Parametric model of VT area functions: vowels and consonants

Fant, G. and Båvegård, M.

journal: TMH-QPSR
volume: 38
number: 1
year: 1997
pages: 001-020

http://www.speech.kth.se/qpsr
Parametric model of VT area functions: vowels and consonants*

Gunnar Fant and Mats Båvegård

Abstract
This is an extension of earlier work on vocal tract area function modelling (Fant, 1960, 1992, 1993; Lin, 1990) retaining a minimum of three independent control parameters in a more complex, physiologically oriented model providing a flexible choice of detail structure. Systematic perturbation analysis of vocal tract boundary conditions and of values of the independent control parameters have been carried out. These data supplement calculated transfer functions and nomograms of formant frequencies as a function of the place, $X_c$, and the area, $A_c$, of a vocal tract constriction and the lip-parameter $l_0/A_0$. Compensatory articulations providing almost the same formant frequency patterns with substantially different choice of control parameters are discussed.

Additional vocal tract area function structures and control parameters for simulating consonantal articulations are introduced. Essential properties of laterals, nasals and retroflex articulations have been simulated.

Inversion experiments on steady state vowels have been successful but there remains to define and test possible constraints for the handling of variations in connected speech.

Introduction

Our modelling is an extension of the earlier work of Fant (1992, 1993). This was based on a set of midsagittal tracings of 13 Swedish vowels from X-ray data including some tomographic studies of the vowels [i], [a] and [u]. The tomographic data (Fant, 1965) revealed the existence of significant air volumes on both sides of the tongue for [a] and [u]. These have been quantified in Fant (1993). Formant frequencies calculated from a 0.5 cm quantization of the VT area function showed a substantial agreement with sound recordings during or immediately before X-ray exposure. The major differences were found in the back vowels [u], [o] and [ɛ] were the calculated $F_2$ came out too high which was found to be related to an overestimation of the constriction area $A_c$ in the VT area function or to a more precise articulation in the sound recording which was not always synchronous with the X-ray exposure.

The modelling of the VT area functions employed the three traditional independent parameters, $X_c$ and $A_c$ for constriction location and area and a lip parameter $l_0/A_0$. The detailed area function was constructed by a concatenation of 6 successive segments and a Sinus piriormis cavity shunting the outlet of the larynx tube. The constants describing the shapes and sizes of these segments were given default values uniquely determined by the three main control parameters. Some of these constants can be released and given the status of secondary parameters for adjustment of the model to a specific speaker.

Examples of normal covariation of control parameters and shape constants exemplified in Fant (1993) are the positive correlation of inter-incisor distance with $A_c$ in front vowels and with $X_c$ in back vowel which reflects increasing jaw opening. In back vowels, the degree of asymmetry of the pharyngeal constriction varies in accordance with the location of $X_c$. As expected, there is also a positive covariation between the overall vocal tract length and the degree of lip-rounding $l_0/A_0$.

In order to conform with natural constraints, the model was divided in three parts with respect to the location of the $X_c$ coordinate which has been defined as the distance from the incisors. Thus we defined a range of $X_c<$4 cm where all the front vowels were found at $4 < X_c > 2.5$. The second region, $4<X_c<7$, was referred to as midvowels where we located [u] at about $X_c=6.5$. The third region of $7<X_c<14$ housed the back vowels in the order of [o], [a] [æ]. These preferred locations are by now quite well documented in several studies, e.g. Wood (1979), Boe et al. (1992).

* This is a revised edition of an article with the same title submitted as a SPEECHMAPS (Esprit/Bre 6975) Delivery 28 WP2.2 (1995) and included in the Tekn.lic. thesis of Mats Båvegård (1996).
Front vowels ($X_c < 4$ cm)

\[ X_b = X_c + 4.1(A_c + 0.1)^{-0.1} \]
\[ X_p = X_b + 2 \]
\[ X_l = 14.4 + 1.4[X - \exp(-l_0/2A_0)] \]
\[ D_t = [0.5 + 0.75A_c^{1/2}][1 - 0.5 [1 - \exp(-2.1 l_0/A_0)]] \]
\[ A_t = 2.4 D_t^{1.4} \]
\[ A_b = 4.5 \]
\[ A_p = 7.6 - 0.69A_c; \quad A_c < 4.5 \text{ cm}^2 \]
\[ = 1.43A_c - 1.93; \quad A_c \geq 4.5 \text{ cm}^2 \]
\[ A_l = 6.55 - 0.115A_c^{2.07} \]

Segments

Weight function in interval $[1 < X_c < 12]$:

\[ W(X_c) = 0.5(1 - \cos(2\pi(X_c - 1)/11)) \]
\[ p(X_c) = 3(1 - 0.5W(X_c)) \]

1) $X < X_f$

\[ A_{tc} = A_c + (A_t - A_c)[(X_c - X)/X_c]^2 \]

2) $X_f < X < X_b$

\[ A_{cb} = A_c + (4.5 - A_c)[(X_c - X)/(X_c - X_b)]^2 \]

3) $X_b < X < X_p$

\[ A_{bp} = A_p - (A_p - 4.5)[(X_p - X)/(X_p - X_b)]^2 \]

4) $X_p < X < X_l$

\[ A_{pl} = A_p + (A_l - A_p)[(X_c - X)/(X_l - X_p)] \]

5) $X > X_l$

Standard larynx inserted (scaled or nonscaled).

Mid-vowels ($4 < X_c < 7$ cm)

Weight function in interval $[5 < X_c < 8]$:

\[ W(X_c) = 0.5(1 - \cos(2\pi(X_c - 5)/3)) \]
\[ X_m = X_c - 4 \]
\[ X_f = 5(X_c - 4)/3 \]
\[ X_b = X_c + 4.1(A_c + 0.1)^{-0.1}(7 - X_c)/3 \]
\[ + (8.9 - 0.49X_c)(X_c - 4)/3 \]
\[ X_p = X_p + 2(7 - X_c)/3 \]
\[ X_l = 14.4 + 1.4[X - \exp(-l_0/2A_0)] \]
\[ + 0.5\exp(-A_c/3) W(X_c) \]

Back vowels ($7 < X_c < 14$ cm)

