Swedish vowels and a new three-parameter model

Fant, G.

journal: TMH-QPSR
volume: 42
number: 1
year: 2001
pages: 043-049

Swedish vowels and a new three-parameter model[*]

Gunnar Fant

Abstract

Vocal tract area functions of 13 Swedish vowels have been derived from midsagittal tracings of X-ray pictures, supported by a limited tomographic material. With a few exceptions, e.g. F2 of [u†], [o†] and [O], calculated formant frequencies show a substantial agreement with data measured during the X-ray session. The sensitivity of formant frequencies to variations in vocal tract area functions have been studied.

Observed co-variation of overall vocal tract dimensions revealed a number of dependent relationships that enable a prediction of overall length, inter-incisor distance, and asymmetries of cavity shapes from the basic specification of location and area of tongue constriction and the degree of lip-rounding. The new model thus preserves physiological constraints that make it better suited for a future adaptation to consonantal modifications than earlier three-parameter models. Nomograms of formant frequencies for systematically varied model parameters are shown. Problems related to inverse mapping, from formant frequencies to model parameters, are discussed

Introduction

In connection with an X-ray tomographic study of the vowels [**u:**], [**ɑ:**] and [**i:**] (Fant, 1964) processed by Sundberg (1969) a set of lateral views were obtained for the same male subject sustaining the vowels **[u:] [o:] [ɔ] [ɑ:] [a] [æ:] [ɛ] [e:] [i:] [œ:] [y:] [ʉ:] [ø:] [œ:] [ɵ].**

These have been used on various occasions to illustrate the Swedish vowel system, see [8], but have not been processed earlier.

Now, with the growing interest in articulatory synthesis, (Lin,1990) this material should constitute a valuable source for articulatory modelling, adding to the rather meagre available data on vowel specific VT area functions which for a long time has been dominated by the Russian vowels, (Fant, 1960).

The Swedish vowel system

Swedish has a quite rich vowel system (Fant, 1973, 1983), the orthographic base of which is three back vowels /O/, /Å/ /A/, three front vowels /I/, /E/, /Ä/, and three rounded front vowels /Y/, /U/, /Ö/. These occur in pairs of long and short vowels, thus in all 18 phonemes. Within a pair there usually exists a quality difference, which might be small or absent as in the pre-[r] allophones [ae:] and [ae] of the phoneme /Ä/ and in the pre-[r] allophones [oe:] and [oe] of the phoneme /Ö/. Of the 13 vowels selected for our study, see Figure 1, three are phonemically short [**ɔ**], [**a**], [**ɵ**] but have F-patterns significantly different from corresponding long phonemes and may accordingly be sustained. The individual vowels in Figure 1 have been oriented approximately according to their position in an F1 versus F2 plot

Observe the front vowel character of [**u:**] articulated with a lip opening more narrow than in [**y:**] and with an extra apical elevation. Whilst the [**ʉ:**] historically has advanced to an extreme high front vowel, its phonemical mate, the short vowel [**ɵ**] is quite close to the back vowel [**o:**]. In standard Swedish, the distance between the vowels [**ʉ:**] and [**ɵ**] is greater than within any other pair of a long and a short vowel.

[*] This is a revised version of Fant G (1992): Vocal tract area functions of Swedish vowels and a new three-parameter model, Proc. ICSLP-92, Vol. 1,. 807-810.

The overall relations between vowels as seen in Figure 1 are typical .One exception is the fairly large lip opening of **[y:]**. One has also reason to suspect that the articulatory targets of **[u:] [o:] [ɔ]** attained during the X-ray exposures were somewhat relaxed compared to the subject's reference conditions. This pertains both to lip opening and back tongue constrictions, and could explain the small difference between **[o:]** and **[ɵ]**.

