A method to determine the area function of speech based on perturbation theory

Yu, Z.

journal: STL-QPSR
volume: 34
number: 4
year: 1993
pages: 077-096

http://www.speech.kth.se/qpsr
A METHOD TO DETERMINE THE AREA FUNCTION OF SPEECH BASED ON PERTURBATION THEORY

Zhenli YU

ABSTRACT

A new method to determine the vocal tract area function of speech from given measured formants based on a new perturbation algorithm is proposed. Two models of VT area function with parameters that have direct relation to formant frequencies are used. A new perturbation algorithm with emphasis on arbitrary area changes and an automatic procedure is proposed. A corresponding computer program is developed. Evaluating tests for the six Russian vowels and continuous speech are carried out with satisfactory results.

I. INTRODUCTION

The search of the Inverse Vocal Tract Transform (IVTT), i.e. to determine the area function of the vocal tract (VT), both for isolated vowels and for the vowels in connected speech, from given measured frequency parameters, in specific from formant frequencies, is well-known to be a difficult subject, especially if we want to do this work automatically. Among several reasons, one is the difficulty of modelling the area function of VT with model parameters that have an apparent mathematical relation to the spectrum parameters. Much attention has been attracted to this challenge (Schroeder, 1967; Mermelstein, 1967; Ladefoged, 1978; Lin, 1990, etc.). From the view of the mathematical relation, the perturbation method proposed by Schroeder seems to be an optimal approach.

According to his perturbation method, the area function can be determined using the relation between the perturbation (small change of area function) and the shift of the resonance formants from a uniform tube. In this way, the area function is presented by a special form of Fourier cosine series:

\[
A(x) = A_0 - A_0 \sum_{n=1}^{N} \left[ a_n \cos(2n-1)\pi x/L \right]
\]

where \( A_0 \) is the area of a uniform tube, \( a_n \) the coefficient of the nth odd cosine term. It means that \( A(x) \) can be seen as a uniform tube with changes as the sum of \( N \) terms. The \( N \) terms area change are called 'perturbation' of the area function. For a small perturbation, a simple relation between the coefficients and the corresponding resonance formants of the area function is

\[
a_n = -0.5 \cdot \frac{\delta f_n}{f_n}, \quad n=1,2,\ldots,N.
\]

where \( n \) denotes the resonance mode, \( f_n \) the nth mode formant of the uniform tube and \( \delta f_n \) the shift of \( f_n \) caused by the \((2n-1)\)th cosine perturbation term. This implies
that a small perturbation of the area function to a uniform tube with a single odd cosine
term will vary only one resonance frequency while leaving the others unchanged. Thus,
it proposes a possible way to infer the area function $A(x)$: if the $N$ resonance
frequencies and their shifts of a uniform tube are specified, the coefficients of the $N$
odd cosine terms can be determined and therefore $A(x)$ is determined. It should be
pointed out that this simple relation is only true for a small perturbation. In fact,
however, most real vowels have the formants far from that of the uniform tube. So the
situation of large perturbation, i.e. large area change from uniform tube must be
considered for real world IVTT.

In this paper, we will propose a method of determining the VT area function both of
isolated vowels from specific formants and the vowels in continuous speech from
connected sequences of measured formants, based on a new VT modelling and
perturbation algorithm developed from Schroeder's theorem. Special attention will be
paid to the large perturbation and the automatic realisation of the IVTT. The study
contains the following aspects: the modelling of VT area function, the robust algorithm
for an arbitrary area change, and an evaluating test for some isolated vowels and for
continuous speech.

II. MODELLING OF VT AREA FUNCTION

1. A simple model

As we know, any area function can generally be expanded to a certain kind of Fourier
series. For instance, we can expand it to the Fourier odd cosine series, purposely with
neglecting the error resulting from the uncompleted expansion. Thus, the VT area
function $A(x)$ can be expressed as

$$A(x) = A_0 - A_0 \sum_{n=1}^{N} \{ p_n \cos[(2n-1)\pi x / L] \}$$

(3)

where $p_n$ is the coefficient of the $n$th odd cosine term, $x$ is the distance from the
glottal and $L$ the whole length of the VT. Using connected sections with equal length
$x_0$, the $x$ can be expressed as $i \cdot x_0$ and $L$ as $l \cdot x_0$. We have

$$A(i) = A_0 - A_0 \sum_{n=1}^{N} \{ p_n \cos[(2n-1)\pi i / l] \}$$

(4)

