ON THE ONE-TO-ONE CORRESPONDENCE OF SPECTRAL WITH GEOMETRICAL CHARACTERISTICS OF THE LPC VOCAL TRACT MODEL.

E.Hennes

1. INTRODUCTION

In acoustic phonetics the vocal tract is often approximated by means of N successive cylindrical tubes of lengths ℓ_i and cross-sectional areas S_i , i=0,1,..,N-1 (fig 1). Each tube-section is a one-dimensional wave transmitter. The wave is characterized by the sound pressure p(x,t) and the volume velocity u(x,t). The N-tube is terminated by a glottal acoustical resistance Z_g and a complex radiation load Z_r [Flanagan,1972].

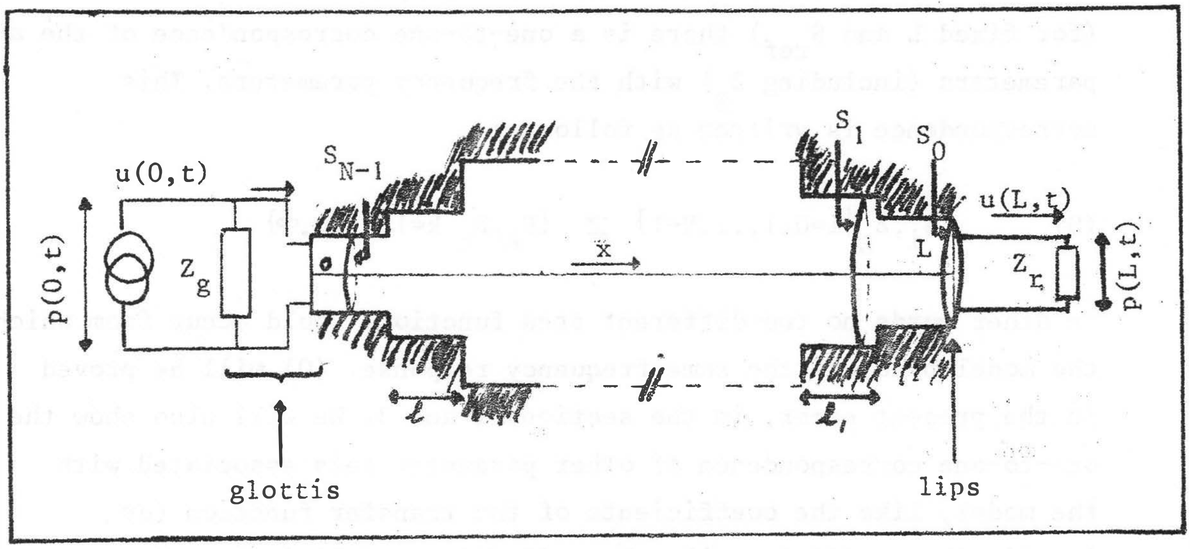

fig 1. N-tube model of the vocal tract.

The linear prediction model of the vocal tract [Markel & Gray,1976] is a simplified version of Flanagan's model. All section of the N-tube are of equal length. The wave transmission in the sections is lossless and the radiation load Z_r is zero. All dissipation of energy is concentrated in the glottal resistance Z_g. The model is fully characterized by its total length L = N.ℓ, the glottal resistance, and the cross-sectional areas S_i, i=0,1,..,N-1. From these parameters N characteristic frequency parameters, being the first N/2 formant frequencies and their correspon-ding bandwidths, can be predicted. All higher formant frequencies and corresponding bandwidths are determined by these N characteristic frequencies (see section 2).

Conversely we might ask whether it is possible to determine the geometry parameters from these frequencies. This possibility is not self-evident. The lossless twintube model [Mol,1970] ,for instance,shows that there are at least two different twintubes with the same formant frequencies, even if the length L and a reference area $S_{ref} \varepsilon \{ S_0, S_1, .., S_{N-1} \}$ are the same for both tubes. Symmetries in lossless tubes often result in identical frequency responses [Schroeder,1967].

The linear prediction model provides us with a means to calculate a set of areas (including Z_g) from the characteristic frequencies if L and S_{ref} are given. To carry out this calculation it is necessary for the glottal resistance to be finite, not zero, i.e. the bandwidths B_k must not be zero (k=1,2,..). These areas determine an N-tube that has the above-mentioned frequency characteristics. In most of the literature [c.f. Wakita,1973 ; Markel & Gray,1976] it is assumed that (for fixed L and S_{ref}) there is a one-to-one correspondence of the area parameters (including Z_g) with the frequency parameters. This correspondence is written as follows :

(0) $$\{S_i, Z_g \mid i=0,1,..,N-1\} \rightleftarrows \{F_k, B_k \; k=1,2,..,\infty\}$$

In other words no two different area functions would occur from which the model predicts the same frequency response. (0) will be proved in the present paper, in the sections 2 and 3. We will also show the one-to-one correspondence of other parameter sets associated with the model, like the coefficients of the transfer function (or 'prediction coefficients'), the reflection coefficients and the zeros of the transfer function. So Wakita's method of estimating the area function of the vocal tract from the speech wave is justified from a theoretical point of view. Actually Wakita uses the prediction coefficients as input for the calculation because these can directly be obtained by linear prediction of the (preemphasized) speech signal.

In section 2 the relation between different sets of parameters is discussed. In section 3 the one-to-one correspondence of the prediction coefficients with the reflection coefficients is proved. In section 4 some models related to the linear prediction model, such as the resistive radiation load model and the lossless N-tube model, are discussed.

2. THE LINEAR PREDICTION MODEL

2.1 Description of the transfer function.

Acoustically the one-dimensional vocal tract model can be described with the volume velocity function $u(x,t)$, which is defined as the product of $S(x)$, the cross-sectional area of the tract, and $v(x,t)$, the particle velocity in the x-direction :

$$(1) \qquad u(x,t) = v(x,t) \,.\, S(x)$$

For each tube section $u(x,t)$ obeys the one-dimensional wave equation. The general solution of this equation is a linear combination of two waves, travelling in the positive and negative x-directions respectively. Continuity of sound pressure ans volume velocity between the sections is assumed. The sound pressure at the lips is set equal to zero. The pressure-velocity relation at the glottal side of the tract is determined by the glottal resistance Z_g. Under these conditions $u(x,t)$ can be solved.