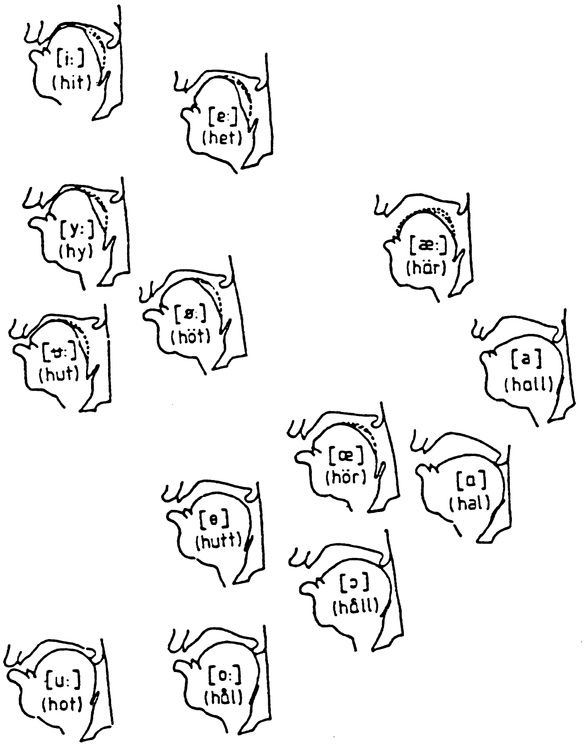

Figure 1. Lateral view of sustained Swedish vowels. From Fant (1964, 1983).

F-pattern calculations

The subject was a man of age 30, an amateur singer with barytone voice, and as judged from the collected data, of average head size. For the processing of dimensions we followed the system of Fant (1960). The basic coordinate system in the sagittal plane was thus not the usual one with horisontal layers slicing the pharynx and a joining system of radial lines for the mouth. Such a system may produce projection errors with respect to wavefronts. Instead, an outline connecting estimated center points of wavefront slices was constructed. As a zero co-ordinate along this centre line we choose its intersection with a line through the upper and lower incisors.

The total length of the vocal tract, from the assumed plane of radiation at the lips to the glottis, was typically 19.5 cm for rounded vowels and 17.5 cm for completely unrounded vowels. A part of this difference, about 1 cm, was associated with a larynx lowering in rounded vowels, a well known phenomenon, see e.g. Wood (1986).

The conversion from distance d(x) in the sagittal plane to cross-sectional area A(x) was performed on the basis of power function expressions, in part derived from tomographic data for the subject's **[u:] [ɑ:]** and **[i:]**, in part from earlier studies, (Sundberg, 1969; Sundberg et al., 1987; Lin, 1990).

$$A(x) = a\, d(x)^b \qquad (1)$$

For the lip section in the range of $d< 1.7$ cm, we choose $a=1.8$, and $b=2.5$, and for $d>1.7$ cm, $a=5$ and $b=0.6$.

For the mouth cavity we adopted an initial estimate of $a=2.4$ and $b=1.4$, which was derived from Fant (1960) and was found to be very close to the (Sundberg et al., 1987) data. However, when tested against the tomographic data for the subject's **[u:]** and **[ɑ:]** we found that it was necessary to add a correction for air columns on both sides of the tongue. These are quite prominent, see pictures in Fant (1964). For the vowel **[u:]** they add about 35% to the mouth cavity volume. A similar correction was applied to all back vowels.

When applying the power function to pharynx measures it was found to be necessary to divide the range into three regions; one of $d<1.75$ cm with $a=2$ and $b=1.6$, an intermediate region of $1.75<d<2.5$ cm with $a=2.8$ and $b=1$, and a limiting higher region of $d> 2.5$cm with $a=3.7$ and $b=0.7$. Pharynx cross-sectional areas derived from these expressions are somewhat smaller than those reported in Sundberg (1969) but larger than those in Sundberg et al. (1987).

For the larynx tube, we adopted area functions derived from the tomographic data of **[u:] [ɑ:]** and **[i:]**, adjusted to the particular side views .The overall length varied from 2.2 to 2.5 cm and the input area was close 1.7 cm^2. A standard sinus piriformis cavity was inserted.

Calculations were performed from detailed equivalent network representations of area functions (Lin, 1990). All known losses were incorporated. The wall impedance was introduced by an R, L shunt at a level 4 cm above the vocal folds (Fant et al., 1976). This simple representation appears to provide a more realistic overall function than a distributed impedance.

The sound recordings from the X-ray session were not complete in all details, and because of noise interference we had to rely on averages of phonations during the subject's rehearsal prior to exposure.