From equation (4), we can determine $A(i)$ by selecting different values of $p_n$ with
certain constant $N$ and $l$. In the other words, we can use a set $\{ p_n \}$ as the modelling
parameters of the VT area function. Although such a model of VT area function is
rough, the advantage of it is that the modelling data $\{ p_n \}$ has a direct simple relation to
the resonance frequencies in the case of small perturbation and therefore can be
determined from measured formants by a perturbation algorithm developed below in
section III. From Eq.(1), (2), and (3), we get
in the case of small perturbation. We will show later that this relation is still useful in our approach to large perturbation algorithm. It should be noticed that the significance of this modelling is not the modelling itself but the easy extraction of the model parameters from measured formants, which will make the IVTT possible. In the following sections of this paper we will abbreviate this simple model as SM. Fig. 1 shows an example of an area function produced by the SM model parameters.

2. An improved model with constant larynx area

With the knowledge of the measured VT area data of the Russian vowels (Fant, 1960) which showed the fact that the larynx part of the VT with the length about 2 cm has the area about 2 cm$^2$. Based on this special feature, we adopt a constant area for the larynx part, i.e. $A(x) = A_{lar}$, for $x=L_{lar}$. And the area function producing formula equation (4) is substituted by

$$A(i) = \begin{cases} 
A_{lar}, & \text{for } i \cdot x_0 \leq L_{lar} \\
A_0 - A_0 \sum_{n=1}^{N} \{p_n \cdot \cos[(2n-1)\pi \cdot i / I] \}, & \text{for } i \cdot x_0 > L_{lar} 
\end{cases}$$

(6)

Usually, $A_{lar}$ is selected as 2 cm$^2$ and $L_{lar}$ as 2 cm. The behaviour of such a constraint will be discussed later. This model will be abbreviated to SML. Fig. 2 shows an example of an area function produced by the SML model parameters.

III. PERTURBATION ALGORITHM

A new algorithm of perturbation, developed from Schroeder's theorem, with emphasis to arbitrary area change is proposed in this research. The automation and flexibility of the perturbation will be specially considered.

1. Basic procedure

Before we introduce our perturbation algorithm, a basic procedure of perturbation according to Schroeder's theorem is described as following steps:

**Step 1:** Calculate the first $N$ formants $\{f_0n\}$ of a uniform tube.

**Step 2:** Specify the first $N$ target formants $\{f_{tn}\}$, and compute the error $\{\Delta f_n\}$ between $\{f_0n\}$ and $\{f_{tn}\}$.

**Step 3:** Suppose the initial perturbation, i.e. area change, in terms of $\{p_n\}$

$$p_n = -0.5 \frac{\Delta f_n}{f_0n}, \quad n=1,2,\ldots,N$$

(7)

**Step 4:** Derive the area function $A(i)$ from $\{p_n\}$ by Eq. (4) or Eq. (6).
Step 5: Find the formants of the perturbed area function $A(i)$ by direct VT transform. There are many successful approach to do this work, and the result will depend on what method of transform and formant finding are used. In our proposal, the 'impedance method' developed by Liljencrants and Fant (1975) is used.

Step 6: Infer a new area function to have its corresponding formants matching the target formants.

First, it starts by computing $\{\Delta f_n\}$, the error between the arrived formants $\{f_n\}$ and the targets $\{f_{tn}\}$. A decision is made here by comparing the sum of the absolute relative error

$$\varepsilon = \sum_{n=1}^{N} |\Delta f_n/f_n|$$ (8)

If $\varepsilon$ is less than a given threshold, the perturbation procedure ends up with successful result. Otherwise, it goes on.

Second, try to determine the new value of the area perturbation $\{p_n\}$ by adding an increase of perturbation (i.e. the adjustment of perturbation) $\{\Delta p_n\}$ to the previous value of perturbation $\{p_{np}\}$

$$p_n = p_{np} + \Delta p$$ (9)

and

$$\Delta p_n = -0.5 (f_n - f_{tn}) / f_n$$
$$= -0.5 \frac{\Delta f_n}{f_n}, \quad n=1,2,\ldots,N$$ (10)

Then, return to step 4 and repeat the process such that an iteration of perturbation is formed.