We can look upon the vocal tract as a linear filter. The volume velocities at the glottis and the lips are the input and output signals respectively (fig 1). The transfer function of this filter can be written as the Laplace Transform (see App. I) :

$$(2) \qquad H(s) = C(z) \,/\, A(z) \,,$$

where $z = \exp(sT)$,

$T = 2\ell/c$,

c = velocity of sound ,

$C(z) = k.z^{-\frac{1}{2}N}$ is a phase function of constant amplitude k ,

$A(z)$ = polynomial in z^{-1} of degree N.

Hence $H(s)$ is a function of $\exp(sT)$. So the frequency response $H(2\pi jF)$, where F is the frequency, is a function of $\exp(2\pi jFT)$, and consequently it is periodical with period $\Delta F = 1/T$. This periodicity is typical of the model, for it is a consequence of the fact that the section length

ℓ is constant. The periodicity of the frequency response implies a discrete impulse response [Brigham,1974]. From the notation $z = \exp(sT)$ it is clear that $\mathcal{H}(z) := \mathcal{H}(\exp(sT)) = H(s)$ is the z-transform of the impulse response. Note that this result is obtained without sampling the impulse response. $\mathcal{H}(z)$ comprises the frequency response over the whole frequency range from zero to infinity. The frequency amplitude response of the filter is within a constant $k = |C(z)|$ determined by $A(z)$.

2.2 Derivation of the transfer function from the area function.

The polynomial $A(z)$ contains the information about the tube areas and the glottal resistance in the following way. The reflection coefficients are defined as :

$$\mu_i := \frac{S_{i-1} - S_i}{S_{i-1} + S_i} \quad ; i=1,2,..,N. \tag{3}$$

with $S_N := \rho c/Z_g$.

ρ is the density of the air. In practice S_N is of the order 0.4 cm^2, and $\mu_N > 0$. The glottal termination is represented by a semi-infinite tube of area S_N (fig 2). The volume velocity $u_N(t)$ in this tube can be compounded of the forward and backward travelling waves $u_N^+(t)$ and

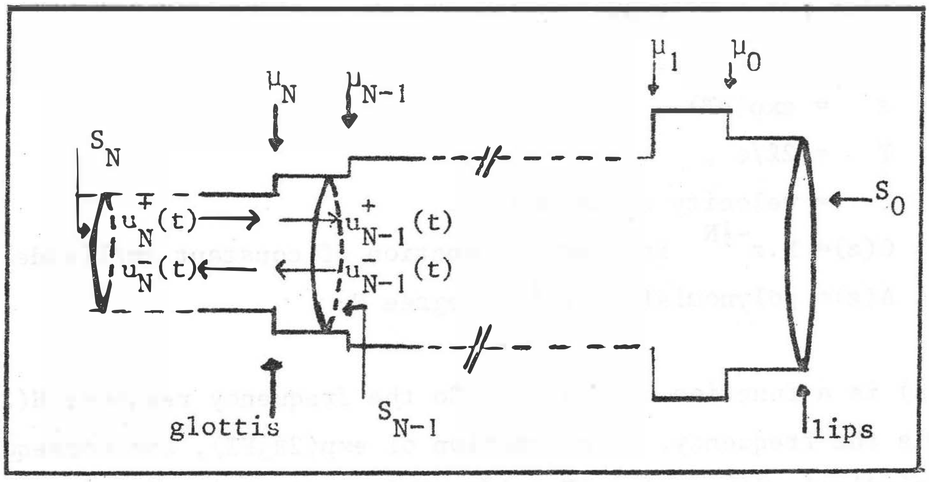

fig 2. Model with the artificial S_N-section.

$u_N^-(t)$ respectively. $u_N^+(t)$ represents half of the glottal volume velocity $u_g(t)$ [Markel & Gray,1976,p.71]. If $u_g(t) = 0$ in a certain interval of time, then $u_N^-(t)$ is the flow through the glottal resistance. Thus in fig 2 this flow is the diffracted part of the backward travelling wave $u_{N-1}^-(t)$ in the S_{N-1}-section. The energy carried by this diffracted wave represents the loss in the glottal resistance. The two limits $Z_g = 0$ and $Z_g = \infty$ correspond to $S_N = \infty$, $\mu_N = -1$ and to $S_N = 0$, $\mu_N = 1$ respectively. In other words they correspond to a totally open and closed tube ending at the glottal side respectively. In these cases there is total reflection of waves at the glottis and there is no loss of energy. Tubes with lossless terminations are referred to as lossless tubes in this paper. Only the closed glottis model ($\mu_N = 1$) is of interest as a lossless vocal tract model, for μ_N is never valued below 0.

According to Markel & Gray the function A(z) can be obtained with a recurrence relation. This reads :

$$
\text{(4)}\qquad
\begin{aligned}
A_m(z) &= A_{m-1}(z) + \mu_m \cdot z^{-m} \cdot A_{m-1}(1/z) \;, \quad m=1,2,..,N. \\
A_0(z) &= 1 \;, \\
A(z) &= A_N(z) \;.
\end{aligned}
$$

The functions $A_m(z)$ are polynomials in z^{-1} of degree m, with real coefficients. If $a_{m,i}$ denote the coefficients of z^{-i} (i=0,1,2,..,m), then $z^{-m} \cdot A_m(1/z)$ is obtained from A(z) by reversing the order of the coefficients :

$$
\text{(5)}\qquad z^{-m} \cdot A_m(1/z) = \sum_{i=0}^{m} a_{m,m-i} \cdot z^{-i} \;, \quad m=1,2,..,N \;.
$$

Thus the recurrence relation (4) can also be written as

$$
\text{(6)}\qquad
\begin{aligned}
a_{m,i} &= a_{m-1,i} + \mu_m \cdot a_{m-1,m-i} \;, \quad i=1,2,..,m-1 \;. \\
a_{m,0} &= 1 \;, \quad a_{m,m} = \mu_m \;, \quad m=1,2,..,N.
\end{aligned}
$$

According to the fundamental theorem of algebra A(z) can be factorized :

$$
\text{(7)}\qquad A(z) = 1 + \sum_{i=1}^{N} a_{N,i} \cdot z^{-i} = \prod_{i=1}^{N} (1 - z_i z^{-1}) \;.
$$

There is a one-to-one correspondence of the N zeros z_i with the prediction coefficients $a_{N,i}$ $(i=1,2,..,N)$. As $a_{N,i}$ are real coefficients the zeros either occur in complex conjugate pairs, or they are real.