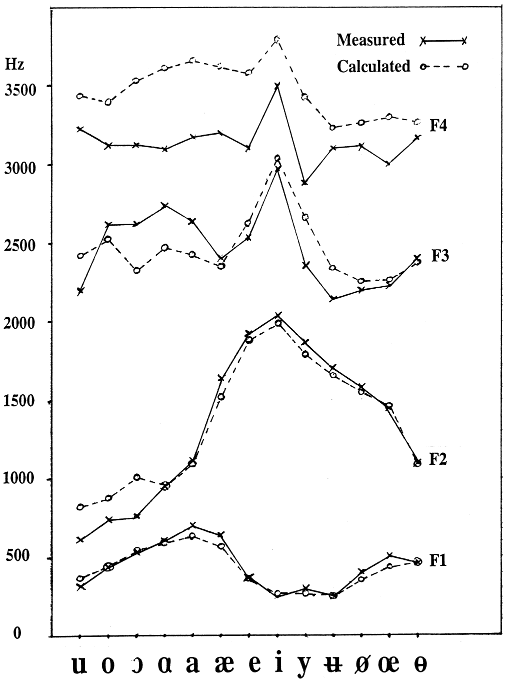

Figure 2. Measured and calculated formant frequencies.

Formant frequencies were measured from broadband spectrograms. Measured and calculated data are shown in Figure 2. Except for the back vowels **[u:]** **[o:]** **[ɔ]** the overall fit in F1 and F2 is quite good with an error of predicted minus measured formant frequencies averaging -22 Hz (SD=27 Hz) in F1 and -42 Hz (SD=42 Hz) in F2.

Except for **[u:]** where F3 is too high, calculated F3 values of back vowels tend to come out too low. F3 of the rounded front vowels **[y:]** and **[ʉ:]** are somewhat too high, reflecting insufficient lip rounding during the X-ray session. Calculated F4 values were about 200 Hz too high, which is a general finding.

As judged from a perturbation analysis, the errors in F2 of the three back vowels, about 200 Hz, reflect a combination of insufficient lip rounding and insufficient narrowing at the tongue constriction of an order of a factor 1.5 to 2.0. These large differences can not likely be explained from systematic errors in the power functions for area determinations. The most likely explanation lies in a relaxed articulation during the exposure. A corresponding, but even greater difference in F2 was observed in the MRI study of Baer et al. (1991).

Systematic analyses of factors related to assumptions concerning vocal tract configurations have been made. Thus, the effect of neglecting the air columns on the sides of the tongue in the vowel **[u:]** is to increase F2 by about 100 Hz and to decrease F3 by about 200 Hz.

The sinus piriformis cavities cause a lowering of all formant frequencies by a small amount, usually less than 70 Hz, whilst the wall impedance has an opposite effect, essentially confined to F1. The central grove in front vowels needs to be taken into account. If neglected, cross-section areas tend to be overestimated and calculated F2 values come out about 80 Hz too low. An increase of the length of the larynx tube by 2.5 mm or introducing a corresponding radiation inductance load at its termination will mainly effect F4 by a lowering of the order of 100 Hz.

A new three-parameter model

Earlier three-parameter models of VT area functions, employing a symmetrical tongue hump constriction between front and back cavities of constant cross-sectional areas are rather stereotype. The constricted region is either given a constant area (Fant, 1960) or a parabolic (Stevens & House, 1955) or a catenoidal (Fant, 1960) or a cosine shape (Lin, 1990). The need for an asymmetry of the constricted region was pointed out by Lin (1990).

Our ambition has been to include as much as possible of physiological realism retaining the three basic parameters, X_c and A_c specifying the location and minimum area of the tongue constriction, and l_o/A_o specifying the length over area ratio of the lip opening. A detailed study of the area functions of the thirteen vowels revealed a number of dependent relations to other descriptors such as jaw opening, ratios of front to back cavity maximum areas, overall length and contriction asymmetry.