2. A new proposal of perturbation algorithm

However, the above procedure is a very simple and rough one. We have developed a perturbation algorithm with special emphasis in three aspects, which will make the algorithm reliable and flexible to succeed with the IVTT task.

A. Cross-sensitivity matrix

In a general case, the perturbation may be large. So, the conclusion that a small perturbation of the area function of a uniform tube with an odd cosine term will vary only one resonance frequency while leaving the others unchanged will no longer be correct. The so called cross-influence between different resonance mode now comes up, and the relation between area perturbation and formant shift will definitely be more complicated.

We introduce a matrix
and suppose the relation between \( \{A_p,\} \) and \( \{A_{fn}/f_{fn}\} \) to be

\[
\begin{bmatrix}
\Delta f_1 / f_1 \\
\Delta f_2 / f_2 \\
\vdots \\
\Delta f_N / f_N
\end{bmatrix}
= \begin{bmatrix}
b_{11} & \cdots & b_{1N} \\
b_{21} & \cdots & \cdots \\
\vdots & \ddots & \vdots \\
b_{N1} & \cdots & b_{NN}
\end{bmatrix}
\begin{bmatrix}
\Delta p_1 \\
\Delta p_2 \\
\vdots \\
\Delta p_N
\end{bmatrix}
\]

(12)

It is clear that the element \( b_{ij} \) of the \( B \) is a sensitivity factor showing how the \( i \)th mode resonance formant is influenced by the \( j \)th cosine term perturbation. With this meaning, we call \( B \) the cross-sensitivity matrix. There are two important properties of \( B \). One is that each element of \( B \) is perturbation-dependent (i.e. area-dependent or formant-dependent). And the other is that in the case of a small perturbation of a uniform tube, it degrades to a diagonal matrix with the diagonal element

\[
b_{ij} = \begin{cases}
0, & \text{for } i \neq j \\
-2, & \text{for } i = j
\end{cases}
\]

(13)

To find out the value of \( b_{ij} \), for certain \( p_n \), we need to calculate each formant shift \( (\Delta f_i / f_i)_j \), the \( i \)th formant shift caused by the \( j \)th perturbation term, by specifying a testing perturbation increase \( \{\Delta p_n\} \)

\[
\Delta p_n = \begin{cases}
\Delta p_j, & \text{for } n = j \\
0, & \text{for } n \neq j
\end{cases}
\]

(14)

Then, we obtain

\[
b_{ij} = (\Delta f_i / f_i)_j / \Delta p_j, \quad i=1,2,\ldots,N
\]

(15)

Repeat \( j \) from 1 to \( N \), then all \( b_{ij} \) are obtained. Because the final goal is to find the inferring perturbation increase \( \{\Delta p_n\} \) for given \( \{\Delta f_{fn}/f_{fn}\} \), we derive a matrix \( Y=\{y_{ij}\} \), the inverse matrix of \( B \), and have an inverse solution of Eq. (12)

\[
\begin{bmatrix}
\Delta p_1 \\
\Delta p_2 \\
\vdots \\
\Delta p_N
\end{bmatrix}
= \begin{bmatrix}
y_{11} & \cdots & y_{1N} \\
y_{21} & \cdots & \cdots \\
\vdots & \ddots & \vdots \\
y_{N1} & \cdots & y_{NN}
\end{bmatrix}
\begin{bmatrix}
\Delta f_1 / f_1 \\
\Delta f_2 / f_2 \\
\vdots \\
\Delta f_N / f_N
\end{bmatrix}
\]

(16)
It should be pointed out that because the B is area-dependent, B as well as Y should be updated after each perturbation to save accuracy. This is particularly necessary in case of large $\{\Delta f_n\}$. Of course, a new problem of too much computation will occur. To deal with this, a step-by-step procedure is proposed. If the $\{\Delta f_n\}$ is larger than a specified distance threshold, it is divided into linear steps and the B and Y is consequently re-computed for each step. This means the whole procedure goes to the formants target via several sub-targets.