2.3 The relation between the resonances and the zeros of the transfer function.

The vibrational modes of the vocal tract model can be derived from the zeros z_i of $A(z)$ according to (App I) :

$$(8) \qquad z_i = \exp(-\pi B_k T + 2\pi j F_k T) \quad , \quad i=1,2,..,N \; , \quad k = 1,2,..\infty$$

Every mode, or resonance, is characterized by the parameter-pair (F_k, B_k), where F_k denotes the k-th formant frequency and B_k its corresponding bandwidth. To every zero z_i corresponds an infinite series of resonances with equal bandwidths but with formant frequencies which differ by integral multiples of $1/T$ from each other. This is the result of the above-mentioned periodicity of $H(j\omega)$ (see section 2.1).

Finally, in section 3 the following properties of the model are used. They follow from the reasonable assumptions that no two formant frequencies are equal, and that $\mu_N > 0$. ')

a. For even N the transfer function has no real zeros.

b. For odd N the transfer function has precisely one real zero on the the negative axis corresponding to a fixed formant frequency $F_{\frac{1}{2}(N+1)} = 1/2T$.

c. The N zeros z_i that characterize the transfer function are determined by N characteristic frequencies. These frequencies can be obtained from measurement of the first [N/2] '') resonances. If N is odd the bandwidth $B_{[N/2]+1} = B_{\frac{1}{2}(N+1)}$ is needed too.

') The proof of these properties is simple and is not given in this paper. If the reader so wishes the author can send them .

'') [N/2]:=ENTIER(N/2) is the largest integer smaller than or equal to N/2, i.e. its truncation.

2.4 Review

The linear prediction model provides us with an infinite series of resonances (F_k, B_k), $k=1,2,..$. These pairs can be derived from N coefficients $a_{N,i}$ of $A(z)$, $i=1,2,..,N$ with (7) and (8). On the assumptions that no two formant frequencies are equal and that $\mu_N > 0$, $a_{N,i}$ can be derived from N measured frequencies by means of (7) and (8) too. In these formulas the zeros z_i of $A(z)$ are used. On the other hand the set $a_{N,i}$ can be calculated from the reflection coefficients μ_i with (6). In their turn the reflection coefficients are calculated from the section areas and the glottal resistance with (3). The whole scheme is illustrated in fig 3.

$$\begin{tikzcd} \{ S_{i-1}, Z_g \} \arrow[r, shift left] & \{ \mu_i \} \arrow[l, shift left] \arrow[r, shift left] & \{ a_{N,i} \} \arrow[l, dashed, shift left] \arrow[d, shift left] \\ \{ F_k, B_k \} \arrow[r, shift left] & \left\{ \begin{array}{c} F_j, B_j; \\ (B_{[\frac{1}{2}N]+1} \\ \text{if } N \text{ odd}) \end{array} \right\} \arrow[l, shift left] \arrow[r, shift left] & \{ z_i \} \arrow[l, shift left] \arrow[u, shift left] \end{tikzcd}$$

$i = 1,2,..,N$,

$j = 1,2,.., [\frac{1}{2}N]$,

$k = 1,2,..,\infty$.

—— : confirmed,

-----: to be confirmed.

fig 3. N-tube parameter relations.

From this figure it can be seen that all mappings are one-to-one if the dotted arrow between $\{\mu_i\}$ and $\{a_{N,i}\}$ holds. In the next section we will examine the conditions necessary for the validity of the dotted arrow.

3. ON THE RELATION BETWEEN THE REFLECTION COEFFICIENTS AND THE COEFFICIENTS OF THE TRANSFER FUNCTION

In this section the recurrence relation (6) is used to prove that from

$$(9)\qquad \begin{cases} \mu_2 \neq -1 & \text{for } N = 2 \\ \mu_2 \neq -1,\ \mu_i \neq \pm 1 \ (i=3,4,..,N) & \text{for } N = 3,4,... \end{cases}$$

it follows that the mapping

$$(10)\qquad \{\mu_i\} \rightarrow \{a_{N,i}\}\quad ,(i=1,2,..,N).$$

is one-to-one, where N=1,2,..... .

Outline of the proof : first the cases N=1 and N=2 are considered (3.1). For N=3,4,... (10) will be proved by the principle of mathematical induction (3.2). For the sake of convenience we rewrite (6), omitting the trivial terms $a_{m,0} = 1$:

$$(11)\qquad a_{m,i} = a_{m-1,i} + \mu_m \cdot a_{m-1,m-i}\quad ,\ (m=1,2,..,N;\ i=1,2,..,m-1)$$

$$a_{m,m} = \mu_m .$$

3.1 <u>The cases N = 1 and N = 2.</u>

<u>N=1</u>. In (11) there is only one step to do, viz. m=1. Thus $a_{1,1} = \mu_1$ and (10) holds. There is no restrictive condition on the reflection coefficient.

<u>N=2</u>. There are two recursive steps :

$$(12)\qquad \begin{cases} m = 1 : & a_{1,1} = \mu_1 \\ m = 2 : & a_{2,1} = \mu_1 \cdot (1+\mu_2)\ , \\ & a_{2,2} = \mu_2 \end{cases}$$

If $\mu_2 \neq -1$ the two equations (12) have one non-trivial solution for μ_1 and μ_2. On this condition $a_{2,i}$ and μ_i (i=1,2) determine each other uniquely.

3.2 The case N > 2.

We use the principle of mathematical induction.

