SPEECH SYNTHESIS WITH A REFLECTION-TYPE LINE ANALOG

av Johan Liljencrants, tekn lic

Akademisk avhandling

som med tillstånd av Kungl Tekniska Högskolan i
Stockholm framlägges till offentlig granskning
för avläggande av teknisk doktorsexamen måndagen
den 26 augusti 1985 kl 10.00 i Kollegiesalen,
Administrationsbyggnaden, KTH, Valhallavägen 79,
Stockholm. Avhandlingen försvaras på engelska.

Stockholm 1985

Dissertation

which by permission from the Royal Institute of
Technology, Stockholm, will be submitted to public
examination for the degree of Doctor of Science,
on Monday 26, August 1985, 10.00 hours, in the
Staff Committee Hall, Administration building,
KTH, Valhallavägen 79, Stockholm. The dissertation
will be defended in English.

Stockholm 1985

ABSTRACT

With the Kelly-Lochbaum reflection-type line model, the vocal tract is
modelled as a uniaxial set of equally long line segments of different
impedance. The paper deals with practical aspects of its use for direct
speech synthesis in the time domain, including effects from time variation,
losses, and nonlinear phenomena. Existing models to account for time
variation in the segment areas are brought out as dynamic corrections to the
static scattering, and are extended for the analogy of flow partial waves. A
dichotomy between criteria of continuity in pressure and flow versus conti-
nuity in force and velocity is discussed. The dynamic scattering models of
both kinds are approximated with a simple method of adjusting partial waves
locally. The function of the models is illuminated in a number of dynamic
tests, one working from a concept of resonance mode isolation. Both types of
dynamic models appear to be useful, but for different purposes. The first is
relatively insensitive to undersampling of area data while the second is
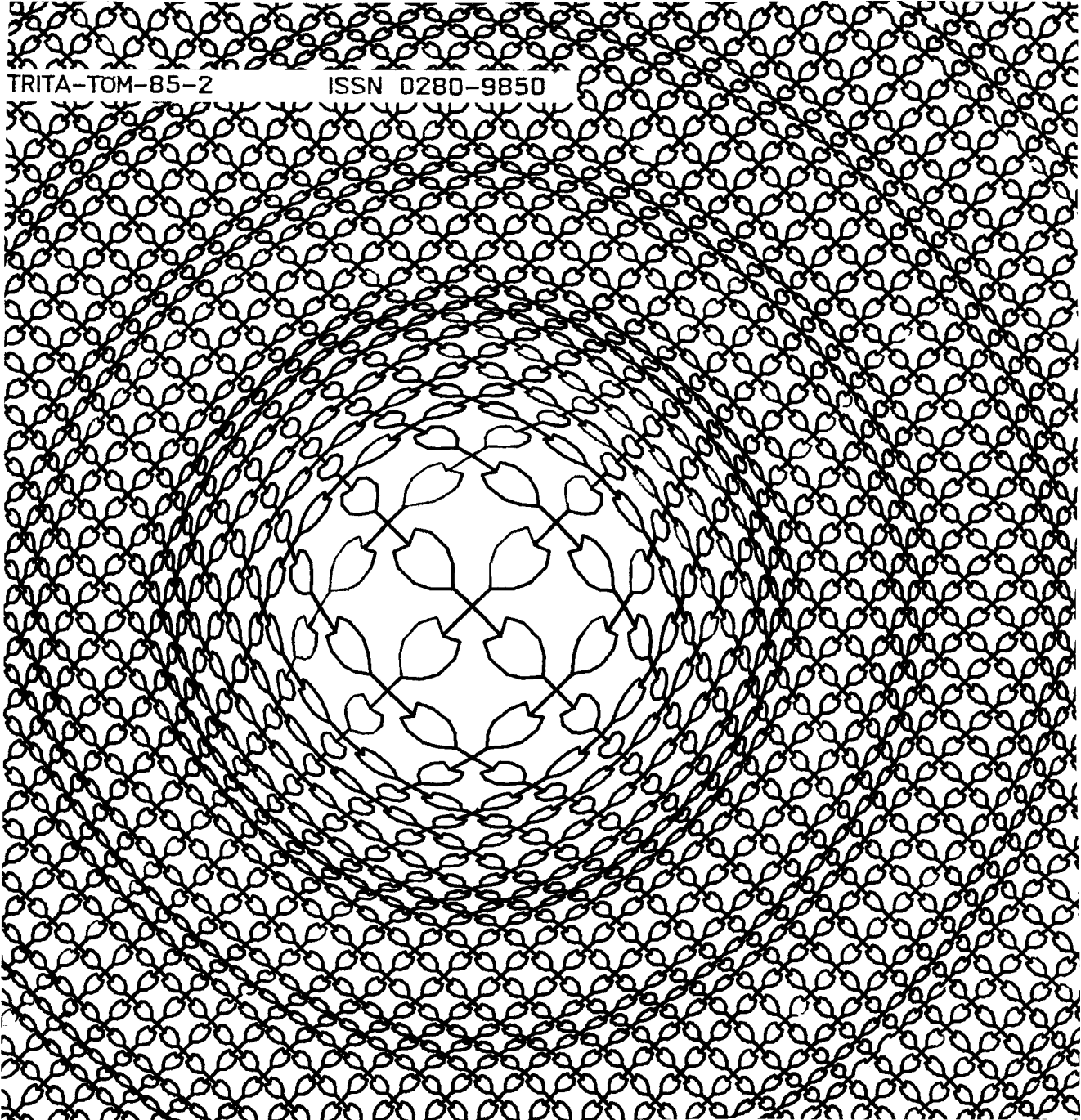closer to physical reality.