**B. Initialisation of perturbation**

Because the B and Y is area-dependent, it is good to start the perturbation at some convenient value of $\{p_n\}$, and then measure the B and Y and therefore the $\{\Delta p_n\}$ which is optimal to arrive the target formants quickly. We call this initialisation of the perturbation. Otherwise the procedure will spend much time and risk failing the automatic iterative perturbation.

As a good solution, we use a pre-prepared perturbation code book. The code book contains a number of vectors, components of the code book. Each vector consists of a value of initial perturbation $\{p_n\}$ and its corresponding formants $\{f_n\}$ and $\{y_{ij}\}$. The code book is generated by setting a series reasonable $\{p_n\}$ and deriving the area function from them, and then getting the formants $\{f_n\}$ by direct transform and computing the inverse cross-sensitivity matrix $\{y_{ij}\}$ by the method described in section III.2.A.

Once the code book is prepared, the procedure can start from an optimum $\{p_n\}$ and $\{y_{ij}\}$ as initialisation perturbation. The option is chosen by finding a nearest formant $\{f_n\}$ matching to the target formants $\{f_{tn}\}$. When the optimal $\{f_n\}$ is found, the $\{p_n\}$ and $\{y_{ij}\}$ in the same vector of the code book is determined as the initial perturbation.

For a good performance, the code book should cover reasonable number of combinations of $\{p_n\}$ so that the initial perturbation value will be close to the goal and enable the procedure to succeed more quickly. However, an over large code book will be not realistic because of its memory requirement. A good proposal to generate a code book is that first we make a book as large as possible, and then compress it into a smaller one by the vector quantization (VQ) technique. The VQ method used here is 'binary clustering' (Makhoul et al., 1985).

**C. Constraint of unreasonable area function**

Sometimes, an unreasonable area function, most often with a negative area, will be derived either by over perturbation, i.e. too large area change, or by a peculiar combination of perturbation terms. If this happens, the perturbation procedure will be led to a wrong direction and consequently fail. To avoid this, we make two constraints for the area function produced by $\{p_n\}$. First, the increase of perturbation $\{\Delta p_n\}$ is adjusted to half, i.e. multiplied by a factor of 0.5, and then try to produce a new area function again. Second, if the area is still negative after the adjustment, it is constrained to a very small value, e.g. 0.1 cm$^2$. If the two attempts do not respond with a good result, it implies that the automatic perturbation procedure will not succeed. In this case, a possible remedy is to do the work by manual mode perturbation. Some
attempts such as, for instance, changing the VT length, adjusting certain coefficient alone carefully and gradually while leaving the others unchanged may be effective. The technique of the manual mode will introduced later in this section.

D. A computer program of inverse VT transform: IVTTPG

According to the algorithm described above, we developed an inverse VT transform program, named IVTTPG, in C code for Apollo. The structure of the main part of the program is shown by Fig. 3. It provides both automatic and manual operation modes. In the automatic mode, the target formants can be input either from keyboard or from a data file which contains connected formants series and allows the perturbation result to be written back. In the manual mode, the program provides the function of specifying target formants and perturbation increase either absolutely or relatively by the mouse technique. This function is particularly useful as a remedy way when the automatic iteration can not give a satisfactory result. Moreover, there is a function of plotting the N terms perturbation (area change) profile and the derived VT area function (the inferred VT area function), the corresponding formants and the error between the formants and the targets on the screen. The determined area function can be output in data form. Another important ability of the IVTTPG is to change the VT length. This ability is particularly useful since the human vowels are produced by different vocal tract length.

IV. Test example of determination of Russian vowels

The original area function data of the six Russian vowels in Fant (1960) are used to derive the target formants for IVTT. The direct VT transform to estimate the formants of the original area function is taken by the program 'TRACTTALK' (Lin, 1993), an in-house software for speech production and synthesis. It should be explained that we do not use the formants data measured by Fant (1960) directly as target since we will evaluate the determined area function by direct VT transform by 'TRACTTALK'. Both the SM and SML models are used in the test.

1. Determination of the VT area function

The test starts with finding the first four formants of the original area data by 'TRACTTALK'. To make it simple and identical, the VT is supposed to be lossless tube (hard wall) and the lip radiation as free. Looking at the first four formants found as target formants, IVTTPG performs the perturbation procedure and determines the inferred area function. Fig. 4 to Fig. 9 show the determined area functions with SM and Fig. 10 to Fig. 15 with SML. Fig. 16 to Fig. 21 give the comparison between SM and SML model.