As shown in Figure 3, the modelling employs somewhat different conditions for three major regions of the X_c scale, a "front" region of X_c located less than 4 cm from the teeth, a "mid"

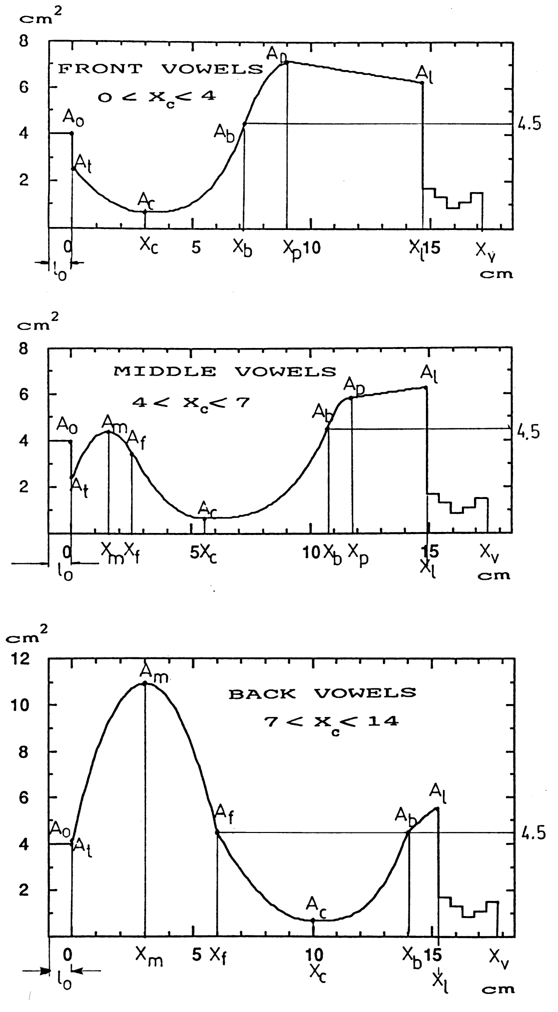

Figure 3. The new three-parameter model.

region at coordinates between X$_c$=4cm and X$_c$=7 cm, and a "back" region at X$_c$ greater than 7 cm. In the latter region, we find all back vowels The vowel **[u:]** occupies a position at the border between the mid and back regions. The vowel **[ɵ]** is close to **[o:]** but has greater A$_x$. As expected, the vowel **[æ:]** fitted best into the back category with the largest X$_c$, i.e. a constriction in the bottom of the pharynx.

The area functions have been divided into a number of segments, six for front vowels, eight for mid vowels and seven for back vowels. A$_o$ extending from X=-1 to X=0 is the lip area. Instead of the default value l$_o$=1 cm one may choose a l$_o$ covarying with A$_o$, retaining the desired l$_o$/A$_o$. The area between the front teeth, A$_t$, is derived from the sagital distance dt through a power function with a=2.4 and b=1.4. A$_m$, at X=X$_m$, is the maximum mouth cavity area in mid and back vowels. In front vowels, there is merely a gradual transition between A$_t$ and A$_c$. The coordinate X$_b$ is defined by A(x)=4.5 cm^2, which is the value approached for the entire area function when A$_c$=4.5 cm^2, i.e the neutral state. A(x)=4.5 cm^2 also defines the coordinate X$_f$, located posterior to X$_m$. In mid vowels the point (X$_f$,A$_f$) on the area function is located in the decent from maximum mouth area A$_m$ to constriction minimum area A$_c$. The (X$_p$/A$_p$) point ensures a suitable shape of the pharynx cavity.

In front vowels, the teeth opening, dt was found to be correlated with increasing A$_c$ and A$_o$. In back vowels, we found a positive correlation of dt with X$_c$, i.e. the jaw opens as the tongue constriction moves towards the bottom of the pharynx. The linear regression equation

$$dt = 1.8X_c\ 0.4 \qquad (2)$$

(r=0.94)

fits well except for **[æ:]** which was produced with 0.5 cm greater dt than predicted.

The coordinates X$_f$ and X$_b$ of back vowels show accurate linear relations to X$_c$, from which we may derive an asymmetry index

$$S_a = (X_c\text{-}X_f)/(X_b\text{-}X_c) \qquad (3)$$

which covers a substantial range, S$_a$=0.37 at X$_c$=7 and 2.4 at X$_c$=13.