Weight function in interval $[5 < X_c < 8]$:

\[ W(X_c) = 0.5(1 - \cos(2\pi(X_c - 5)/3)) \]
\[ X_m = 3.0 \]
\[
X_f = 0.1397(7 - X_c)^2 + 0.1583(7 - X_c) + 5
\]
\[
X_b = 8.9 + 0.51X_c
\]
\[
X_l = 14.4 + 1.4[1 - \exp(-l_0/2A_0)] + 0.5\exp(-A_c/3)W(X_c) - (X_c - 7)/10
\]
\[
D_t = 0.153X_c - 0.04[1 - \exp(-l_0/2A_0)] + 0.153X_c - 0.04 \quad X_l = 14.5 \text{ cm}
\]
\[
A_c = 2.4D_t^{1.4}
\]
\[
A_m = 4.5 + (1.13X_c - 5.7)(4.5 - A_c)/3.35
\]
\[
A_f = 4.5
\]
\[
A_b = 4.5
\]
\[
A_l = [6.55 - 0.115A_c^2.07][1 - [(X_c - 7.0)/5]^2] + 4.5[(X_c - 7.0)/5]^2
\]

**Segments**

Weight function in interval \([1 < X_c < 12]\):

\[
W(X_c) = 0.5(1 - \cos(2\pi(X_c - 1)/11))
\]

\[
p(X_c) = 3(1 - 0.5W(X_c))
\]

1) \(X < X_m\)

\[
A_m = A_m - (A_m - A_1)(X_m - X_m/X_m)^2
\]

2) \(X_m < X < X_f\)

\[
A_{mf} = A_m - (A_m - 4.5)(X_m - X_m)/(X_f - X_m)^2
\]

3) \(X_f < X < X_c\)

\[
A_{fc} = A_c + (4.5 - A_c)(X_c - X_c)/(X_c - X_f)^2
\]

4) \(X_c < X < X_b\)

\[
A_{cb} = A_c + (4.5 - A_c)(X_c - X_b)/(X_c - X_c)^2
\]

5) \(X_b < X < X_l\)

\[
A_l = 4.5 + (A_l - 4.5)(X_b - X_b)/(X_l - X_b)
\]

6) \(X > X_l\)

Standard larynx inserted (scaled or nonscaled).

Valid equations:

- \(X_f > X_l\): Equation 1, 2 and 6
- \(X_c > X_l\): Equation 1, 2, 3 and 6
- \(X_b > X_l\): Equation 1, 2, 3, 4 and 6

**The revised model**

**Objectives**

The basic outline of the revised model, Fig. 1a, b and c, with its three prototypes for front, mid and back vowels, is the same as in the earlier version but the defining equations and constants have been elaborated in several ways. One object has been to allow for a more realistic modelling of the vowel [u], i.e. to secure a sufficiently low \(F_3\). Another is to produce more realistic prototypes of [ae] modelled alternatively as a front or a back vowel. We are also using a larger default value of the inter incisor distance \(D_t\), which determines the cross-sectional area of the teeth passage and a new default equation for the reduction of \(D_t\) with increasing lip rounding. These model aspects will be discussed in greater detail in the following sections.

**Peripheral constraints**

Our standard radiation impedance is that of Stevens, Kasowski & Fant, also quoted in Badin & Fant (1984) and Lin (1990).

We have means of implementing two types of wall impedance:

1) A distributed impedance of

\[
Z_w(x) = R_w(x) + j\omega L_w(x)  \quad (1)
\]

with \(R_w = 660\) and \(L_w = 1.4\) acoustical units per cm² surface area.

2) An area independent, lumped impedances of type

\[
Z_w = R_w + j(\omega L_w - 1/\omega C_w)  \quad (2)
\]

inserted at two places, 2 cm above the outlet of the larynx tube and 2 cm behind the incisors. The magnitudes are tabulated in table 1.

<table>
<thead>
<tr>
<th>(R_w)</th>
<th>(L_w)</th>
<th>(C_w)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
These data were used by Lin (1990) and derive from Fant et al. (1976). Assuming a vocal tract of volume 75 cm$^3$ and the two inductance in parallel the closed tract resonance becomes $F_w=210$ Hz. In practice we use a scale factor of 35% larger neck impedance which reduces $F_w$ to 189 Hz. For an [i]-type vocal tract of volume 50 cm$^3$ and with the 35% scale factor for the neck impedance, we obtain the same value of $F_w=189$ Hz. These values derived from simple Helmholtz resonator conditions agree well with experimental values from the complete vocal tract simulation, which gave 191 Hz for the [i] and 193 Hz for the neutral tube.

According to Fant et al. (1973), a group of female subjects showed 14% higher $F_w$ values than for men. This is related to 30% larger wall inductance and as much as 65% smaller VT volumes (female volume = 0.6 male volume). Also the main difference in VT volume as well as in wall inductance was confined to the pharynx. These data confirm with general female/male shape factors and could have some use in deriving male-female VT transforms.

The close tract bandwidth $B_w=R_w/2\pi L_w$ is close to 75 Hz for males and 24% higher for females. The $L_wC_w$ wall resonance is 40 Hz. The $C_w$ has no effect on the vocal tract resonance modes but is included for future extensions to dynamic effects in stop production.

It is important to keep track of the magnitude of $F_w$ since this is the limiting low frequency value of $F_1$ and has a trading relationship with $A_c$. This can be a problem in inversion.

The model may optionally combine the output from the lips with that radiated from the walls which has no effect on formant frequencies, but may cause some interference in the transfer function of very closed articulatory configurations, (Lin, 1990, p. 29). It has a definite role in the simulation of sounds produced with complete vocal tract closure, but may be faked by a minimum $A_c$ value.
At other area discontinuities, e.g. at the teeth and at the velar constriction of [u] or in any consonantal constriction, we have omitted the end correction inductance, quantified by Fant (1960) and experimentally tested by Sundberg et al. (1992), since it can be traded for an \( A_c \) decrease and we anyhow have difficulties in accurately estimating constriction areas and estimating the degree of abruptness of area changes.

At the front end, the tract is terminated by a tube of 1 cm length anterior to the teeth, i.e. at coordinates of \(-1 < X_c < 0\). This standard lip section has an area which is the inverse of the lip parameter \( l_0/A_0 \). Values of \( l_0/A_0 \) smaller than 0.1 are excluded. Calculations can be performed with varying unit lengths in the equivalent network. The standard is 0.5 cm but the sectioning can be done as finely as with 1 mm units. The basic program was developed by Lin (1990).