From Fig. 4 to Fig. 15, we can observe that all of six determined area functions are close to the original area function, except the part where the original area have abrupt change to large value, for instance in the case of the vowel /u/. The big error in this case is because the inferred area only contains limited spectrum information which can be thought as an effect of low-pass filtering. Fig. 10 to Fig. 15 show better improvement by SML model, especially in the part of larynx. Another interesting phenomenon can also be observed from Fig. 16 to Fig. 21 that the larynx constraint
Ala of SML model is less than the area in the same part of SM model in each vowel, and the area in lip of SML model is less than that of SM model, too.

Table 1 gives parameter data of the inferred area function. From Table 1, we can find the fact that the parameters of the two models are little different. Generally, \( p_n \) of the SML are less than that of the SM model. This means that the larynx constraint results in less perturbation of the area function.

Table 1. Parameter data of the inferred area function of six Russian vowels

<table>
<thead>
<tr>
<th></th>
<th>SM parameter</th>
<th>SML parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( p_1 )</td>
<td>( p_2 )</td>
</tr>
<tr>
<td>/a/</td>
<td>0.7091</td>
<td>-0.3650</td>
</tr>
<tr>
<td>/o/</td>
<td>0.7183</td>
<td>-0.5326</td>
</tr>
<tr>
<td>/u/</td>
<td>-0.1558</td>
<td>-0.8906</td>
</tr>
<tr>
<td>/y/</td>
<td>-0.7656</td>
<td>0.2849</td>
</tr>
<tr>
<td>/i/</td>
<td>-0.5041</td>
<td>0.7188</td>
</tr>
<tr>
<td>/e/</td>
<td>-0.2376</td>
<td>0.4679</td>
</tr>
</tbody>
</table>

2. Analysis of the formants error

To evaluate the two models from the view of formants, we re-estimate the formants of the inferred area function for both models by direct VT transform using 'TRACTTALK'. Table 2 and Table 3 give the re-estimated formants, and the error of them between SM model to the original and the error of them between the SML model to the original.

From Table 2 and Table 3, we can first find that both models have small formant errors for the six vowels, with the biggest one of 4.33 per cent of SML model of the vowel /a/, and the smallest one of zero of SM model of the vowel /y/. It means the algorithm and the two models perform well with the IVTT subject. Second, some difference can be observed from the errors between the formants of inferred area of the two models to the formants of the original area function, in the same condition of lossless VT and free radiation at lip. The errors of SML are bigger than that of SM, except in the case of vowel /o/. And the average error of six vowels of SML model is also bigger than that of SM model. However, this not necessarily means that the SML model is worse than the SM model because the result is only obtained in the condition of lossless VT and free radiation. If the tract wall condition and radiation condition have changed the result might be different. This falls outside our present scope and is not further detailed.
Table 2. Comparison of formant errors for the inferred urea function with SM and SML models