N=3. The first two recursive steps are similar to the case N=2 (see 3.1). With (12) the third step is realized :

$$
\begin{aligned}
a_{3,1} &= \mu_1\cdot(1+\mu_2) + \mu_3\cdot\mu_2 , \\
a_{3,2} &= \mu_3\cdot\mu_1\cdot(1+\mu_2) + \mu_2 , \qquad (13)\\
a_{3,3} &= \mu_3 .
\end{aligned}
$$

Inspection of these equations shows that they are only independent if

$$
\begin{aligned}
\mu_3 &\neq 1,-1 , \\
\mu_2 &\neq -1 . \qquad (14)
\end{aligned}
$$

Thus with these restrictions there is a one-to-one relation between $a_{3,i}$ and μ_i , i=1,2,3.

N -1 $\Rightarrow$ N. Consider an N-tube model which is restricted to (9). This tube has a transfer function $A_N(z)$. From this we construct an (N-1)-tube with transfer function $A_{N-1}(z)$, omitting the glottal section with area S_N. The shorter tube can be looked upon as a vocal tract model with glottal section S_{N-1}. For this tube (9) holds too and the induction statement is according to formula (10) :

$$(15) \qquad \{ a_{N-1,i} \} \leftrightarrows \{ \mu_i \} \qquad , i=1,2,..,N-1.$$

From (11) it follows that the transfer functions of the two tubes are related by :

$$
\begin{aligned}
a_{N,i} &= a_{N-1,i} + \mu_N\cdot a_{N-1,N-i} \qquad ,i=1,2,..,N-1. \\
a_{N,N} &= \mu_N . \qquad (16)
\end{aligned}
$$

$a_{N,N}$ and μ_N being equal, they determine each other uniquely. It now only remains to prove that there exists no other function $A'_{N-1}(z)$ with coefficients $a'_{N-1,i}$ that gives the same coefficients $a_{N,i}$ in (16).

The mapping

$$\{ a_{N-1,i} \} \rightarrow \{ a_{N,i} \} \quad , \; i=1,2,..,N-1.$$

is linear according to (16) and is expressed by N-1 homogeneous linear equations. The mapping is only one-to-one if the determinant of the matrix of coefficients is unequal to zero :

$$(17) \qquad \begin{vmatrix} 1 & 0 & \cdots & \cdots & 0 & \mu_N \\ 0 & 1 & & & \mu_N & 0 \\ \vdots & & 1 \; \ddots & \cdot^{\cdot^{\cdot}} \; \mu_N & & \vdots \\ \vdots & & \cdot^{\cdot^{\cdot}} & \ddots & & \vdots \\ \vdots & & \mu_N & 1 & & \vdots \\ 0 & \mu_N & & & 1 & 0 \\ \mu_N & 0 & \cdots & \cdots & 0 & 1 \end{vmatrix} \neq 0$$

The form of the center of the matrix depends of whether N is even or odd. In both cases we can see that (17) implies that $\mu_N \neq 1,-1$. Thus putting $a'_{N-1,i}$ into (16) under this restriction, we can only get $a_{N,i}$ if $a'_{N-1,i} = a_{N-1,i}$.

This completes the proof.

4 N-TUBE MODELS RELATED TO THE LINEAR PREDICTION MODEL

4.1 Model with a resistive radiation load.

For area functions with a mouth area S_0 of 5 cm^2 or more it is rather the radiation load than the glottal resistance that is responsible for the bandwidths, especially for those of the higher formants (> 2 kHz) [Flanagan, 1972, p.63]. The simplest model is a constant and resistive load R_r, which is the real part of Z_r. Analogous to the representation of the glottal termination in section 2.2 R_r can be represented by a semi-infinite tube with cross-sectional area $S_{-1} = \rho c/R_r$ connected to the S_0-section (fig 4a). This tube only transmits a wave travelling away from the lips, representing the loss of energy in R_r. In practice $S_{-1} \gg S_0$. Between these sections a reflection coefficients μ_0 is defined analogous to formula (3) :

$$(18) \qquad \mu_0 := \frac{S_{-1} - S_0}{S_{-1} + S_0} \quad ,$$

Note that $\mu_0 > 0$. Formula (4) for the calculation of the transfer function applies also to the model with both resistive glottis and radiation load. The index m has to be taken from 0 to N and the initial condition becomes $A_{-1}(z) = 1$.

In Appendix II it is shown that N-tubes of this kind with oppositely ordered or oppositely signed reflection coefficients have identical volume velocity transfer functions (within a constant). So there are four acoustically equivalent tubes, which are characterized by :

a. $(\mu_0, \mu_1 \ldots, \mu_N)$,

b. $(\mu_0', \mu_1', \ldots, \mu_N') = (\mu_N, \ldots, \mu_1, \mu_0)$,

c. $(\mu_0'', \mu_1'', \ldots, \mu_N'') = (-\mu_0, -\mu_1, \ldots, -\mu_N)$,

d. $(\mu_0''', \mu_1''', \ldots, \mu_N''') = (-\mu_N, \ldots, -\mu_1, -\mu_0)$.

As the coefficients with indices 0 and N have to be greater than zero (see below form. (3) and (18)), the tubes c and d are not interesting as vocal tract models. The change of sign makes these coefficients negative.