**Nomograms**

A representative set of nomograms are documented in Figs 4 and 5. Four values of \( A_c \): 4.5, 1.5, 0.65 and 0.2 cm\(^2\) are included. In Fig. 4 \( l_0/A_0 = 0 \). The bottom part shows the effects of removing the Sinus piriformis cavity. There is a shift in \( F_5 \) less close to \( F_4 \) and in the posterior region, the place of minimal \( F_2 \) \( F_1 \) distance is advanced by about 1 cm. The effects of Sinus piriformis will be discussed in more detail in the section on perturbation analysis. (With the SP retained there is an instability of \( F_5 \) in the extreme, but physiologically irrelevant, combination of \( 13 < X_c < 14 \) and very small \( A_c \) values). The three nomograms in Fig. 5 pertain to \( l_0/A_0 = 0.5, 2 \) and 8. With increasing lip rounding, the \( F_2 F_3 \) proximity region is advanced. The main features of Figs 4 and 5 conform with those of earlier published nomograms, e.g. of Fant (1960).

![Figure 4. Nomograms of \( F_1, F_2, F_3, F_4, F_5 \) at varying \( X_c \) and \( l_0/A_0 = 0 \). \( A_c = 0.2 \) cm\(^2\) (solid), 0.65 cm\(^2\) (dashed), 1.5 cm\(^2\) (dotted), 4.5 cm\(^2\) (semi-dashed). Top: Sinus piriformis included, Bottom: without Sinus piriformis.](image-url)

Fig. 6 is a close up of the front part of Fig. 4 with an \( A_c \) of 0.05 cm\(^2\) included. At \( X_c \) around 4 this would correspond to the target of a consonant \([j]\). The bottom part of the figure shows the VT area and transfer functions (with +6 dB/oct for radiation transfer added) for \( A_c = 0.05 \) and \( A_c = 0.65 \) at \( X_c = 4 \), the latter representing a non-extreme \([i]\). Observe the \( F_4 \) \( F_5 \) proximity and the clustering of \( F_3, F_4 \) and \( F_5 \) in the prominent high frequency region, In \([j]\), the distance between \( F_3 \) and \( F_5 \) is merely 650 Hz. This clustering, which basically derives from the horn shaped outline of the front half of the vocal tract, is representative of real speech and should be noted in formant synthesis and inversion.
Fig. 5. Nomograms of $F_1$, $F_2$, $F_3$, $F_4$, $F_5$ at varying $X_c$. $A_c=0.2$ cm$^2$ (solid), 0.65 cm$^2$ (dashed), 1.5 cm$^2$ dotted), 4.5 cm$^2$ (semi-dashed). From top: $l_0/A_0 = 0.5$, $l_0/A_0 = 2$ and $l_0/A_0 = 8$.

Fig. 7 illustrates two nomogram regions where we may locate the vowel [ae], which can be considered either as a front vowel with high $A_c$ or as an extreme back vowel. Typical coordinates could be $X_c=4$ with $A_c=6$ or $X_c=13.5$ with $A_c=2$.

In spite of the differences in the areafunctions, there is a functional similarity which can provide the same $F_1$, $F_2$ and $F_3$. With reference to the model properties of back vowels (Fig. 1), it should be observed that the posterior part of the model becomes truncated when $X_c$ approaches $X_l$, i.e. the coordinate of the entrance to the larynx tube. The area functions and corresponding transfer functions of a front and a back [ae] show an overall similarity.

Finite rounding conditions are illustrated in the nomograms of Fig. 8. In the top diagram, $l_0/A_0 = 3$ and three different values of $A_c$ are included, 0.3, 0.65 and 1.3 cm$^2$.

In the lower diagram $A_c$ is set to the constant value of 0.325 cm$^2$ and $l_0/A_0=2$, 4 and 8. In the upper diagram at $X_c$ around 3 and $A_c$ of the
Fig. 6 Top: Anterior part of the nomogram in Fig. 4 expanded to include $A_c = 0.05 \text{ cm}^2$ (solid line). Bottom: Area functions and transfer functions of [i], $A_c = 0.65 \text{ cm}^2$, $F_1 = 329$, $F_2 = 2090$, $F_3 = 2978 \text{ Hz}$ and [j], $A_c = 0.05 \text{ cm}^2$ with $F_1 = 214$, $F_2 = 2282$, $F_3 = 3457 \text{ Hz}$, both with $X_c = 4 \text{ cm}$.

order of 1 cm$^2$ we could locate the Swedish vowel [u] which is a true front vowel. The normal location for the Swedish [u:] with a low $F_3$ and very low $F_2$ is at $X_c$ around 6.5. A variant with higher $F_3$ and occasionally higher $F_1$ is located in the back vowel region at $X_c$ around 10 cm which earlier has been noted by Boe et al. (1992), Savariaux et al. (1995). Potentially, [u] vowels can be produced with $X_c$ anywhere between 6 and 10.

Area functions and transfer functions of two [u]-prototypes with similar formant patterns are illustrated in the bottom part of Fig. 8.

Fig. 7. Nomograms adopted for the $X_c/A_c$ localisation of [æ]. In the front vowel region $A_c = 6, 4$ and 2 cm$^2$. In the back vowel region $A_c = 2, 3, 4 \text{ cm}^2$. Typical front vowel version is $X_c = 4, A_c = 6$ with $F_1 = 648$, $F_2 = 1595$ and $F_3 = 2450 \text{ Hz}$. Typical back vowel version is $X_c = 13.5, A_c = 2$ with $F_1 = 654$, $F_2 = 1588$ and $F_3 = 2452 \text{ Hz}$.

A limited number of changes in model constants functioning as secondary parameters are illustrated in Fig. 9. One problem we have had is to secure a sufficiently low $F_3$ of [u] in the region of 2400 Hz or lower. A simple recipe is to increase the length of the pharynx but an anatomically more motivated alternative is to reduce the effective length of the tongue constriction. In our recent modelling, this has been achieved by a reduction of the power function for the posterior contour of the constriction which has been changed from $p_s = 3$ to $p_s = 1.5$. 

$\text{Fig. 7. Nomograms adopted for the } X_c/A_c \text{ localisation of [æ]. In the front vowel region } A_c = 6, 4 \text{ and } 2 \text{ cm}^2. \text{ In the back vowel region } A_c = 2, 3, 4 \text{ cm}^2. \text{ Typical front vowel version is } X_c = 4, A_c = 6 \text{ with } F_1 = 648, F_2 = 1595 \text{ and } F_3 = 2450 \text{ Hz}. \text{ Typical back vowel version is } X_c = 13.5, A_c = 2 \text{ with } F_1 = 654, F_2 = 1588 \text{ and } F_3 = 2452 \text{ Hz.}$
One of the typical features of a female vocal tract is the relatively shorter pharynx (Fant, 1975). As illustrated in the bottom part of Fig. 9, the consequence for the vowel [i] is a substantially higher F₂ while F₃ is less affected. This is to be expected from the well known back cavity affiliation of F₂ of [i].