| vowel | original area | SM rea | SML area
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>formants</td>
<td>formants</td>
<td>error (%)</td>
</tr>
<tr>
<td>/a/</td>
<td>670</td>
<td>665</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>1176</td>
<td>1165</td>
<td>0.94</td>
</tr>
<tr>
<td></td>
<td>2531</td>
<td>2514</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td>3754</td>
<td>3731</td>
<td>0.56</td>
</tr>
<tr>
<td>total</td>
<td>2.92</td>
<td></td>
<td></td>
</tr>
<tr>
<td>/o/</td>
<td>530</td>
<td>527</td>
<td>0.57</td>
</tr>
<tr>
<td></td>
<td>950</td>
<td>948</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td>2416</td>
<td>2410</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>3441</td>
<td>3434</td>
<td>0.20</td>
</tr>
<tr>
<td>total</td>
<td>1.23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>/u/</td>
<td>252</td>
<td>251</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>603</td>
<td>600</td>
<td>0.50</td>
</tr>
<tr>
<td></td>
<td>2364</td>
<td>2353</td>
<td>0.47</td>
</tr>
<tr>
<td></td>
<td>3679</td>
<td>3659</td>
<td>0.54</td>
</tr>
<tr>
<td>total</td>
<td>1.91</td>
<td></td>
<td></td>
</tr>
<tr>
<td>/i/</td>
<td>265</td>
<td>265</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>1613</td>
<td>1622</td>
<td>-0.56</td>
</tr>
<tr>
<td></td>
<td>2402</td>
<td>2410</td>
<td>0.33</td>
</tr>
<tr>
<td></td>
<td>3457</td>
<td>3476</td>
<td>-0.22</td>
</tr>
<tr>
<td>total</td>
<td>1.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>/e/</td>
<td>228</td>
<td>229</td>
<td>-0.44</td>
</tr>
<tr>
<td></td>
<td>2240</td>
<td>2246</td>
<td>-0.27</td>
</tr>
<tr>
<td></td>
<td>3262</td>
<td>3272</td>
<td>0.34</td>
</tr>
<tr>
<td></td>
<td>3770</td>
<td>3780</td>
<td>0.27</td>
</tr>
<tr>
<td>total</td>
<td>1.32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>/e/</td>
<td>431</td>
<td>432</td>
<td>-0.23</td>
</tr>
<tr>
<td></td>
<td>1997</td>
<td>2003</td>
<td>-0.30</td>
</tr>
<tr>
<td></td>
<td>2888</td>
<td>2897</td>
<td>-0.31</td>
</tr>
<tr>
<td></td>
<td>3824</td>
<td>3840</td>
<td>-0.42</td>
</tr>
<tr>
<td>total</td>
<td>1.26</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3. Comparison of formant errors for the inferred area function with SM and SML model

<table>
<thead>
<tr>
<th></th>
<th>/a/</th>
<th>/o/</th>
<th>/u/</th>
<th>/i/</th>
<th>/e/</th>
<th>average</th>
</tr>
</thead>
<tbody>
<tr>
<td>SM</td>
<td>2.92</td>
<td>1.23</td>
<td>1.91</td>
<td>1.11</td>
<td>1.32</td>
<td>1.26</td>
</tr>
<tr>
<td>SML</td>
<td>7.02</td>
<td>1.14</td>
<td>2.04</td>
<td>2.36</td>
<td>1.41</td>
<td>1.40</td>
</tr>
</tbody>
</table>
V. An example of determination of area function parameters of continuous speech

We have also determined the area function of the vowels contained in continuous speech using the program IVTTPG. A speech data file 'WD131.FMC' which contains connected sequences of estimated formants frame by frame.

The IVTTPG read the target formants \( f_n \) from the kth frame and ran the perturbation iteration procedure automatically. Once the area function was determined, the parameters \( p_n \) and the error between the formants found and the target formants were written back to the data file. Two thresholds were set for ending the automatic procedure. One was the error threshold. If the error was less than the threshold, the iteration for the current frame ended. The other was the maximum iteration times for each frame. If the iteration repeated times more than the threshold, it ended. In this case, the perturbation result might not be successful. However, the IVTTPG was able to choose the best \( p_n \) with smallest formants error, it had arrived in the iteration as the final result.

Fig. 22 shows the profiles of the target formants, the determined parameters of area function and the error. Table 4 gives their corresponding datum. Fig. 23 is the spectrogram of the original continuous speech.

From Fig. 22 and Table 4 we can observe that most of 418 frames have very small errors, less than 1 per cent. The biggest error is 26 per cent. There are three parts of the frame sequence, timed 1.6055 sec. to 1.6380 sec., 3.5417 sec. and 4.1758 sec., where the errors bigger than 5 per cent occur. By comparing the formants data of these parts in Table 4 with the spectrogram in the same parts in Fig. 22, we can find that the estimation of the formants in these frames are unreliable. For example, at the frame timed 3.5417 sec., the first formant is estimated as 78 Hz, while the spectrogram shows that this is not correct. The incorrect estimation of formant will provide a combination of target formants which is impossible to produce by a reasonable VT area function. So we should not expect the IVTTPG to get a good inferred area function for such a target.

Also, it should be pointed out that the VT length was defined to be constant, i.e. 17.5 cm, in this test. This is another reason that the IVTTPG could not do successful work for the frames where the vowels have VT length much longer or shorter than 17.5 cm. For example, the vowel /u/ has the VT length about 19.5 for ordinary human speaker. And the formants in the frames timed 1.8516 sec. to 1.9186 sec. seem to be derived from /u/, according to our knowledge of the frequency behaviour. Thus, it is understandable if big errors in these frames occur.