The many loss phenomena in the vocal tract are reviewed and displayed in
several forms. They are formalized in a consistent way to make it easy to
include them in the reflection model. Series and shunt losses in the line
model are expressed as separate and additive corrections to the scattering
equations. Special attention is given to the "jet loss," related to the
kinetic pressure drop. This is optionally handled as a variable generator or
as a loss resistance, and is used as a central element in simulating the
glottis as an integral part of the line model. Frequency-dependent losses,
including those into the vocal tract wall, are treated using z transforms.
Also incorporated is a z domain model of lip radiation impedance. Generation
of noise at constrictions is theoretically reviewed. With support from a
nozzle-blowing experiment, a practical procedure is elaborated. Programs
implementing the features are briefly described in context with a system of
service programs.

The general shape of the vocal tract area function may be modelled in terms
of a few Fourier coefficients. This was found in an earlier paper, appendix
to the dissertation, to be an efficient means of describing the mid-sagittal
tongue profile. (Author)

SPEECH SYNTHESIS WITH A REFLECTION-TYPE LINE ANALOG

Johan Liljencrants

Speech Communication
  and Music Acoustics
Royal Institute of Technology
S-100 44 STOCKHOLM SWEDEN

SCIENCE
AND
ART

CONTENTS

0.1 SUMMARY

The simple Kelly-Lochbaum (1962) line model is illuminated and augmented with a number of tricks to render a comprehensive research tool for simulation studies of the vocal tract. Several physical phenomena relevant to speech synthesis are reviewed. They are formalized in a homogeneous manner in order to make it easy to apply the results to the reflection line analog. When possible those tricks have been put in the form of corrections to the standard scattering equations. This approach makes it particularly easy to connect and disonnect features for research studies.

Existing models to account for time variation in the areas by Maeda (1977) and Strube (1982) are brought out as dynamic corrections to the static scattering, and also extended for the analogy of flow partial waves. A dichotomy between criteria of continuity in pressure and flow versus continuity in force and velocity is discussed. The dynamic scattering models of both kinds are approximated with a novel simple method of adjusting partial waves locally. The function of the models is illuminated in a number of dynamic tests. Both types of dynamic models appear to be useful, but for different purposes. The first is relatively insensitive to undersampling of area data while the second is closer to physical truth.

Methods of connecting parallel branches to the line are outlined. This makes it possible to implement nasals and their complex patterns of pole-zero pairs in addition to the all-pole characteristics of the single line.

Series and shunt losses in the line model are put as separate and additive corrections to the scattering equations, and a similar treatment is shown for generators. Also the non-linear Bernoulli effect is handled this way.

The many loss phenomena in the vocal tract are reviewed and displayed in several forms. They are formalized in a consistent way to make it easy to include them in the reflection model, and some quantitative data is compiled. Frequency dependent losses, including those into the vocal tract wall, are treated using z transforms. Also the z domain model of lip radiation impedance of Laine (1982) is incorporated. Special attention is given to the 'jet loss', related to the kinetic pressure drop. This is used as a central element in simulating glottis as an integral part of the line model.

Automatic generation of noise at constrictions is theoretically reviewed. With added support from a nozzle blowing experiment a practical procedure is defined in detail.

Programs implementing the features are briefly described in context with a surrounding system of service programs.

0.2 INTRODUCTION

In modelling the speech production process within the vocal tract we can already from its early days distinguish the uses of line analogs and terminal analogs. The line analogs then form models that represent direct spatial mappings of the speech organs. The line acts as a pseudo-stationary filter that operates on a source signal fed to its input. From circuit theory we can determine some suitable properties of that filter as seen from its terminals. The interest is then centered on resonance frequencies and bandwidths that define the speech formants. Here we recognize the class of terminal analogs. These model the behaviour as seen from the outer world, but their internal structure and operation is only indirectly related to the vocal tract.

For vowel production one can get very far with the terminal analogs since the natural process mainly offers a one-dimensional wave propagation problem. With consonants the modelling process is much more complicated. One reason is that in many sounds like nasals we have to account for partly parallel paths of wave propagation. Another is that the sound source may be located at several different places, as with fricative and plosive sounds. Both involve that the transfer function will have zeroes as well as poles. All these things can be handled in terminal analogs as shown by Fant (1960). In practical synthesis work it is however rather intricate to find lucid strategies to fabricate consonants, and especially the important transitions to neighbouring sounds. One consequence has been that the theoretically well founded cascade formant synthesizer is usually augmented with parallel branches. The approach with all formant resonators in parallel is therefore also a useful alternative, especially to produce good synthetic speech, replicated from an original.