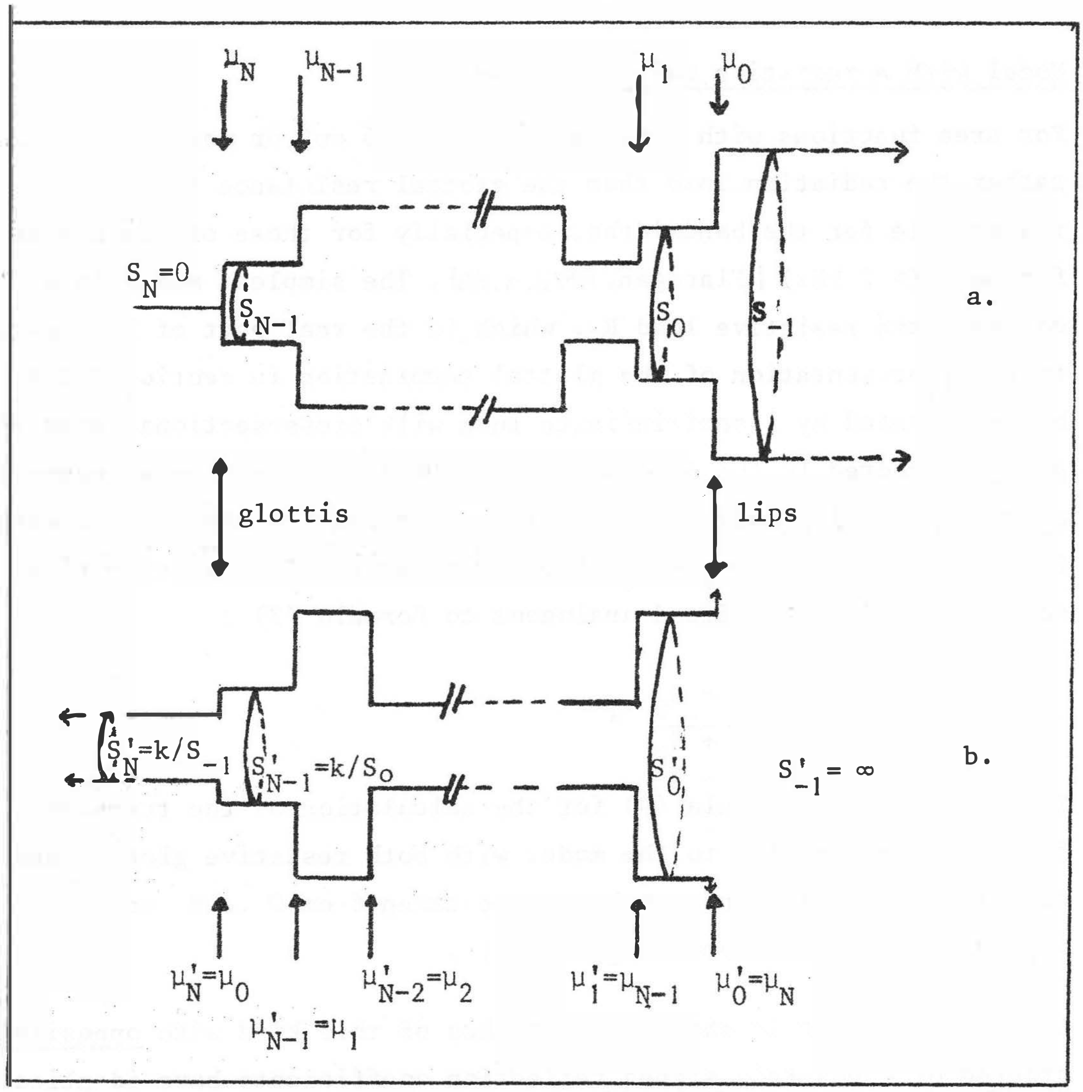

fig 4. Two N-tubes with oppositely ordered reflection coefficients.

Let us consider the tubes a and b with $\mu_N=1$ (fig 4a and 4b). Tube a with areas $S_{-1}, S_0, .., S_N$ is a 'resistive radiation load model' with no glottal losses ($S_N=0$). Tube b is a 'resistive glottis model' with no radiational losses, as $\mu'_0=1$. The areas of tube b are the oppositely ordered reciprocals of those of tube a : $(S'_{-1}, S'_0, .., S'_N) = (k/S_N, .., k/S_0, k/S_{-1})$. Or $S'(x) = k/S(L-x)$. k is an arbitrary constant of dimension $[\text{length}]^4$. This can be verified by formula (3), the definition of μ_i. Model a is preferred by Wakita [1973] because of better results. Model b has been used by Atal and Hanauer [1971]. The models are identical in transfer function and reflection coefficients save for their order. This accounts for and is in agreement with the comments of Strube [1977,p.236] on the contradictory results af Atal and Wakita. From the previous section we know that the area function and Z_g can be estimated from the transfer function if $\mu_N \neq 1$ in the resistive glottis model. When the resistive radiation load model is preferred we only need to place the reflection coefficients in the opposite order. In this case a unique solution of the area function is guaranteed if $\mu_0 \neq 1$ (or $R_r \neq 0$). Atal [1970] posed this also in his San Diego talk. In both cases the the total tube length and a reference area have to be known.

4.2 Model with both resistive glottis and radiation load.

Here the terminations at the glottal and mouth sides are represented by two tubes having characteristic impedances $Z_g \neq \infty$ and $R_r \neq 0$ respectively, as we have seen in section 4.1. From the symmetry property discussed in this section we know that the determination of the reflection coefficients from the transfer function will lead to at least two different sets of oppositely ordered coefficients. In order to get a unique solution of the area function knowledge of for instance μ_0 is needed, which is a function of S_0 and R_r. Even then uniqueness is not ensured, as it might happen that $\mu_0 = \mu_N$, and then the order of the coefficients is not known.

There is no direct method for the calculation of the area function of this model. Iterative procedures or other techniques have to be used. It is to be expected that the result of such a calculation will be in between those obtained when the two models of the previous sections are used.

4.3 More complicated models.

The radiation load Z_r is in fact complex, and a function of frequency [Flanagan, 1972, p.36]. For models which include Z_r and other details of the sound transmission through the vocal tract, it is very difficult or impossible to prove the one-to-one correspondence of area function and transfer function. They are too complicated. Even if it is proved for a certain model, it does not always make sense in practice. For, two clearly different area functions can have transfer functions that are so close together that they cannot be ditinguished experimentally [see Atal,1978 ; Strube,1977].

4.4 The lossless model.

We consider the case $\mu_0 = \mu_N = 1$ and $\mu_i \neq 1$ $(i=1,2,..,N-1)$. This corresponds to $R_r = 0$ and $Z_g = \infty$; i.e. the model is completely lossless. From section 3 we know that the coefficients $a_{N-1,i}$ are uniquely determined by $\mu_1, \mu_2, .., \mu_{N-1}$, because none of these reflection coefficients is ± 1. From $a_{N-1,i}$ we get the coefficients $a_{N,i}$ of the transfer function with (16). The number of independent prediction coefficients in A(z) is limited, because from (16) with $\mu_N = +1$ it follows that :

(20) $$a_{N,i} = a_{N,N-i} \quad , \; i=1,2,..,N-1 \;.$$

From (20) it follows that for odd N (N-1)/2 values of the parameter set $a_{N,i}$, (i=1,2,..,N) determine the whole set. For even N this number is N/2. Combination of these cases gives ENTIER(N/2)=:[N/2] independent parameters that determine the transfer function of a lossless N-tube. The number of independent parameters that determine the area function is twice as large, even if the length and a reference area are kept constant. Every lossless N-tube is a member of an ensemble with (N-1) - [N/2] = $[\frac{1}{2}(N-1)]$ degrees of freedom, whose elements all have the same frequency response, length and reference area. From the symmetry property of section 4.1 we already know that S(x) and S'(x) = k/S(L-x) belong to the same ensemble, as $\mu_0 = \mu_N$. These conclusions are the same as those of Bonder [1979], found by means of the so called N-tube formula.