Another secondary parameter of some importance is the inter-incisor distance D₁ which in our model determines the cross-sectional area, $A_t = 2.4D_1^{1.4}$, at and immediately behind the teeth. Our recent model has larger $D_1$ values than the earlier version which is an improvement. The $D_1$ measure is, as earlier stated, mainly influenced by the jaw opening but has been observed to also follow the degree of liprounding. The amount of decrease of $D_1$ with increasing $l_0/A_0$ in our formula has allowed reasonable inversions. The effect of a 40% reduction of $D_1$ in the region of $X_c < 7$ and $l_0/A_0 = 0$ is very small. With liprounding included, the most notable effects of the decreased $D_1$ is an increase of about 100 Hz in $F_3$ of rounded front vowels such as [u] and [ø].

**Perturbations**

Results from a systematic variation of boundary conditions are illustrated in Table 2 which shows differences between standard conditions and the system after a removal or a change in one of the boundary elements, the Sinus piriformis, the VT wall load or the length of the larynx tube. Our standard wall load is a lumped impedance with the 1.35 scale factor of the posterior load. The alternative of a distributed wall impedance is included in the tabulation. The larynx tube perturbation is here defined as
Fig. 8. Top: Nomograms with $l_0/A_0 = 3$ and $A_c = 0.3\text{ cm}^2$ (solid line), $0.65\text{ cm}^2$ (dashed) and $1.3\text{ cm}^2$ (dotted). Middle: Nomograms with $A_c = 0.325\text{ cm}^2$ and $l_0/A_0 = 8$ (solid line), 4 (dashed) and 2 (dotted). The area functions and transfer functions below represent a midvowel [u] at $X_c = 6.5\text{ cm}$, $A_c = 0.325$ and $l_0/A_0 = 8$ with $F_1 = 272$, $F_2 = 632$, $F_3 = 2242\text{ Hz}$ and a backvowel [u] at $X_c = 10.5\text{ cm}$, $A_c = 0.325$ and $l_0/A_0 = 4$ with $F_1 = 304$, $F_2 = 587$, $F_3 = 2561\text{ Hz}$.
Fig. 9. Top: [u] with a variation of the shape of the part posterior to $X_c$. Middle: [u] with varying pharynx length. Bottom: [i] with varying pharynx.
Table 2. Systematic variation of boundary conditions of four different vowels.

<table>
<thead>
<tr>
<th></th>
<th>ΔF1</th>
<th>ΔF2</th>
<th>ΔF3</th>
<th>ΔF4</th>
<th>ΔF5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sinus piriformis</td>
<td>i</td>
<td>-9</td>
<td>-109</td>
<td>-13</td>
<td>-4</td>
</tr>
<tr>
<td></td>
<td>æ</td>
<td>-18</td>
<td>-99</td>
<td>-111</td>
<td>-5</td>
</tr>
<tr>
<td></td>
<td>a</td>
<td>-43</td>
<td>-122</td>
<td>-9</td>
<td>-45</td>
</tr>
<tr>
<td></td>
<td>u</td>
<td>-10</td>
<td>-12</td>
<td>-118</td>
<td>0</td>
</tr>
<tr>
<td>Lumped walls</td>
<td>i</td>
<td>76</td>
<td>3</td>
<td>60</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>æ</td>
<td>21</td>
<td>8</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>a</td>
<td>50</td>
<td>20</td>
<td>5</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>u</td>
<td>87</td>
<td>31</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Distributed walls</td>
<td>i</td>
<td>55</td>
<td>8</td>
<td>12</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>æ</td>
<td>15</td>
<td>6</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>a</td>
<td>23</td>
<td>13</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>u</td>
<td>57</td>
<td>21</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>Larynx length</td>
<td>i</td>
<td>-2</td>
<td>39</td>
<td>-27</td>
<td>-225</td>
</tr>
<tr>
<td></td>
<td>æ</td>
<td>-3</td>
<td>-25</td>
<td>-71</td>
<td>-204</td>
</tr>
<tr>
<td></td>
<td>a</td>
<td>-8</td>
<td>-17</td>
<td>-5</td>
<td>-190</td>
</tr>
<tr>
<td></td>
<td>u</td>
<td>-2</td>
<td>-2</td>
<td>-45</td>
<td>-96</td>
</tr>
</tbody>
</table>

The difference between normal full length and a decrease by 0.5 cm.

The vowels have been selected from the set of long Swedish vowels of Fant et al. (1969), Fant (1975) submitted to inversion (Båvegård & Fant, 1995).

Some main trends can be noted. As expected, the wall impedance load increases all formant frequencies and the effect is greater for the lumped than the distributed version. In the latter case, formant frequencies can approximately be estimated from the formula

$$F_n = (F_{ni}^2 + F_w^2)^{0.5}$$

where $F_{ni}$ represents hard wall conditions (Fant, 1960, 1972; Badin & Fant, 1984; Lin, 1990).

The influence of the wall load is mainly confined to $F_1$ except for the lumped version where the posterior load also affects $F_2$ of back vowels and the anterior load affects $F_3$ of [i].

The presence of Sinus piriformis lowers all formant frequencies to various degrees. As with the lumped wall impedance, the effects can be predicted from formant-cavity affiliations. The Sinus piriformis lowers $F_1$ of [a] by about the same amount, of the order of 50 Hz, as it is raised by the lumped wall load. Other notable effects of the Sinus piriformis are the lowering of $F_2$ of [a] and of [i] and of $F_3$ of [u]. The greatest effect is on $F_5$ of [æ], [i] and [æ] which are lowered by more than 200 Hz.

The most prominent effect of a shortening of the larynx tube is an increase in $F_5$ and to a less extent $F_4$. Conversely, a lengthening of the larynx or the introduction of an end-correction will cause a decrease of $F_4$ and $F_5$.

An inclusion of the air on both sides of the tongue in [u] caused a 100 Hz lowering of $F_2$ (Fant, 1993).