At all times the line analog has been there in the background to supply necessary theoretical and practical data, but it has not until more recently been much used for direct synthesis, mainly because of its higher cost in computations. The inherent merit of it is the direct relation to the vocal tract which reflects in generally fewer and more understandable control parameters, but the balancing drawback is that these articulatory data are hard to measure, very few of them can be reliably found from the speech wave only.

From available types of simulated lines this paper will deal with the so called reflection type, working in the time domain. The phenomenon of wave reflections is since generations familiar to radio and radar engineers, using the Smith chart to establish standing wave ratios and reflection coefficients. Kelly and Lochbaum (1962) applied this concept in making a vocal tract analog from a number of line segments having equal length but different impedances. In the current search for better vocal tract analogs this so-called 'K-L' model has attracted continued interest, for instance by Maeda (1977), Degryse (1981), Strube (1982), and Titze (1983).

Neighbour disciplines covering many features of the K-L model are linear prediction and wave filter theory.

Linear prediction theory (Markel and Gray (1976)) has become very popular because it offers a set of straightforward computational methods for direct transformation between the domains of time waveforms, frequency spectra, formant data, and, by analogy, pseudo area functions. This theory in its basic form assumes a minimum-phase system although progress is made toward generalizations in several respects. Connected with this is filter implementations using lattice structures based on reflection coefficients. These coefficients have a particularly simple relation to a pseudo area function. The basic lattice makes all-pole filters only. It can be augmented with taps to take care of zeroes as well, but then the direct connection to the geometry of the vocal tract is lost.

Wave digital filter theory (Fettweis (1971)) treats discrete circuit elements and lines working from the concept of wave frequency which includes the feature of forward and backward partial waves. Some topics of this paper have a correspondence in this theory. In wave digital filter theory line segments are for instance joined with 'adaptors' that handle the partial wave scattering. Also reactive elements can be implemented as input impedances of lines having real terminations at the far end.

0.3 THE REFLECTION-TYPE LINE ANALOG

The Kelly-Lochbaum model of the vocal tract can be seen as an uniaxial set of equally long tubes, but with different cross-sectional areas. The area changes at the tube joints define a set of reflection coefficients. Several more recent improvements of this model are reviewed and some novel tricks will be presented. The primary goal has been to develop a flexible research tool for simulations of details in the speech production, and a secondary to get hints on the construction of a practical high-quality speech synthesizer.

The K-L approach is very attractive because it combines conceptual and computational simplicity with a very close relation to the physical world. It is of great interest that it is so easy to inspect the simulated flows and pressures at will at any place and time. But the model also has a number of weak points to be examined and cured.

First, in the reflection model the whole concept of frequency just is not there, it has a pure time domain approach and it applies to an ideal line. Of course a composite model with several segments does have frequency properties, but these are not lucid from the derivation proper. Frequency dependent elements can also be added using wave filter methods. A terminologically slightly different approach that will be used here is to interface elements formulated in the z domain as components in the

reflection coefficients.

Secondly, it does not properly cover losses in the vocal tract, which among other things determine the resultant formant bandwidths. In the original K-L approach losses were introduced as small reductions of partial wave amplitudes as the waves went through the tube sections. Additional discrete losses were added at the ends of the line to model energy absorption in the glottal and radiation impedances. Much of this paper will deal with ways to step by step account for the many loss mechanisms of the vocal tract. Some of those are crucial to make the model work at all with narrow constrictions and closures.

In the third place the transmission line analogy as adopted from electrical circuit theory does not inherently cover some inertial phenomena that are of importance in the acoustical line. The moving air (as opposed to electricity) has a nonzero density that will show inertial effects different from those of an inductance. A well known consequence is the Bernoulli effect: when air is accelerated in a constricted passage static pressure power is converted into kinetic power which at instances may create local downstream static underpressures. The supply pressure is then proportional to the particle velocity squared. One topic here is how to account for this. It is done with two different approaches, one with a direct view on the Bernoulli equation, and one more indirect treating the kinetic pressure drop as coming in a loss resistor.

Fourth, the model applies to a stationary line. In speech synthesis using terminal analogs the formant resonators are derived from stationary theory. This fact will usually not cause any serious problems. The spurious signals generated by parameter changes can be minimized to acceptable levels by use of proper design and preemphasis. This capitalizes on the resonators making up a filter that concentrates the spurious signals to the same frequency band as the wanted signals. Experience shows the situation to be more difficult with a line analog. The goal of this study is a wideband analog for both vowel and fricative synthesis in the same line. If such a line is excited with a lowpass filtered source signal and several formants are moved by gradually changing the shape of the line, then the high frequency band will easily be filled with intolerable spurious signals. These now stand out because they are not filtered away by the line and they are not masked by the wanted signal. It becomes a problem when areas are updated at a lower rate than the signal sampling frequency, something very desirable in order to economize computations. It is found that this problem is eased by a dynamic model given by Maeda (1977). Another interesting aspect of a dynamic model is its capacity of parametric pumping, the excitation of the wave system from temporal impedance changes in the line. This appears to be covered by a competing dynamic model by Strube (1982). These two dynamic models are reformalized and compared, and their correspondents in a flow wave analogy are shown, an alternative to the conventionally used pressure wave analogy. Also a simplistic approximation to these models is presen-

ted, a method of dynamically adjusting the partial waves. A number of tests on the various dynamic models are reported.