The zeros z_i of A(z) are on the unit circle. For both z_i and $1/z_i$ are zeros :

(21) $$A(z_i) = 0 \quad \Rightarrow \quad z^N \cdot A(z_i) = A(1/z_i) = 0 \quad .$$

By (8) it is confirmed that the resonance bandwidths are zero.

SUMMARY OF RESULTS

Most items of this paper are well-known to phoneticians who work with the linear prediction model with energy losses in the glottal resistance Z_g and in the resistive radiation load R_r. We have verified the one-to-one correspondence of the parameter sets associated with the model, especially of the reflection coefficients with the prediction coefficients (see fig 2). This correspondence justifies the estimation of the area function of the N-tube from the formant frequencies and their corresponding bandwidths (or from the prediction coefficients). The essential conditions are :

1. Either $Z_g \neq \infty$ and $R_r = 0$ (no radiative losses) or $Z_g = 0$ and $R_r \neq 0$ (no glottal losses). An equivalent condition is : One of the two reflection coefficients μ_0 and μ_N is 1, the other one is smaller than 1.
2. A reference area S_{ref} is known. This might also be the area of the artificial section (representing R_r or Z_g).
3. The length L of the tube is known.
4. The tube areas are finite.
5. the first [N/2] formant frequencies and their corresponding bandwidths are known. For odd N also the $\frac{1}{2}(N+1)$-th bandwidth has to be known. An equivalent condition is that the N coefficients of the transfer function are known. These can be estimated from an N-th order linear prediction of the (preemphasized) speech wave.

The completely lossless N-tube ($Z_g = \infty, R_r = 0$). Every tube of this kind is a member of an ensemble with $[\frac{1}{2}(N-1)]$ degrees of freedom, whose elements all have the same formants, length and reference area. The bandwidths are zero. The transfer function is determined within a constant by [N/2] independent parameters.

APPENDIX I. On the general form of the transfer function of the N-tube model.

According to Flanagan [1972,p.26] the Laplace-transformed acoustic pressure p and volume velocity u at each side of an N-tube section of constant area S_i are related by (fig 5a) :

$$\text{(A1)} \qquad \begin{pmatrix} u_i(s) \\ p_i(s) \end{pmatrix} = \begin{pmatrix} \cosh(s\ell/c) & -\frac{S_i}{\rho c}\cdot\sinh(s\ell/c) \\ -\frac{\rho c}{S_i}\cdot\sinh(s\ell/c) & \cosh(s\ell/c) \end{pmatrix} \begin{pmatrix} u_{i-1}(s) \\ p_{i-1}(s) \end{pmatrix}$$

where i denotes the section number, counted from the lips, i=0,1,..,N-1, ℓ is the section length and c is the velocity of sound. Actually Flanagan's model is more complicated. It deals with internal losses in the tube. In the linear prediction model these losses are ignored. (A1) is a simplified version of Flanagan's formula. When the tube terminates in the glottal resistance Z_g and in the constant resistive radiation load R_r the boundary conditions become (see also fig 1) :

$$\text{(A2)} \qquad u_g(s) = p_g(s)\, Z_g + u_{N-1}(s) \quad , \quad u_{-1}(s) = p_{-1}(s)/R_r \ .$$

$u_g(s)$ and $p_g(s)$ are the glottal volume velocity source and glottal pressure.

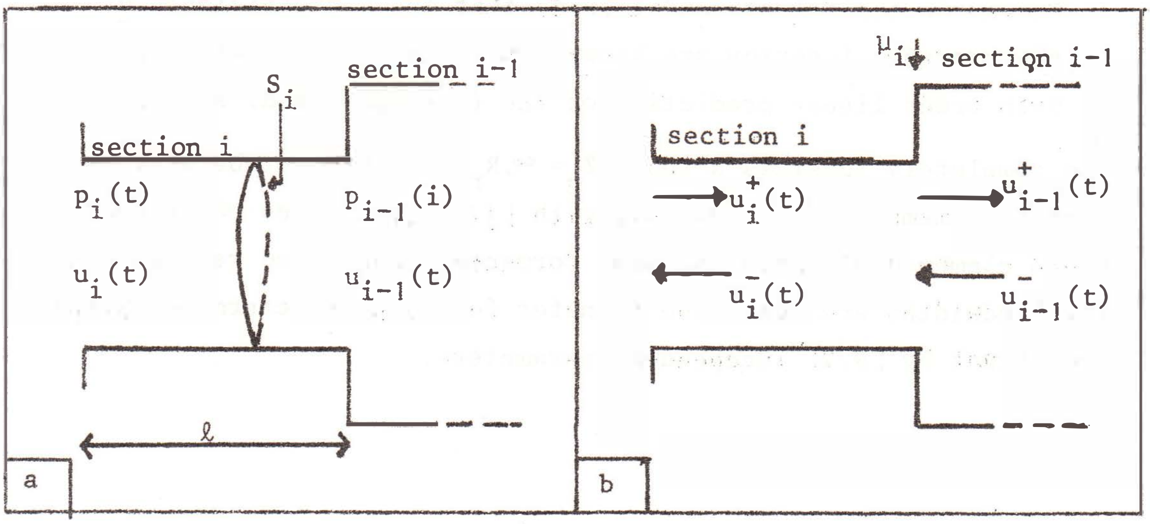

Fig 5. Two equivalent descriptions of the acoustical behaviour of a tube section.

