a given subject, at a given place and for a given articulatory setting, a vowel identification method was defined. By taking a holistic view of the articulatory profile, articulatory interrelations and constraints in the vocal tract could be used: dimensions from the mid pharynx, velar region and anterior oral cavity determined frontness and openness, and the lip depth determined rounding. From combinations of these characteristics, a number of vowel subgroups arose. This phonetically motivated prediction strategy was applied successfully to the MR images and also to X-ray images collected previously [2].

Figure 2. Sagittal view of axial and coronal MR planes and their absolute mean area prediction deviation as calculated by ✘ classic and ♣ vowel dependent distance-to-area equations.



4. AREA FUNCTIONS

The acoustically relevant midline was determined by connecting the mid points of the sagittal distances, as in [5]. A fixed larynx tube for each subject was placed according to careful estimation of the larynx height. The length of the lip section was set in relation to the front teeth, and its area

was set at the lip tangent position. The areas along the vocal tract were predicted from distances perpendicular to the midline, either using subject and place specific equations, or subject, place and vowel specific equations.

5. ACOUSIC RESULTS AND METHOD DESPECIFICATION

The acoustic response of the area functions was calculated using Formflek [7], taking internal length corrections into account, and then correcting the first formant for impedance of yielding vocal tract walls. The results were generally in good agreement with observations, but there was little difference between the prediction strategies (Figure 4). Hence, area functions based on more accurately predicted articulatory data did not produce significantly better formants. When there were deviations in the predictions, they were mainly of the same kind in the results of both strategies.

If more detail has negligible acoustic effects, the question arises, as to how much detail can be excluded before it severely affects the acoustic results. To investigate this, formant response was evaluated from area functions without place dependence, then without speaker dependence. As a last evaluation step, 193 pharyngeal distance and area points from 6 speakers collected from the literature ([4]: cinefilm via fiberoscope, 2 levels of the pharynx, 1 male subject, Swedish vowels /u o ɒ a/, 8 distance area data points; [11]: axial computed tomographs from 4 levels in the pharynx, 1 male and 1 female speaker of Swedish, vowels /u i ɒ ø/, 32 distance area data points; [12]: axial MR images at 11 levels in the low pharynx, 1 male speaker of Akan and 1 male speaker of American English, vowels /i ɪ e ɛ u ʊ/, 132 distance area data points; [9]: CT scans from the pharynx, 1 male speaker of French, vowels /i a u/, 21 distance area data points) were included in the data set. Only pharyngeal data were added, because it was available, and because oral and labial data could be expected to add less noise. These data were added to the set of 582 data points from the present study, and a power function was derived (Figure 3, right). The acoustic response from the resulting area functions turned somewhat poorer with each generalization condition (Figure 4, for lack of space only the most general condition is illustrated for comparison with the more detailed ones), but formant patterns for individual vowels were still characteristic.

6. DISCUSSION

Satisfactory acoustic patterns could be predicted from mid-sagittal profiles using specific as well as general conversion rules for obtaining the area functions, while satisfactory articulatory predictions were more dependent on specification. This asymmetry appears to reflect insensitivity of the three lowest vowel formants to articulatory detail, but the results should be seen in light of the kind of detail the conversion actually affects, and the method used for acoustic evaluation.

The conversion method was indeed originally found to be speaker specific [5], which studies evaluating it always confirm, e.g. [11]. Its impact on the area function is however limited to *absolute* area size. More acoustically crucial factors can be directly derived from the sagittal view, and they were kept constant between evaluation strategies: the mid line and vocal tract termination points did not change, which practically assured cavity lengths were preserved. The sagittal distances were also the same between conditions, and despite the problems of multiple areas to a given distance discussed in the introduction and illustrated in Figure 1**,** their capability of maintaining *relative* size identity in the area prediction must be interpreted as very strong. A more general conversion equation flattens out extreme values (Figure 3, right), especially large ones since they were fewer than the small areas in the data set, and hence big areas were decreased more than small areas were increased, which is acoustically convenient.

As for the acoustic evaluation, observed formants from an initial part of the vowel were set as reference, i.e. disregarding intrinsic errors in formant measurements and the dilemma of finding the representative patterns of sustained vowels. An acoustic reference demands a perfect sound evaluation method of area functions, which is being developed and calibrated from details in this sort of study, rather than confirming detail in it. At present, it is still possible to object to the method of modelling the lips, or the larynx, or the data sampling along the tube, or the midline derivation, or the response of the tube as a set of cylinders rather than of complex shapes.

This work was carried out within the framework of creating articulatory modeling tools primarily for use in linguistic research. Apparent acoustic resistance against some ariticulatory detail in vowels might seem as a welcome simplification factor in such a model, while in fact it is only mirroring the bluntness of the work. Small articulatory changes have crucial acoustic and perceptual effects if made at the right places [10], at the right times [6], in the right setting [8] and so on, and a useful model must be fit to adequately represent this. What is still needed is more detail on dynamic aspects of articulatory parameters conditioning the area shapes, and continuous refinement of methods for hifi acoustic evaluation of articulatory data. No model built on approximations will respond well to detail.