Finally it could be remarked that the model does not account for pumping in the pseudo DC sense, namely that an area change implies a volume change. The displaced amount of air will then invoke additional flows in the system. This can be accounted for in an elementary way by connection of external flow generators.

A guiding principle has been that added features should appear as corrections to the original expressions. The highlight is then that you can devise solutions to individual partial problems and in an easy way connect or disconnect them from the already existing construction. Admittedly this principle has its dangers, one may have to validate that the combination of two added features do not interact as to give undesired artifacts.

Having the fundamental arsenal of tools it is relatively straight forward to construct a complete vocal tract analog. Separate sections will deal with loss magnitudes, a lip radiation impedance model, glottal wave shaping to account for the source-filter interaction, and also automatic noise generation.

No explicit modelling of the mechanics of the vocal tract or glottis has been attempted here though it is directly feasible. The model is instead supplied with forcing area functions.

Table 0.1 CONSTANTS AND VARIABLES

All quantities are in terms of SI units: kg, m, s, N, etc. Constants pertaining to air in the vocal tract recomputed from Flanagan (1965):

ρ	$= 1.14 \ kg/m^3 = N*s^2/m^4$	density at 37°C, 100% RH
c	$= 350 \ m/s$	speed of sound at 37°C, 100% RH
μ	$= 1.86*10^{-5} \ Ns/m^2$	viscosity at 20°C, 100 $kPa = kN/m^2$
λ	$= 22.9*10^{-3} \ W/m*deg$	thermal conductivity at 0°C
C_p	$= 1000 \ \ \ \ J/kg*deg$	thermal capacitivity at 0°C, 100 kPa
η	$= 1.4$	adiabatic constant

Generally used variables:

l, d	length, diameter in m
A	area in m^2
P	pressure in $Pa = N/m^2$ (1 kPa = 9.87 $\approx$ 10 cm H_2O)
U	flow in m^3/s
v	particle velocity in m/s
τ	sampling interval in s
z^{-1}	z-transform unit delay operator
Z, R	acoustic impedance, resistance in Ns/m^5
k	reflection coefficient

Other variables are commented in the text.

Tab 0.1

1. BASIC TOOLS OF THE REFLECTION TYPE LINE ANALOG

The scattering equations for the lossless line. Pressure and flow analogies. Generalization into a dynamic line, existing models and a simplistic proposal. Dynamic tests on the models. Mode isolation. Practical considerations using the equations. Branching the line model. Introducing series and shunt losses as separate correction terms added to the lossless model. Accounting for frequency dependence of viscosity and heat conduction losses. Inclusion of a source in the line, and application to the kinetic pressure drop.

1.1 STATIC LINE MODEL

The reflection-type line analog described by Kelly and Lochbaum (1962), the K-L model, is a very simple-minded yet effective way to simulate wave propagation in the vocal tract. The tract model is set up as a sequence of N uniform tubes of equal length $l=c/2*\tau$, where τ is the sample interval, and different areas A_n. It stems from the general solution to the wave equation that says that the pressures and flows along the line can be put in terms of a forward going and a backward going wave system. Within each cylinder each partial wave propagates undisturbed, but when it reaches a joint it is split into two parts. One part goes on in the same direction into the next tube and the other part is reflected back into the old tube in the opposite direction.

To begin we look at the conventional formulation as shown in fig 1.1.
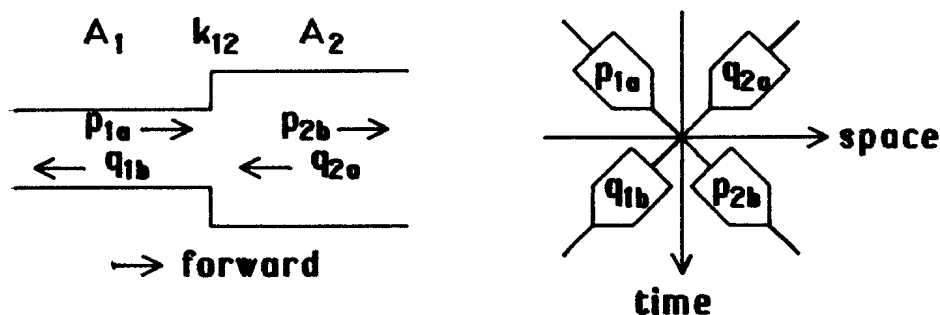
Fig 1.1. Tube joint with wave scattering. To the left physical arrangement, to the right space-time diagram of partial waves.