Results from a perturbation of $A_c$, $X_c$ and $l_q/A_0$ within the total ensemble of Fant et al. (1969) vowels appear in Table 3. It shows the differential change in $F_1$ $F_2$ and $F_3$ for a variation of $A_c \pm 20 \%$ and $X_c \pm 0.25 \text{ cm}$. These are the mean of plus and minus perturbations. The lip parameter is perturbed by adding 0.5 to the $l_q/A_0$ value. The main impression is that of stability with moderate perturbation effects. An exception is the high sensitivity of $F_3$ of [i] and [y] for a change in $X_c$. This is related to the location of their $X_c$ at 4.0 and 4.2 respectively, just anterior to the coordinate where $F_3$ meets $F_2$.

Here, the steepness of the $\delta F_3/\delta X_c$ is enhanced by the small $A_c=0.2 \text{ cm}^2$ which in turn is required by the low $F_1$. An increase of $F_1$ by 20 Hz only, would require an increase of $A_c$ by about 50%. An anterior shift of $X_c$ of [i] from 4.0 to 3.5 and a somewhat higher $A_c$, which seems to be more representative, would have greatly decreased the perturbation sensitivity without a major change in the formant pattern. The effect of lip rounding is rather low in [i]. A removal of the $l_q/A_0=0.36$ value would have increased $F_1$ by 7 Hz, $F_2$ by 50 Hz and $F_3$ by 110 Hz. A labial constriction has a greater effect in more open vowels where it overrides the $A_c$ constriction. The neutral vowel [a] has been trapped in the mid-vowel range at $X_c=6.4$, $A_c=1.6$ and $l_q/A_0=0.1$. However, a doubling of $A_c$ has a rather small effect, +45 Hz in $F_1$, +25 Hz in $F_2$ and -24 Hz in $F_3$. These examples
Table 3. Parameter perturbations of Swedish vowels, see text.

<table>
<thead>
<tr>
<th>Ac</th>
<th>Xc</th>
<th>lo/Ao</th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
<th>ΔΔAc</th>
<th>ΔΔXc</th>
<th>ΔΔlo/Ao</th>
<th>ΔΔF1</th>
<th>ΔΔF2</th>
<th>ΔΔF3</th>
</tr>
</thead>
<tbody>
<tr>
<td>u</td>
<td>0.28</td>
<td>6.96</td>
<td>7.33</td>
<td>282</td>
<td>588</td>
<td>2353</td>
<td>0.8</td>
<td>1.7</td>
<td>-3.3</td>
<td>16.8</td>
<td>-0.3</td>
</tr>
<tr>
<td>o</td>
<td>0.30</td>
<td>11.10</td>
<td>1.50</td>
<td>400</td>
<td>672</td>
<td>2416</td>
<td>0.4</td>
<td>-7.3</td>
<td>-35.5</td>
<td>23.3</td>
<td>-3.2</td>
</tr>
<tr>
<td>a</td>
<td>0.78</td>
<td>10.83</td>
<td>0.29</td>
<td>606</td>
<td>926</td>
<td>2563</td>
<td>2.4</td>
<td>0.2</td>
<td>-112.7</td>
<td>33.6</td>
<td>-10.8</td>
</tr>
<tr>
<td>æ</td>
<td>2.00</td>
<td>13.50</td>
<td>0.00</td>
<td>654</td>
<td>1589</td>
<td>2459</td>
<td>-1.0</td>
<td>-10.7</td>
<td>-179.7</td>
<td>7.6</td>
<td>44.5</td>
</tr>
<tr>
<td>æ</td>
<td>6.00</td>
<td>4.00</td>
<td>0.00</td>
<td>621</td>
<td>1629</td>
<td>2441</td>
<td>7.9</td>
<td>-1.4</td>
<td>-125.9</td>
<td>20.3</td>
<td>44.5</td>
</tr>
<tr>
<td>e</td>
<td>2.40</td>
<td>4.50</td>
<td>0.00</td>
<td>482</td>
<td>1824</td>
<td>2657</td>
<td>24.3</td>
<td>6.3</td>
<td>-58.4</td>
<td>-51.8</td>
<td>-60.9</td>
</tr>
<tr>
<td>e</td>
<td>0.60</td>
<td>4.50</td>
<td>0.00</td>
<td>348</td>
<td>2181</td>
<td>2847</td>
<td>12.1</td>
<td>11.4</td>
<td>-24.1</td>
<td>-38.0</td>
<td>-40.1</td>
</tr>
<tr>
<td>i</td>
<td>0.21</td>
<td>4.02</td>
<td>0.36</td>
<td>256</td>
<td>2158</td>
<td>3094</td>
<td>6.8</td>
<td>6.8</td>
<td>-9.9</td>
<td>-9.5</td>
<td>-55.8</td>
</tr>
<tr>
<td>y</td>
<td>0.28</td>
<td>4.24</td>
<td>0.88</td>
<td>266</td>
<td>2085</td>
<td>2644</td>
<td>6.7</td>
<td>10.8</td>
<td>-7.0</td>
<td>-29.8</td>
<td>-46.2</td>
</tr>
<tr>
<td>u</td>
<td>1.02</td>
<td>3.25</td>
<td>2.53</td>
<td>287</td>
<td>1653</td>
<td>2268</td>
<td>7.4</td>
<td>0.9</td>
<td>-6.4</td>
<td>-44.5</td>
<td>-37.0</td>
</tr>
<tr>
<td>ø</td>
<td>1.11</td>
<td>4.84</td>
<td>0.54</td>
<td>375</td>
<td>1756</td>
<td>2377</td>
<td>11.1</td>
<td>10.9</td>
<td>-21.7</td>
<td>-43.8</td>
<td>-2.3</td>
</tr>
<tr>
<td>ø</td>
<td>1.60</td>
<td>6.40</td>
<td>0.10</td>
<td>501</td>
<td>1510</td>
<td>2501</td>
<td>11.4</td>
<td>10.4</td>
<td>-53.8</td>
<td>5.9</td>
<td>-6.4</td>
</tr>
</tbody>
</table>
illustrates some of the compensatory phenomena which complicate the inversion.

**Inversion**

The inversion tool provides a rapid access to three-parameter data of any set of tabulated formant frequencies.

![Graph showing formant frequencies](image)

**Fig. 10. Inversion of Swedish vowels (Fant et al. 1969).** Top: Measured formant frequencies (solid line). Codebook lookup frequencies (dashed) and after the optimisation procedure (dotted). Bottom: Vocal tract parametric representation with codebook data (dashed) and after the optimisation procedure (dotted).