Figure 3. Left: example of speaker, place and vowel specified equations, largely avoiding multiple areas to a given distance. Right: equation based on data from 5 studies. Profuse instances of multiple areas to a given distance.

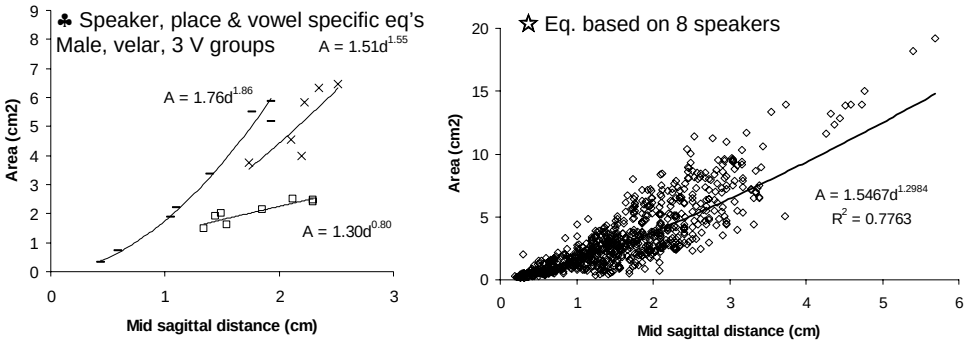
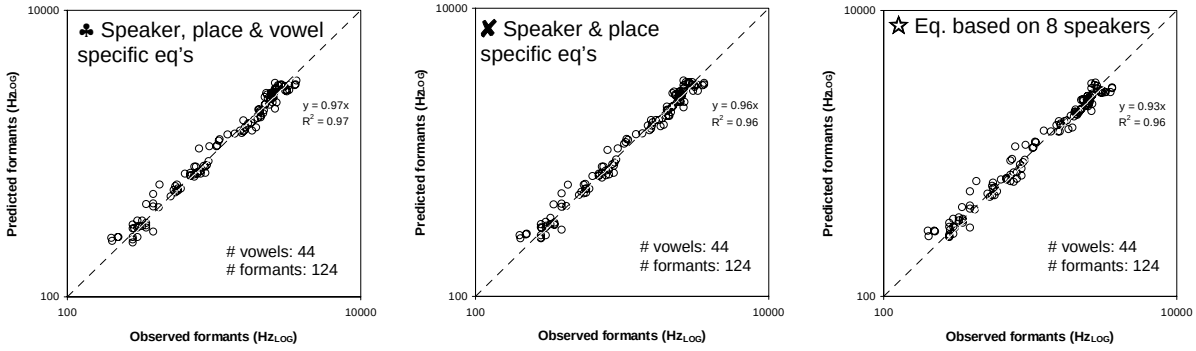


Figure 4. Observed and predicted formants, female and male speaker. From left to right: decreasing area function detail.



7. REFERENCES

[1] Badin, P., Makarov, I. S., Sorokin, V. N. 2005. Algorithm for Calculating the Cross-Section Areas of the Vocal Tract. *Acoustical Physics* 51, 38-43.

[2] Ericsdotter, C., Stark, J., Lindblom, B. 1999. Articulatory coordination in coronal stops: Implications for theories of coarticulation. *Proc. XIVth ICPhS, San Francisco,* 1885-1888.

[3] Ericsdotter, C. 2005. *Articulatory-Acoustic Relationships in Swedish Vowel Sounds.* Doctoral thesis. Stockholm University.

[4] Gauffin, J., Sundberg, J. 1978. Pharyngeal constrictions. *Phonetica* 35, 157-168.

[5] Heinz, J. M., Stevens, K. 1965. On the relations between lateral cinoeradiographs, area functions, and acoustics of speech. *Proc. 4th International Congress on Acoustics,* Stuttgart, A44.

[6] Kluender, K. R., Coady, J. A., Kiefte, M. 2003. Sensitivity to change in perception of speech. *Speech Comm.* 41, 59-69.

[7] Liljencrants, J. Formf.c. C-program. TMH-KTH.

[8] Ohala, J. J. 1993. The phonetics of sound change. In C. Jones. *Historical Linguistics: Problems and Perspectives.* London, Longman, 237-278.

[9] Perrier, P., Boë, L.-J., Sock, R. 1992. Vocal Tract Area Function Estimation from Midsagittal Dimensions with CT Scans and a Vocal Tract Cast: Modeling the Transition with Two Sets of Coefficients. *JSHR* 35, 53-67.

[10] Stevens, K. N. 1989. On the quantal nature of speech. *J. Phon* 17, 3-45.

[11] Sundberg, J., Johansson, C., Wilbrand, H., Ytterbergh, C. 1987. From sagittal distance to area. *Phonetica* 44, 76-90.

[12] Tiede, M. 1996. An MRI-based study of pharyngeal volume contrasts in Akan and English. *J. Phon* 24, 399-421.

[13] Tiede, M., Yehia, H., Vatikiotis-Bateson, E. 1996. A shape-based approach to vocal tract area function estimation. *Proc. 1st ESCA Workshop in Speech Production Modeling: From Control Strategies to Acoustics,*Autrans, 41-44.