Here we see a specific joint number 1 where the two tubes number 1 and 2 of areas A_1 and A_2 meet. The subscripts 1 and 2 then denote the space dimension. The "forward" direction is defined as the direction of rising subscripts. A forward pressure wave p and a backward pressure wave q travel across the joint. In a sampled system these are normally regarded as impulse wave packages. Before they reach the joint they have the

Fig 1.1   BASIC TOOLS   1 - 1

magnitudes p_{1a} and q_{2a}. They meet at the joint and interact in some way, and the result of that interaction is the waves p_{2b} and q_{1b} leaving the joint. The time dimension is thus denoted by the a and b subscripts.

In classical literature the forward and backward waves are mostly denoted by the same symbol and superscripts '+' and '−' to show the direction. This tends to make the typography a bit involved. In the wave filter literature one goes to the other extreme and marks them by the subscripts only, 'a1', 'a2' etc, so there is no direct indication as to their direction. My notation here is hopefully a good compromise.

In each tube, the acoustic impedance, that is, the ratio of pressure P to flow U is, with n=1, 2

$$Z_n = \rho * c / A_n \qquad\qquad Ns/m^5 \qquad\qquad (1.11)$$

In the following I will frequently take the liberty to represent the tubes alternatingly by their areas or their impedances, whichever gives the simpler expressions.

At any time and place the total pressure P is the sum of the actual partial waves

$$P_n = p_n + q_n \qquad\qquad N/m^2 \qquad\qquad (1.12a)$$

Instead taking the difference between the partial waves, and excerting this pressure on the local line impedance will give the net flow through that tube section

$$U_n = (p_n - q_n) / Z_n \qquad\qquad m^3/s \qquad\qquad (1.12b)$$

The problem is now to find the "scattering equations", that is what are the resultant p_{2b} and q_{1b} expressed in the incident p_{1a}, q_{2a}, and the areas or impedances involved.

This is resolved from two continuity criteria. The first is that the total pressure in tube 1 must be the same as in tube 2, averaged over the infinitesimal time of interaction

$$P_1 = p_{1a} + q_{1b} = P_2 = p_{2b} + q_{2a} \qquad\qquad (1.13a)$$

The second continuity criterion is that the net flow also is the same in both sections

$$U_1 = (p_{1a} - q_{1b})/Z_1 = U_2 = (p_{2b} - q_{2a})/Z_2 \qquad\qquad (1.13b)$$

Using (1.11) this can also be written as

$$(p_{1a} - q_{1b}) * A_1 = (p_{2b} - q_{2a}) * A_2 \qquad\qquad (1.13c)$$

From (1.13a) and (1.13c) we solve the required quantities as

$$P_{2b} = (\quad 2*A_1 *p_{1a} + (A_2-A_1)*q_{2a} \quad) / (A_1+A_2) \qquad (1.14a)$$
$$q_{1b} = (\quad (A_1-A_2)*p_{1a} + \quad 2*A_2 *q_{2a} \quad) / (A_1+A_2)$$

Now a reflection coefficient is defined, equivalently in terms of areas or impedances as

$$k_{12} = \frac{A_1-A_2}{A_1+A_2} = \frac{Z_2-Z_1}{Z_2+Z_1} \qquad\qquad k_{12} = -1 \ldots +1 \qquad (1.15)$$

The reflection coefficient obviously contains a convention as to the direction. Also the different order of subscripts in the numerators using areas or impedances should be noted. The relations that give the resultant forward p_2 and backward q_1 are mostly put as

$$P_{2b} = (1+k_{12})*p_{1a} - \quad k_{12} *q_{2a} \qquad\qquad (1.16a)$$
$$q_{1b} = \quad k_{12} *p_{1a} + (1-k_{12})*q_{2a}$$

or, with some rearrangement to minimize computations, the "one multiplier lattice"

$$t_{12} = k_{12}*(p_{1a} - q_{2a})$$
$$P_{2b} = p_{1a} + t_{12} \qquad\qquad (1.16b)$$
$$q_{1b} = q_{2a} + t_{12}$$

These "scattering equations" is the result that is useful for direct simulation of the total behaviour of the waves in a line composed from several tubes. Starting such a procedure p_n and q_n are given for all the tubes at one instant of time. Working with equations (1.16) through all the joints we get an updated set of p_n and q_n valid for the next sampling time interval.