We continue to describe the model in terms of plane volume velocity waves u^+ and u^-, travelling in the positive and negative x-directions respectively (fig 5b). Within a single section the waves are of constant amplitude. The actual volume velocity u_i is the difference of the two compounding volume velocities. The pressure p_i is the sum of the two pressure waves p^+ and p^-, which are proportional to u^+ and u^- respectively. The constant of proportionality is $\rho c/S_i$, the characteristic impedance of the section. Thus the transformation is :

(A3) $$u_i = u_i^+ + u_i^- \quad , \qquad p_i = \frac{\rho c}{S_i} \cdot (u_i^+ + u_i^-) \quad .$$

For the sake of readability the argument s is omitted. Putting (A3) into (A1) with $z = \exp(2\ell s/c)$, the definition of the hyperbolic functions and that of the reflection coefficients (3) gives :

(A4) $$\begin{vmatrix} u_i^+ \\ u_i^- \end{vmatrix} = \frac{z^{\frac{1}{2}}}{1+\mu_i} \begin{pmatrix} 1 & -\mu_i \\ -\mu_i z^{-1} & z^{-1} \end{pmatrix} \begin{pmatrix} u_{i-1}^+ \\ u_{i-1}^- \end{pmatrix}$$

The termination at the glottal and lip sides is represented by tubes of area $S_N = \rho c/Z_g$ and $S_{-1} = \rho c/R_r$ respectively (fig 6).

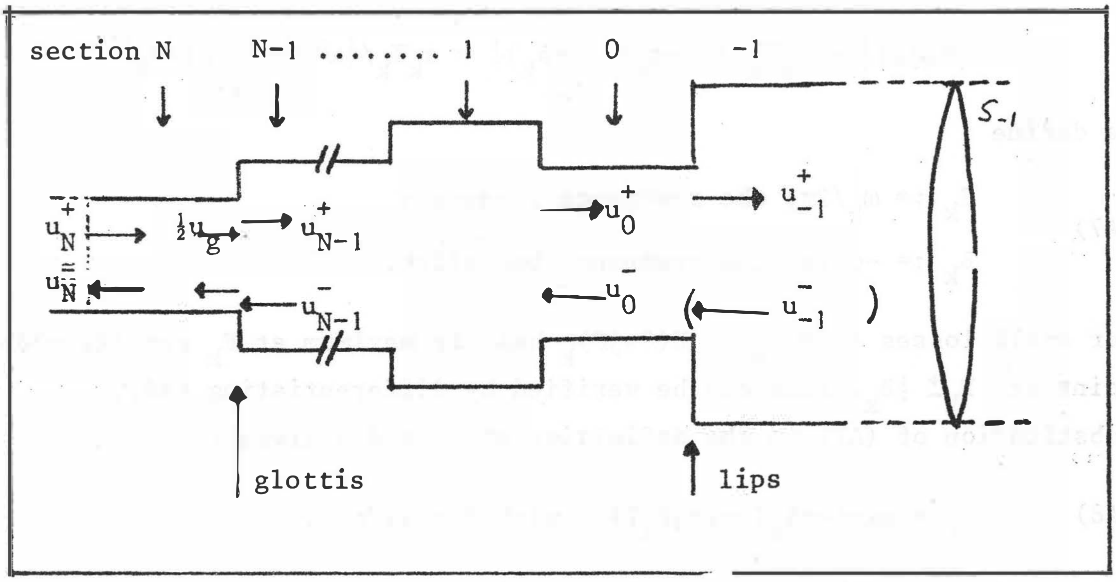

Fig 6. The whole model in travelling wave notation, glottal and radiative termination included.

Execution of the N matrix multiplications gives :

$$\text{(A5)}\qquad \begin{pmatrix} u_N^+ \\ u_N^- \end{pmatrix} = \prod_{i=0}^{N} \frac{z^{\frac{1}{2}}}{1+\mu_i} \begin{pmatrix} 1 & -\mu_i \\ -\mu_i z^{-1} & z^{-1} \end{pmatrix} \begin{pmatrix} u_{-1}^+ \\ u_{-1}^- \end{pmatrix} = \frac{z^{\frac{1}{2}(N+1)}}{\prod_i (1+\mu_i)} \begin{pmatrix} A(z) & B(z) \\ C(z) & D(z) \end{pmatrix} \begin{pmatrix} u_{-1}^+ \\ u_{-1}^- \end{pmatrix} .$$

As there is no backward travelling wave in the S_{-1}-section the volume velocity at the lips is that of the forward travelling wave u_{-1}^+. A and B are polynomials in z^{-1} of degree N, C and D are polynomials in z^{-1} of degree N+1. As we know from section 2.2 the glottal volume velocity u_g is twice the forward travelling wave entering the tube : $u_g = 2z^{\frac{1}{2}}u_N^+$. The factor $z^{\frac{1}{2}} = \exp(\ell s/c)$ results from the time difference of u_N^+ and u_g. As there is no source at the front end u_{-1}^- is set equal to zero and the volume velocity at the lips is u_{-1}^+. the transfer function can now be written as :

$$\text{(A6)}\qquad \frac{u_{-1}^+}{u_g} = \tfrac{1}{2}z^{-\frac{1}{2}} \cdot \frac{u_{-1}^+}{u_N^+} = \frac{2z^{-\frac{1}{2}N} \cdot \prod_i (1+\mu_i)}{A(z)} = \text{const.}\ \frac{z^{-\frac{1}{2}N}}{A(z)} .$$

This is formula (2) in section 2.1.

In order to determine the resonances of the system the zeros s_k of H(s) are written as $s_k = \sigma_k + j\omega_k$. Every vibrational mode k has an amplitude characteristic according to :

$$|H(j\omega)| = s_k\bar{s}_k / |(j\omega - s_k)(j\omega + \bar{s}_k)| = s_k\bar{s}_k / |\omega^2 + \sigma^2 - 2j\omega\sigma_k|$$

We define

$$\text{(A7)}\qquad \begin{aligned} F_k &:= \omega_k/2\pi, \text{ the resonance frequency} \\ B_k &:= -\sigma_k/\pi, \text{ the resonance bandwidth.} \end{aligned}$$

For small losses ($\omega_k \gg \sigma_k$) $H(2\pi jF)_k$ has its maximum at F_k and its -3dB-point at $F_k \pm \frac{1}{2}B_k$. This can be verified by differentiating (A6). Substitution of (A7) in the definition of s_k and z gives :

$$\text{(A8)}\qquad z_i = \exp(-\pi B_k T + 2\pi j F_k T) \text{ , with } T = 2\ell/c \quad .$$

This is formula (8) in section 2.3. z_i is one of the zeros of A(z).