The interior length of the vocal tract from the teeth to the entrance to the larynx tube, X$_l$, is kept constant 15.5 cm in back vowels and is varied somewhat with l$_o$/A$_o$ in front vowels and also with X$_c$ in midvowels. The larynx tube is represented by five 0.5 cm sections of standard values. At its outlet in the pharynx it is connected to a sinus piriformis cavity, also of 2.5 cm length, and with an area linearly varied from 3 to 0 cm^2. The individual segments within an area function are in most instances modelled by parabolic functions. An exception is the segment between X$_c$ and X$_b$, where a third order power function was applied, and the intervall between X$_p$ and X$_l$ which was represented by a straight line.

The region of midvowels has on the whole been designed to provide a suitable transition between the very different front and back vowel regions. Some guidance was attained from the area function of a medio-palatal consonant [g], and from the [u:] located close to X$_c$=7. A problem has been to ensure maximally

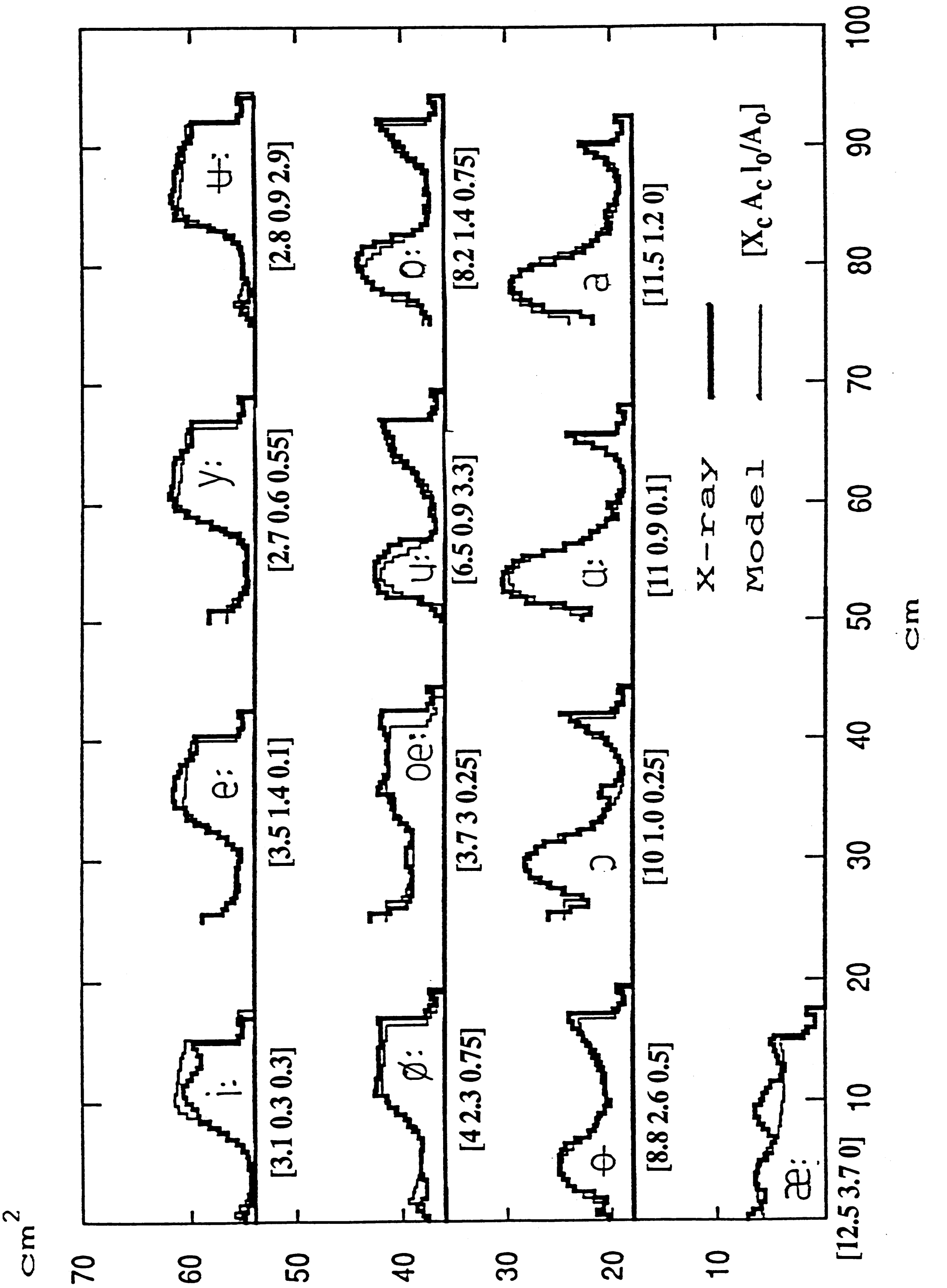

Figure 4. Area functions from the X-ray study and from the model.