1.2 FLOW WAVES

It is not always recognized that one can make an exactly analogous derivation as above, but start with an alternate outset. Instead of the pressure waves p and q, let us use _flow_ waves, and call them r forward and s backward. To get the local net pressures and flows the correspondents to (1.12) will then be

$$P_n = (r_n + s_n) * Z_n \qquad \text{N/m}^2 \qquad (1.21)$$
$$U_n = r_n - s_n \qquad \text{m}^3/\text{s}$$

The continuity criteria (1.13a) and (1.13b) will then, with exactly the same meaning come out as

$$P_1 = (r_{1a} + s_{1b}) * Z_1 = P_2 = (r_{2b} + s_{2a}) * Z_2 \qquad (1.22)$$
$$U_1 = r_{1a} - s_{1b} \qquad = U_2 = r_{2b} - s_{2a}$$

Solving as above we get the _flow_ scattering equations

$$u_{12} = k_{12} * (r_{1a} + s_{2a})$$
$$r_{2b} = r_{1a} - u_{12} \qquad (1.23)$$
$$s_{1b} = s_{2a} + u_{12}$$

k_{12} as in (1.15) is identical with the one for the pressure wave analogy. Whether to use the pressure waves p, q or the flow waves r, s for simulation purposes is much an arbitrary choice. The only obvious difference in the scattering equations for the two cases is that of a few signs.

The flow wave representation may have an intuitive advantage as the waves can be visualized as puffs of air travelling along the line. Thus the reflections at the joint can be pictorially represented as in fig 1.2.
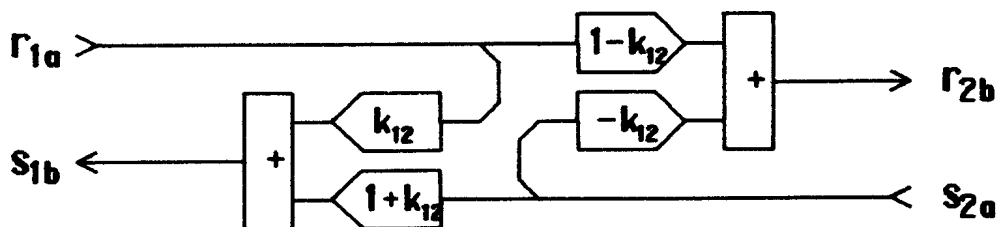

Fig 1.2. Flow wave splitting at a tube joint.

Fig 1.2   BASIC TOOLS 1 – 4

This diagram, oriented in the space domain, shows in a direct way the physical splitting and merging of the partial waves. The forward wave r_{1a} comes from the left in tube 1. When it reaches the joint it is split into two parts. A fraction $k_{12}*r_{1a}$ of it is reflected back to form a partial component of the resulting backward wave s_{1b}. The remainder of the forward wave is $(1-k_{12})*r_{1a}$ and goes on into tube 2 as a component in r_{2b}. Similarly the backward wave s_{2a} is split when it reaches the joint. The reflected portion of it will be the other component of r_{2b} and amounts to $-k_{12}*s_{2a}$. The negative sign is because of the opposite direction of the incident wave in relation to the direction k_{12} was defined. This simplistic reasoning is possible with the flow waves that involve transport of matter. In the case of pressure waves it <u>does not</u> hold, the waves then transport a state of matter rather than matter itself.

Alternatively the equations (1.23) inspire the equivalent flowchart of fig 1.3, now oriented in the time domain.
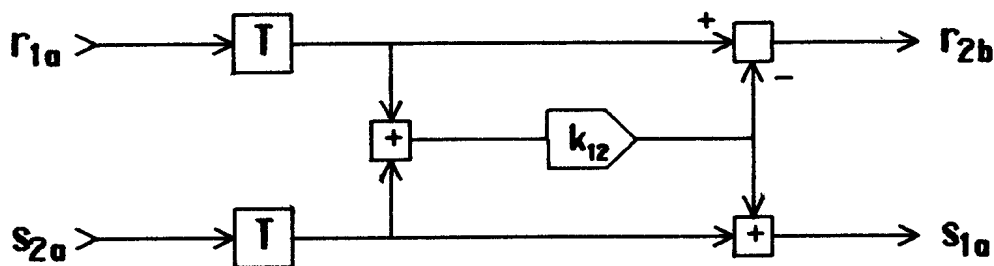
Fig 1.3. Flow diagram of scattering equations.

It may also be interesting to note that one could use partial waves pertaining to a mechanical analog rather than an acoustic, something I have not seen applied to vocal tract simulation. These waves are then velocity waves, let us call then v and w for the forward and backward directions, or <u>force</u> waves, f and g.

Let us put up the initial definitions for total pressure and net flow and compare them to our earlier correspondents (1.12) and (1.21) to have this quartet:

pressure		flow	
$P = p + q$	(1.12)	$P = (r + s) * \rho c/A$	(1.21)
$U = (p - q) * A/\rho c$		$U = r - s$	
velocity		force	
$P = (v + w) * \rho c$	(1.24)	$P = (f + g) / A$	(1.25)
$U = (v - w) * A$		$U = (f - g) / \rho c$	

Fig 1.3   BASIC TOOLS   1 - 5

For these velocity and force wave analogies we can again derive the scattering equations. We can however directly infer that the left pair will give identical scatterings as (1.16) because the tube specific area A is congruently placed in (1.12) and (1.24), while the constant ρc cancels out in the derivation. Similarly the right pair gives identical scattering results, like (1.23).