APPENDIX II. N-tubes with equivalent transfer functions.

The next three theorems are valid for N-tubes as described by form. (A5) and (A6) in Appendix I. These tubes have a glottal reflection coefficient μ_N and a radiative one, μ_0.

THEOREM 1. Two N-tubes with oppositely signed reflection coefficients have within a constant identical transfer functions.

Proof. Let the two tubes be characterized by the reflection coefficients $(\mu_0,\mu_1,..,\mu_N)$ and $(\mu_0',\mu_1',..,\mu_N') = (-\mu_0,-\mu_1,.., \mu_N)$. The following general property of matrices is used :

$$\text{(A9)}\qquad \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} e & f \\ g & h \end{pmatrix}=\begin{pmatrix} i & j \\ k & l \end{pmatrix} \Rightarrow \begin{pmatrix} a & -b \\ -c & d \end{pmatrix}\begin{pmatrix} e & -f \\ -g & h \end{pmatrix}=\begin{pmatrix} i & -j \\ -k & l \end{pmatrix}$$

This property can easily be verified by execution of the multiplications. Applying (A9) to (A5) in App.I we get for the accented tube :

$$\text{(A10)}\qquad \begin{pmatrix} u_N^+ \\ u_N^- \end{pmatrix}' = \frac{z^{\frac{1}{2}(N+1)}}{\prod_i(1-\mu_i)} \begin{pmatrix} A(z) & -B(z) \\ -C(z) & D(z) \end{pmatrix}\begin{pmatrix} u_{-1}^+ \\ u_{-1}^- \end{pmatrix}'.$$

The resulting transfer function, obtained by putting $u_{-1}^{-\prime}=0$, is proportional to $1/A(z)$, i.e. proportional to the transfer function of the original tube (formula (A6)).

THEOREM 2. Two N-tubes with oppositely ordered areas have within a constant identical transfer functions.

Proof. Let the area functions be $(S_{-1},S_0,..,S_N)$ and $(S_{-1}',S_0',..S_N') = (S_N,..,S_0,S_{-1})$. Consider fig. 6. For reasons of symmetry it is allowed to change the roles of glottis and lips. We imagine the lips at the lefthand side, and u_N^+ is set equal to zero. The volume velocity at the lips is now $u_L= z^{-\frac{1}{2}}u_N^-$. $z^{-\frac{1}{2}}$ is the correction for the time difference between u_L and u_N^-. The glottal source is put at the righthand side : $u_g= 2u_{-1}^-$. The transfer function of this system can be found by inversion of the transfer matrix in (A5). The determinant of this matrix is the product of all partial matrices : $AD-BC = \prod_i(1-\mu_i^2)z^{-1}$. Thus :

(A11) $$\begin{pmatrix} u^{+}_{-1} \\ u^{-}_{-1} \end{pmatrix} = \frac{z^{\frac{1}{2}(N+1)}}{\prod_i (1-\mu_i)} \begin{pmatrix} D(z) & -B(z) \\ -C(z) & A(z) \end{pmatrix} \begin{pmatrix} 0 \\ u^{-}_{N} \end{pmatrix}$$

The transfer function is within a constant proportional to that of the original tube :

(A12) $$\frac{u_L}{u_g} = \tfrac{1}{2} z^{-\frac{1}{2}} \cdot \frac{u^{-}_{N}}{u^{-}_{-1}} = \frac{1}{\prod_i (1-\mu_i)} \cdot \frac{z^{\frac{1}{2}N}}{A(z)} \ .$$

q.e.d.

THEOREM 3. Two N-tubes with oppositely ordered reflection coefficients have identical transfer functions.

<u>proof</u>. This case is a combination of the former two theorems. Putting the reflection coefficients in the opposite order corresponds to putting the reciprocals of the areas in the opposite direction (see p.12). This is done in two steps :

1. Change of the order of the areas. According to theorem 2 the transfer function is not affected by this change.
2. Inversion of the areas. This corresponds to a change of the sign of the reflection coefficients (see formula (3)). According to theorem 1 the transfer function is again conserved.

The product of the constants of proportionality is 1.

q.e.d.

REFERENCES

Atal, B.S., Hanauer , L. (1971). Speech analysis and synthesis by linear prediction of the speech wave. J. Acoust. Soc. Am. 50,p.637.

Atal, B.S. (1970). Determination of the vocal tract shape directly from the speech wave. J. Acoust. Soc. Am., 47; 65 (A).

Atal, B.S., Chang, J.J., Mathews, M.V., Tukey, J.W. (1978). Inversion of articulatory-to-acoustic transformation in the vocal tract by a computer-sorting technique. J. Acoust. Soc. Am., 63; p. 1535-1555.

Bonder, L.J. (1979). The theory of the lossless N-tube. Inst. of Phonetic Sciences, Univ. of Amsterdam; Report Nr 60.

Brigham, O.E. (1974). The Fast Fourier Transform. Prentice Hall.

Flanagan, J.L. (1972). Speech analysis, synthesis and perception. Springer Verlag, Berlin-Heidelberg-New York. Second edition.

Markel,J.D.,Gray,A.H. jr (1976). Linear prediction of speech. Springer Verlag, Berlin-Heidelberg-New York.

Mol, H. (1970). Fundamentals of Phonetics, vol II : Acoustical models generating the formants of the vowel phonemes. Mouton & co. The Hague.

Strube, H.W. (1977). Can the area function of the human vocal tract be determined from the speech wave? Dynamic aspects of speech production. ed. by M. Sawashima and F.S Cooper, Univ. of Tokyo Press.

Wakita, H. (1973). Direct estimation of the vocal tract shape by inverse filtering of acoustic speech waveforms. IEEE Trans. AU-21, 417,-427.

ACKNOWLEDGEMENTS

We wish to thank J.S.C. van Dijk, E. de Boer and L.J. Bonder for reading the several versions of this paper critically, and H.V. Deighton-van Witsen and J.J.M. Schepers for correcting the English version.