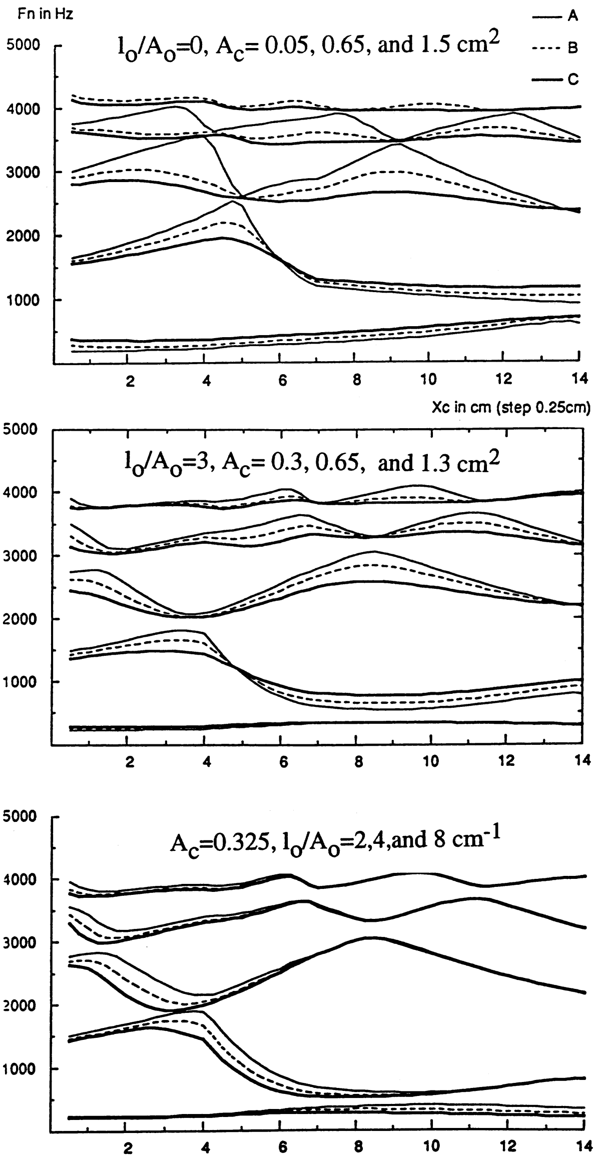

Figure 5. Nomograms of F1, F2, F3, F4, F5 for varying Xc coordinates generated from the new three-parameter model.

continuity at the model boundaries X_C=4 and X_C=7, not only in absolute values but also with respect to derivatives. Some final adjustments remain to be made.

The result of a first order visual matching of model generated area functions to those derived from the X-ray data are shown in Figure 4. The match is on the whole good, but we observe a tendency of a minor underestimation of total length. In **[æ:]** we find an underestimate of the area around X_C=10. The general appearance of the **[æ:]** area function with two internal minima is similar to what has been described by Boe etal. (1992). In terms of X_C, the ordering of the front vowels is **[y:] [ʉ:] [i:] [e:] [œ:] [ɵ]** occupying a region of X_C from 2.7 to 4.0. No vowel was found well within the mid range. An exception was the **[u:]** at X_C= 6.5. The remaining back vowels form the sequence **[o:] [ɵ] [ɔ] [ɑ:] [a]** and **[æ:]** which is in essential agreement with Wood (1979).

Formant frequencies calculated from the model area functions agreed on the whole with those from the X-ray data. In 50% of the cases the model generated formant data matched the subject's phonation better than the X-ray derived data.