VI. SUMMARY

From this study, we can get the following main important conclusions:

1. The simple model of VT area function enables us to establish the relation of the area perturbation and the formants and therefore can make the inverse VT transform subject easier to realise by computer program.

2. The improved VT model from the simple one with a larynx constraint is also meaningful.

3. Based on the idea of cross-effect within different resonance modes in the perturbation process, the cross-sensitive matrix, with which a new formula to
express the relation between perturbation and formants can be established, makes it possible to realise the IVTT especially in case of large perturbation.

4. The code book of perturbation is useful to start the IVTT procedure with an optimum initial perturbation which makes the procedure much faster.

5. The new algorithm of perturbation is very efficient to perform the IVTT and is successfully designed to the computer program IVTTPG. The IVTTPG can provide a powerful tool to determine VT area function from measured formants of vowels. This will be significant to the speech production and speech synthesis research subjects.

At the end of this paper, we should point out some problems to be investigated in more detail.

First, the simple model SM as well as the model with larynx constraint SML should be studied concerning the acoustic properties physically.

Second, in the test for the Russian vowels, we have only used the lossless vocal tract and free radiation at lip conditions to find the formants. These conditions are very simple and can only prove the efficiency of the two models in the perturbation realisation, but not the advantage and perfection of the models themselves. It is necessary to test the models under different conditions.

Finally, the VT length is a stubborn problem in the automatic process of IVTT for continuous speech. There will be infinite answers of the inferred area function as the VT length varies. The question is, how can we select an optimal value of the length for given input of target formants? Although this problem is not a challenge only for our research subject here, it will be significant to consider this problem with attention especially if we try to apply the perturbation algorithm to real speech processing techniques.

ACKNOWLEDGEMENT

This work is a part of my research project as a guest researcher at the Dept. of Speech Communication and Music Acoustics, Royal Institute of Technology (KTH), Sweden, from November, 1992 to November, 1993.

I am extremely grateful to my supervisor, Dr. Johan Liljencrants, for his generous guidance. Without his significant suggestions and help in many ways during my research, this work would not have succeeded.

Thanks also to Professor Gunnar Fant and Dr. Jan Gauffin who have given me kind help with encouragement and suggestions. Special thanks to Dr. Qiguang Lin and Mats Båvegård for offering their program 'TRACTTALK'.

The work has been financially supported by 'Bao Zhao-long and Bao Yu-gang Scholarship for Chinese Student Studying abroad' and by 'The Natural Science Fund of Zhejiang Province, P.R.China' in some way.

Last, but not least, I wish to thank my wife Shangcui and my son Sijing. They have not only overcome the difficulties due to my leaving them, but also encouraged me during all the time.
REFERENCES


Fig. 1 Area function modelling by SM

Fig. 2 Area function modelling by SML
BEGIN

input target formants

perturbation initialisation

computer error $\varepsilon$

$\varepsilon < \text{threshold}?$

yes $\rightarrow$ successful IVTT

no $\rightarrow$ ItNm $> \text{threshold}?$

yes $\rightarrow$ failed IVTT

no $\rightarrow$ $\varepsilon < \text{subthreshold}?$

yes $\rightarrow$ test sensitivity $B$

no $\rightarrow$ infer area perturbation

make area function

find new formants

$\rightarrow$ determined area function $\rightarrow$ END

select best area function

$\rightarrow$ inverse matrix $Y$

$\rightarrow$ END

* ItNm: iteration number

Fig. 3 The structure of the main part of IVTTPG
Fig. 4 - 6. SM area and the original area of /a/, /o/, and /u/.

Fig. 7 - 9. SM area and the original area of /∀/, /i/, and /e/.
Fig. 10 - 12. SML area and the original area of /a/, /o/, and /u/ area cm²

Fig. 13 - 15. SML area and the original area of /æ/, /i/, and /e/
Fig. 16 - 18. SM area and SML area of /a/, /o/, and /u/.

Fig. 19 - 21. SM area and SML area of /∀/, /i/, and /e/.
Fig. 22. Profiles of target formants, determined parameters and errors of the continuous speech
Fig. 23. Original spectrogram of the continuous speech