Figure 10 shows data derived from Fant et al. (1969) which were adopted for the perturbation studies above and Figure 11 the corresponding data of inverting formant frequencies from the X-ray recording in 1956. Apart from some individual dialectal differences there is an overall similarity of formant frequencies and VT-parameters, e.g. the large number of articulations in the region of $X_c=4$ and $X_c=10$. A striking difference is that [æ] became identified as a back vowel in Fig. 10 and as a front vowel in Fig. 11. The X-ray data contained some short vowel prototypes, [ɔ] [u] [u] and the pre-[r] variant of [ø] labeled [œ]. The [ø] is the short mate of the long [u]. The [œ] and [u] were not included in the Fant et al. (1969) data.

![Graph showing formant frequencies](image)

**Fig. 11. Inversion of Swedish vowels produced by the X-ray subject in 1956 (Fant, 1992).** Top: Measured formant frequencies (solid line). Codebook lookup frequencies (dashed) and after the optimisation procedure (dotted). Bottom: The VT-parametric data is that obtained after optimisation.

From the limited experience we now have on inversion it may be concluded that although we generally provide plausible solutions there are many trading relations that could cause a non-uniqueness. However, these are usually confined to articulatory patterns that are not too distant, e.g. in the domain of front vowels. Boundary conditions, such as the choice of wall impedance, the size of the Sinus piriformis, the overall VT length and the terminating areas at the teeth and at the entrance of the larynx tube and maximum cross-sectional
areas within the tract, will affect relative values of the control parameters. However, relational aspects within an ensemble or within a temporal sequence will be fairly well preserved (Båvegård & Fant, 1995).

**Consonants**

The basic principle is to decompose any VT area function into two parts, an overall vocalic part with parameters settings appropriate for the particular coarticulation and a parametrically specified consonantal part which substitutes or modifies parts of the vocalic area function. Temporal organisation will involve the control of covarying and in part dependent vocalic and consonantal area function parameters subject to articulatory constraints and representative time constants. The modelling is still in an initial stage of static representation of a main inventory of consonant categories.

**Consonantal modifications**

The consonantal part of a VT area function is specified by four parameters. Two of these, analogous to $X_C$ and $A_C$ of the vocalic model, have the primary function of specifying the location and degree of consonantal constrictions. The two additional parameters pertaining to the effective length and shape of the consonantal constriction are dependent on the primary consonantal and also on the primary vocalic parameters. As shown in Fig. 12, the outline within a consonantal region and its interface to the vocalic part is shaped by straight lines or sinusoids. The consonantal parameters are defined as follows:

$X_{cc}$ is the coordinate of the constriction, defined by the location of its minimum cross-sectional area or by the center point of a region of complete closure. In velar and palatal constrictives, $X_{cc}$ may coincide with $X_C$ but is still an independent control parameter.

$A_{cc}$ is the cross-sectional area at $X_{cc}$.

$W$ is the width of the local modification projected on the X-coordinate of the vocalic area function. It can be decomposed in two parts, the total region of consonantal modification, $W_1$, and the region of complete closure, $W_2$. The notation $W$ without subscript implies $W_2=0$. In practice, this state is modelled with $A_{cc} = 0$ in a tube of 0.5 cm length.

$K_s$ is the fraction of $W$ located posteriorly with respect to $X_{cc}$.

**Fig. 12. Model of consonantal constriction with three different shapes.** The asymmetrical contour has a tilt factor $k_s = 0.3$. The constriction area $A_{cc}$ is in cm$^2$ and the location $X_{cc}$ is in cm from the teeth. $W$ is the window length in cm.

Labial closure and release are handled by the $I_0/A_0$ parameter of the vocalic system while the rest of the area function may employ a combination of vocalic and consonantal structure.

Modelling secondary regions of vocal tract narrowing in consonants, e.g. a post-velar narrowing as in [r] or a somewhat retracted pharyngeal tongue root in alveolars and retroflex sounds, is still in an initial stage. In these instances, a secondary consonantal region may have to be postulated.

Nasal consonants and nasalised vowels require an additional control parameter which is the velopharyngeal opening subject to constraints with respect to effective length and influence on the inlet to the mouth cavity.

In palatals and velars, the vocalic and consonantal constrictions coincide more or less but are still handled separately. Labial consonant closure and release are handled by the $I_0/A_0$ parameter of the vocalic system which leaves the tongue and thus the rest of the VT area function free to attain a combined vocalic and consonantal pattern, e.g. with a region of consonantal narrowing separate from that of the vocalic constriction as in apical, retroflex articulations.

**Experimental work**

We have tested these principles in the modelling of apical retroflex articulation, laterals, and nasals with labial and alveolar articulation. Front vowel coarticulation was assumed.

**Retroflex articulation**

The apical retroflex articulation illustrated in Fig. 13 is modelling with a vocalic part of $X_C=6.5$ and $A_C=1.3$ combined with a consonantal part, $X_{cc}=3.0$ cm, $A_{cc}=0.1$, $W=4$ cm, $K_s=1$. 
The consonant part is modelled with straight line contours smoothly interfacing to the posterior part and abruptly connecting to the anterior cavity. The resulting transfer function, compared to that of the underlying vowel, displays the typical effect of a rise in F₂ and a lowering of F₃, F₄ and F₅ which brings F₂ and F₃ together close to 2000 Hz.

![Diagram](image)

**Fig. 13.** An apical, retroflex articulation in a front vowel context causes a lowering of F₃ and higher formants. Top: Vocalic area function with consonantal modification. Bottom: Transfer function, with formant frequencies and bandwidths for the consonant.

### Laterals

The modelling of the [l] in Fig. 14 is more complex. It follows the general conditions suggested in Fant (1960).

Two alternatives are provided. One starts from a front vowel configuration with Xₗ=3.0 and Aₗ=2.0. The consonantal parameters are Xₗₗ=2.0, Aₗₗ=0, W=2 cm and Kₛ=1. The lateral passage along the cheeks starts at a coordinate of X= 6.5 cm posterior to the teeth.

The length of the cavity medially behind the apical closure point, acting as a shunt to the lateral passage, is thus 6.5-2 = 4.5 cm. The corresponding zero is located at 3028 Hz which in view of the particular cavity shape is larger than the c/4l = 1960 Hz of a homogenous tube. The additional pole at 2880 is located just above F₃. Changing the Aₗ of the vocalic part from 2.0 to 1.0 cm² does not have any dramatic effects. F₂ shifts from 1675 to 1800 Hz, and F₃, F₄ and the zero are increased by about 60 Hz.