Figure 5 shows nomograms of F1, F2, F3, F4, and F5 as function of X_C coordinates. The top figure pertains to l_0/A_0=0, i.e. no lip rounding and A_C=0.05, 0.65 and 1.5 cm^2. Here we note the maximally high F3=3500 Hz at X_C=4 and A_C=0.05 typical of a **[j]** target. The middle diagram pertains to l_0/A_0=3 and A_C=0.3, 0.65 and 1.3 cm^2. The F2-F3 proximity point has now advanced to the left (Fant, 1960; Wood, 1986). At X_C=6.5 the F-pattern is appropriate for an **[u]**. Here, the main effect of a decrease of A_C is to lower F2. This illustrates the origin of the F2 error of **[u:]** discussed in connection with Figure 2. There is apparently a large range of X_C locations that would provide a satisfactory **[u:]**, as already pointed out by Baer et al. (1991). In the lower diagram, A_C was set to a constant value of 0.325 cm^2 and l_0/A_0 = 2, 4, and 8. Here we may note the influence of lip rounding on F2 of **[u:]**. The F2-F3 proximity range is enlarged which make **[u:]** and **[y:]** insensitive to X_C. These are typical examples of stable regions as proposed by Stevens (1989).

We have applied the algorithms of Lin (1990) for inverse transformation, deriving X_C, A_C, and l_0/A_0 from F1, F2 and F3 of the subject's phonation. The automatic search was in most instances successful, but difficulties were encountered in achieving a correct F3 of **[u:] [ɔ]** and **[æ:]** while maintaining correct F1 and F2. The situation will probably improve by a systematic release of the constraints that tie overall length and mouth opening to the three basic parameters. The model has been extended to consonant articulations (Fant & Båvegård, 1997).

Acknowledgements

This work has in part been supported by grants from the Swedish Board for Technical Development, STU and STUF, and from a grant from Carl Tryggers Stiftelse. A considerable amount of calculation work was contributed by Qiguang Lin.

References

Baer T, Gore JC, Gracco LC & Nye PW (1991). Analysis of vocal tract shape and dimensions using magnetic resonance imaging: Vowels. *J Acoust Soc Am* 90(2): 799-828.

Boe L-J, Perrier P & Bailly G (1992). The geometric vocal tract variables controlled for vowel production: proposals for constraining acoustic-to-articulatory inversion. *Journal of Phonetics*, 20: 27-38.

Fant G (1964). Formants and cavities, *Proc of Vth Int Congress of Phonetic Sciences*. Karger, Basel, 120-140.

Fant G (1960). *Acoustic Theory of Speech Production*. The Hague: Mouton.

Fant G, Nord L & Branderud P (1976). A note on the vocal tract wall impedance. *STL-QPSR, KTH,* 4/1976: 13-20.

Fant G (1973). *Speech sounds and features*. The MIT Press. Cambridge, Mass.

Fant G (1983). Feature analysis of Swedish vowels - a revisit. *STL-QPSR, KTH*, 2-3/1983: 1-19.

Fant G & Båvegård M (1997). Parametric model of the vocal tract area function: Vowels and consonants, *ESPRIT/BR SPEECHMAPS (6975). Delivery 28, WP2.2, 1-30 (1995)*. Also published in *TMH-QPSR, KTH*, 1/1997: 1-20.

Lin Q (1990). Speech production theory and articulatory speech synthesis. *D.Sc. thesis*. Royal Inst of Technology, KTH, Stockholm.

Sundberg J (1969). On the problem of obtaining area functions from lateral X-ray pictures of the vocal tract, *STL-QPSR, KTH*, 1/1969: 43-45.

Sundberg J, Johansson JC, Wilbrand H & Ytterberg C (1987). From sagittal distance to area. A study of transverse, vocal tract cross-sectional area, *Phonetica*, 44: 76-90.

Stevens KN & House AS (1955). Development of a quantitative description of vowel articulation. *J Acoust Soc Am*, 27: 484-493.

Stevens KN (1989). On the quantal nature of speech. *Journal of Phonetics*, 17: 9-34.

Wood S (1986). The acoustical significance of tongue, lip, and larynx manoeuvres in rounded palatal vowels. *J Acoust Soc Am* 80/2: 391-401.

Wood S (1979). A radiographic analysis of constriction locations for vowels, *Journal of Phonetics*, 7: 25-44.