Thus, on one side the pressure and the velocity analogies have identical partial wave scatterings, and on the other, so have the flow and force analogies. The grouping may seem a paradox.

The choice between pressure or velocity analogy is then immaterial, as is the choice between flow or force, except for the scaling constant ρc. But conversely the choice between pressure or flow does have a significance for the operation of the model, and these two cases will be treated in parallel in the following.

One difference between the pressure and flow wave analogies that has an interest in hardware implementations is in the numerical magnitude of the partial waves. In speech the absolute value of total pressure P, the net flow U, and the area A all have a range between zero and some maximum value, maybe 2 kPa, 2 l/s, and 20 cm^2. Using (1.12) and (1.21) with P, U, and A/ρc normalized to their maxima we can solve for the corresponding partial wave amplitudes. Results for all combinations of extreme A, P, and U can be summarized in a 'truth table' like this:

A/ρc	P	U	pressure		flow analogy	
			2 p	2 q	2 r	2 s
1	1	1	2	0	2	0
1	1	0	1	1	1	1
1	0	1	1	−1	1	−1
1	0	0	0	0	0	0
0	1	1	∞	−∞	1	−1
0	1	0	1	1	0	0
0	0	1	∞	−∞	1	−1
0	0	0	0	0	0	0

The point to make is that with the flow analogy the partial waves always have moderate numerical values, but with the pressure analogy there is a risk of numerical overflow at the combination of a significant net flow with a very small area.

In fact, a point of major interest is how the models behave when the area of a section is taken down to zero (or usually, to avoid numerical trouble, a very small value) in order to simulate a closure. To obtain such a cutoff it is inevitable that the relative change in area from one sample interval to the next is very substantial. This can generate artifacts in the signals unless the model is dynamic, that is, in itself accounts for area change with time.

1.3 DYNAMIC LINE MODELS

The problem of developing a dynamic line analog has been addressed by Ruiz (1971), Maeda (1977), and Strube (1982). First the published Maeda and Strube models will be outlined, using the present notation, as in fig 1.1.

The dynamic aspect enters the following way. Simultaneously with the waves reaching the joint then also the impedances (areas) of the tubes change. Thus we have four impedances involved, one for each quadrant in the time-space diagram.

The continuity argument of Maeda is that there is some pressure P and flow U, both continuous at the crossover point. From this and the solution to the wave equation he finds

$$p_{1a} = (P + U*Z_{1a})/2$$
$$q_{2a} = (P - U*Z_{2a})/2$$
$$p_{2b} = (P + U*Z_{2b})/2$$
$$q_{1b} = (P - U*Z_{1b})/2$$

(1.31)

These are quoted here in terms of the impedances Z rather than the areas A which in the following will lead to simpler expressions of k. (1.31) are actually analogous to (1.13), but with the partial waves weighted in relation to the different impedances. Now, by the elimination of P and U and solving for the resulting waves the scattering equations will be

$$p_{2b} = p_{1a} + k_p*(p_{1a}-q_{2a})$$
$$q_{1b} = q_{2a} + k_q*(p_{1a}-q_{2a})$$

(1.32)

arranged here similarly to (1.16b). To get this analogy of form we must however use two different reflection coefficients k_p and k_q, one for each wave direction, where

$$k_p = (Z_{2b}-Z_{1a})/(Z_{2a}+Z_{1a})$$
$$k_q = (Z_{2a}-Z_{1b})/(Z_{2a}+Z_{1a})$$

(1.33)

Introducing the changes of impedance $\Delta Z_1 = Z_{1b}-Z_{1a}$ and $\Delta Z_2 = Z_{2b}-Z_{2a}$ will bring out the difference in Maeda's model compared to the static case, simply as

$$k_p = k_{12a} + \frac{\Delta Z_2}{Z_{1a}+Z_{2a}}$$

(1.34)

$$k_q = k_{12a} - \frac{\Delta Z_1}{Z_{1a}+Z_{2a}}$$

where k_{12a} is identical to the static reflection coefficient k_{12} of eq (1.15) using the impedances Z_{1a} and Z_{2a} in effect <u>before</u> the impedance change. With the notation selected here the expressions for k_p and k_q come out simpler than in the Maeda paper, without other approximations than the intrinsic one that Z is real and can be traded for A using (1.11).

Strube (1982) questions this result and argues on physical grounds that the time average of pressure waves p, q is continuous in space, and that the space average of longitudinal momentum p/Z, q/Z is continuous in time. This can then be put as

$$p_{1a} + q_{1b} = p_{2b} + q_{2a}$$
$$p_{1a}*A_{1a} + q_{2a}*A_{2a} = p_{2b}*A_{2b} + q_{1b}*A_{1b}$$

This again leads to scattering equations with two different reflection coefficients

$$p_{2b} = p_{1a} + k_1*p_{1a} + k_2*q_{2a} \qquad (1.35)$$
$$q_{1b} = q_{2a} + k_1*p_{1a} + k_2*q_{2a}$$

but now with

$$k_1 = \frac{A_{1a} - A_{2b}}{A_{1b} + A_{2b}} = k_{12b} - \frac{\Delta A_1}{A_{1b} + A_{2b}} \qquad (1.36)$$

$$-k_2 = \frac{A_{1b} - A_{2a}}{A_{1b} + A_{2b}} = k_{12b} - \frac{\Delta A_2}{A_{1b} + A_{2b}}$$

k_{12b} is the good old static reflection coefficient, but this time using the areas <u>after</u> the area change.