The lower part of Fig. 14 shows the effect of selecting a more retracted vocalic articulation, Xₗ=6, retaining Aₗ=2 and maintaining the consonantal coordinates and thus the length of the lateral pathway and of the medial shunting system behind the constriction. The latter cavity has now attained a quite different shape with a fronted location of the main volume which accounts for the lower zero at 2330 Hz. F₃ and F₄ have merged to a frequency around 2675 Hz.

These spectral patterns of [l] sounds are in general agreement with experience from speech analysis and with the sweep tone studies of Båvegård et al. (1993).

### Nasals

The labial and the alveolar nasal consonants illustrated in Figure 15 are modelling with a standard nasal tract (Fant, 1985; Lin, 1990) (See also Table 4). A perfect symmetry between the left and right nasal passages and a complete connection of the upper, middle and lower parts is assumed. It is equipped with two shunting cavities modelling as Helmholtz resonators, the Sinus maxillaries at 6 cm from the outlet end with a resonance frequency of 350 Hz, and a smaller cavity, the Sinus frontalis located 2 cm more posterior. The coupling of the sinuses to the nasal tract is much dependent on their impedance as set by the particular length and area of their connections to the rest of the nose. Our standard values of resonator volumes are close to those suggested by Maeda (1982) and recently confirmed by the MRI studies of Dang et al. (1994).

The velo-pharyngeal port is presently controlled by the inlet area alone and an increase will cause a complementary decrease in the inlet to the mouth cavity. Possibly, the configuration of the port and its connecting point might differ from large to very small openings.

The minimal distinction of [m] versus [n] with a fronted tongue positions as brought out in Fig. 15 is entirely related to the difference in the shunting mouth cavity as seen from the level of the velo-pharyngeal port. The greater length for the [m] than for the [n] accounts for lower zeros, i.e. frequencies of zero input impedance to the
mouth cavity as seen from the velo-pharyngeal port. Associated poles are not strictly confined to the mouth cavity but show a similar length dependency. In the transfer functions of Fig. 15, we may identify the labial/alveolar contrast with a pole-zero pair which moves from 1185 Hz respectively 1227 Hz in [m] to 1705 Hz and 1746 Hz in [n]. The next higher pole-zero pair is found in a region around 2390 Hz for [m] and 2910 Hz for [n]. The constant sinus resonances are seen at 350 Hz and 1400 Hz. The 2000 Hz peak in both sounds is a standing wave resonance in the pharynx.

**Summary**

The novelty of our modelling compared to that of Fant (1960), Lin (1990), Fant (1992, 1993) is that a greater physiological relevance and flexibility is attained by a more complex structure retaining a minimum of three independent control parameters, the place of constriction, $X_c$ the constriction area, $A_c$, and the lip aperture parameter, $l_0/A_0$, as in the traditional three-parameter models. A number of additional parameters related to the lengths and shapes of cavity regions and constrictions may be given.
Figure 15. Minimal distinction between an alveolar [n] (solid line) and a labial [m] (dotted line). Top: VT area functions of the two nasal consonants with the nasal branch included. Bottom: Transfer functions with poles and zeros tabulated in Hz.

Table 4. Specification of the nasal area function in terms of tube length, coordinate from the nostrils and the insertion points of the nasal sinuses.
settings for individual vocal tract sizes and configurations. Some of them are constants, and others are treated as dependent variables which reflect a normal covariation with the three basic control parameters, or they can be given values departing from default values in order to increase the specificational accuracy, thus increasing the number of independent variables.

Examples of normal covariation are the positive correlation of inter-incisor distance with A_e in front vowels and with X_c in back-vowels. In back vowels, the degree of asymmetry of the shape of the pharyngeal constriction varies in accordance with how close X_c is located to the upper or the lower part of the pharynx. Another example is that the overall vocal tract length shows a positive covariation with the degree of lip rounding.

In practical use, the model is constrained by the dimensions chosen for the larynx tube and the Sinus piriformis cavities shunting the output of the larynx tube in the pharynx and also by the assumptions concerning the relative magnitude of the vocal cavity wall impedance and its distribution in the vocal tract.

To these constraints, add the specific VT overall length and default settings of the secondary parameters. A detailed perturbation analysis of VT dimensions and of boundary conditions has been performed. A method of inverse transformation for deriving VT parameters from formant frequencies has given promising results for stationary vowels. An extension to a short sentence is discussed in Båvegård & Fant (1995). However, improvements are needed to cope with temporal continuities. Also, the difference between any human VT and that of a particular modelling may cause ambiguities and failures of the model to adapt to a representative set of individual constraints.

The specific improvements, recently introduced in our modelling, pertain to the articulatory regions of the vowels [u:] and [ae:]. Our earlier vowel model lacked the relative shortening of the effective length of the tongue constriction in the velar region which prevented a sufficiently low location of F_3 of [u:] which typically should lie around 2300 Hz rather than at 2700 Hz as generated with the former model which has been used at the ICP. Savariaux et al. (1995). The modelling of [ae:] has been improved. It can be approached either from the front vowel region with an A_e greater than the average tract cross-sectional area, 4.5 cm^2 or from the extreme back vowel region by locating X_c close to the outlet of the larynx tube. The difficulty has been to retain a high F_1 and a low F_3 with an F_2 somewhat higher than the neutral value, typically F_1=650, F_2=1650, F_3=2400 Hz.

We now come close to these values but further improvements may be need.

Inversion experiments for deriving VT parameters from formant frequencies have been carried out. With one set of formant frequencies, the vowel [æ] became identified as a back vowel (Fig. 10), and in another set as a front vowel (Fig. 11). The extension to include consonantal articulations is based on the general philosophy introduced by Öhman (1967) that any consonant can be regarded as superimposed on a vowel configuration which is set by the particular coarticulatory pattern. Conversely, the consonantal part of the articulation must be adapted to the specific vowel configuration. In practice, the coarticulatory pattern will be influenced not only by the specific sequence of phonemic entities but also by the overlayed prosodic pattern, e.g. by the vowel-consonant contrast implied by the relative emphasis. We have not yet been in a position to model such variations but we believe that they could be handled with the present inventory of vocal tract area-function parameters.

The basic consonantal element of our model is a local constriction which substitutes a part of the specific vocalic area-function. In addition, the model is equipped with a nasal tract incorporating two nasal sinuses and there is a lateral bypass simulating the pathway along the cheeks as introduced already in Fant (1960). The consonantal part of a VT area function is specified by four parameters. Two of these, analogous to X_c and A_e of the vocalic model, have the primary function of specifying the location and degree of consonantal constrictions. The two additional parameters, pertaining to the effective length and shape of the consonantal constriction are dependent on the primary consonantal and also on the primary vocalic parameters. The boundary lines interfering consonantal and vocalic parts of the area function may be chosen with different curvature, straight lines or sinusoids.

In palatals and velars, the vocalic and consonantal constrictions coincide more or less but are still handled separately. Labial consonant closure and release are handled by the l/p/A_0 parameter in the vocalic system which leaves the tongue and thus the rest of the VT area function free to attain a combined vocalic and consonantal structure, e.g. with a region of consonantal narrowing separate from that of the vocalic constriction as in apical, retroflex articulations We have exemplified target configurations of labial and alveolar nasals,
laterals, and apical consonants coarticulated with a front vowel. The results are promising but we need more experience from modelling.

Discussion

Vowels

Our revised three-parameter model has gained in overall performance and flexibility to adapt to a wide range of vowel articulations, in the first place for male speakers. This has been achieved by a greater attention to system constants, some of which may be given the role of secondary parameters. Detailed perturbations involving variation of boundary conditions, some of the secondary parameters and the three basic parameters, have added to the general insight conveyed by nomograms. A model of this type is not free from inherent constraints. We encounter some difficulties in realising a complete range of human variations, and conversely, the model may generate patterns beyond human capability. The latter is less of a problem and can be avoided with proper constraints. An insufficient match can also be overcome by a proper choice of system constants and secondary parameters. We do not claim that this is a final version. As experience gained we expect further improvements to be made.

Compensatory forms of articulation providing the same or almost the same formant patterns exist. An example, well known from earlier studies, e.g. Boë et al. (1992), is the existence of mid vowel and back vowel versions of the vowel [u], the latter with a higher F3. The model is capable of producing the vowel [æ] both as a front vowel and a back vowel. The front version has a large Ac and the back version a small Ac but both have similar area functions, which does not disturb the concept of articulatory continuity.

In other instances, compensatory forms may exist in a local, more restricted Xc range in the nomogram, e.g. in the [i] and [y] region. These have to be considered in connection with the influence of VT boundary conditions. In continuous speech, this non-uniqueness can be overcome by imposing a continuity constraint in the inversion. We need more experience in this matter.

The outcome of our inversion experiments are discussed in Båvegård & Fant (1995). They support the typical clustering of place of articulation of Swedish vowels in three distinct regions of front, middle and back vowels.

The wall impedance has an important role in determining the lower bound of F1 which in turn will influence the Ac for a model match of narrow vowels.

Much remains to be done on modelling female vocal tracts. We have demonstrated the well known effect of a shorter pharynx which in [i] affects F2 more than F3. Another feature in this region is the extra high F3 when Ac approaches 0 as in a consonant [j].

From the study of Fant et al. (1973) and Fant (1975) we can draw some general conclusions: The female closed tract resonance frequency Fw is about 25% higher than for males. The front part of the lumped wall impedance is about the same as for males whilst the posterior part is substantially larger. The length of the female pharynx is about 30% shorter than for males but the volume is merely one half of the male volume. The length of the mouth cavity differs by a factor of about 10% and the volume by 30%. The general conclusion about scale factors is that area ratios follow the second power of length ratios and that the total volume of the female vocal tract is merely 60% of the male volume.

According to these data, the inherently smaller back to front area ratio of a female vocal tract might explain the relatively higher F1 in [a] type vowels (Fant, 1975), than what could be expected from a length transformation only.

Consonants

The extension to include consonantal articulations is based on the general philosophy introduced by Öhman (1967) that any consonant can be regarded as superimposed on a vowel configuration which is set by the particular coarticulatory pattern. Conversely, the consonantal part of the articulation must be adapted to the specific vowel configuration.

The basic consonantal element of our model is a local constriction which substitutes a part of the specific vocalic area-function. We have modelled target configurations of labial and alveolar nasals, laterals, and apical consonants co-articulated with a front vowel. The results are promising but we are still in an early phase of studying the static aspects. We need more experience from modelling onset and release phases, e.g. of combined lateral and medial pathways.

A second phase will be to adapt the model to match dynamic patterns of human speech. In practice, the coarticulatory pattern will be influenced not only by the specific sequence of phonemic entities but also by the superimposed prosodic pattern, e.g. by the vowel-consonant contrast implied by the relative emphasis. We have
not yet been in a position to model such variations but we believe that they could be handled with the present inventory of vocal tract area-function parameters.

Articulatory synthesis
It is generally considered that for programming connected speech a parameterisation of VT area functions in terms of constriction locations and shape factors is less effective than a direct factorised articulatory system of type Maeda (1990). One objection is the ambiguity of the $X_c$ parameter in the domain of neutral vowels, and the rapid switching of $X_c$ in the boundary between front and back articulation. However, this could be more of a problem in inversion than in synthesis.

A sequence of phonetic segments should be programmed in terms of spatially distributed and temporally overlapping control functions. In our future work, we will explore the possibility of achieving this within a framework of possible vocal tract area function states. These could be selected to conform with a set of phonetic prototypes of combinations of vocalie and consonantal components constituting primary and secondary targets, the latter intended for a sub-division of trajectories between primary targets. In addition, an inventory of modifiers of place and area of constrictions could be adopted to specify individual and dialectal variations, e.g. the “base of articulation” being more fronted or retracted and the degree of closure being greater or smaller than a norm. At any instant of time, the VT state should be considered as a weighted sum of contributions from earlier and later positioned targets within a prescribed time-window. Weighting functions can be organised so as to make possible both undershoot and overshoot and coarticulation which is a requirement for implementing variations in the prosodic frame such as varying emphasis and hyper-hypo articulation.

The decomposition of an articulatory gesture in a vocalic and a consonantal component, including glottal modifications, facilitates a proper structuring of time constants and thus major aspects of the shape and duration of weighting functions. This concept is of equal importance in the analysis phase of matching a model to human speech.

In summary, our proposed system will provide a shortcut between phonological-phonetic representations and VT area-functions but with an implicit use of knowledge of articulatory configurations and dynamics.

References


Båvegård M & Fant G (1995). From formant frequencies to VT area function parameters, ESPRIT BR SPEECHMAPS (6975), Delivery 29, WP2.3.