The reader is now referred to table 1.1 which contains a comprehensive list of alternatives for the scattering equations, static and dynamic. They are written in a form to make mutual comparisons easy, which to some extent implies that the form is not optimal for computer programming without further rearrangement.

As a parallel set the table also lists the corresponding equations for the flow wave analogy. The detailed derivations are somewhat tedious and are omitted since they can be inferred from the examples above.

Table 1.1. Summary of static and dynamic scattering equations for pressure waves p, q, and flow waves r, s. The equation pairs are written without abbreviations in a form to clearly show mutual differences between cases.

1, 2, n are consecutive subscripts in space, a, b in time.

STATIC REFLECTION COEFFICIENT

$$k_{12} = \frac{A_1-A_2}{A_1+A_2} = \frac{Z_2-Z_1}{Z_1+Z_2} \qquad Z_n = \frac{\rho c}{A_n} \qquad\qquad (1.15)$$

CHANGE WITH TIME IN AREA AND IMPEDANCE

$$\Delta A_n = A_{nb} - A_{na} \qquad\qquad \Delta Z_n = Z_{nb} - Z_{na} \qquad\qquad n = 1, 2$$

STATIC SCATTERING

PRESS
$$p_{2b} = p_{1a} + (p_{1a}-q_{2a})*k_{12}$$
$$q_{1b} = q_{2a} + (p_{1a}-q_{2a})*k_{12} \qquad\qquad (1.16)$$

FLOW
$$r_{2b} = r_{1a} - (r_{1a}+s_{2a})*k_{12}$$
$$s_{1b} = s_{2a} + (r_{1a}+s_{2a})*k_{12} \qquad\qquad (1.23)$$

DYNAMIC: Continuity in P and U

DYNAMIC SCATTERING (Maeda), time index a: early k, Z and A

PRESS
$$p_{2b} = p_{1a} + (p_{1a}-q_{2a})*k_{12a} + (p_{1a}-q_{2a})*\Delta Z_2/(Z_{1a}+Z_{2a})$$
$$q_{1b} = q_{2a} + (p_{1a}-q_{2a})*k_{12a} - (p_{1a}-q_{2a})*\Delta Z_1/(Z_{1a}+Z_{2a})$$

FLOW
$$r_{2b} = r_{1a} - (r_{1a}+s_{2a})*k_{12a} + (r_{1a}+s_{2a})*\Delta A_2/(A_{1a}+A_{2a})$$
$$s_{1b} = s_{2a} + (r_{1a}+s_{2a})*k_{12a} + (r_{1a}+s_{2a})*\Delta A_1/(A_{1a}+A_{2a})$$

SINGLE TUBE WAVE DYNAMIC CORRECTION

PRESS
$$p_b = p_a - (p_a-q_a)*\Delta A/2A_b$$
$$q_b = q_a + (p_a-q_a)*\Delta A/2A_b$$

FLOW
$$r_b = r_a - (r_a+s_a)*\Delta Z/2Z_b$$
$$s_b = s_a - (r_a+s_a)*\Delta Z/2Z_b$$

DYNAMIC: Continuity in PA and U/A, P/Z and UZ, or force and velocity

DYNAMIC SCATTERING (Strube), time index b: late k, A and Z

PRESS
$$p_{2b} = p_{1a} + (p_{1a}-q_{2a})*k_{12b} - (\Delta A_1*p_{1a}+\Delta A_2*q_{2a})/(A_{1b}+A_{2b})$$
$$q_{1b} = q_{2a} + (p_{1a}-q_{2a})*k_{12b} - (\Delta A_1*p_{1a}+\Delta A_2*q_{2a})/(A_{1b}+A_{2b})$$

FLOW
$$r_{2b} = r_{1a} - (r_{1a}+s_{2a})*k_{12b} - (\Delta Z_1*r_{1a}-\Delta Z_2*s_{2a})/(Z_{1b}+Z_{2b})$$
$$s_{1b} = s_{2a} + (r_{1a}+s_{2a})*k_{12b} + (\Delta Z_1*r_{1a}-\Delta Z_2*s_{2a})/(Z_{1b}+Z_{2b})$$

SINGLE TUBE WAVE DYNAMIC CORRECTION

PRESS
$$p_b = p_a - (p_a+q_a)*\Delta A/2A_b$$
$$q_b = q_a - (p_a+q_a)*\Delta A/2A_b$$

FLOW
$$r_b = r_a - (r_a-s_a)*\Delta Z/2Z_b$$
$$s_b = s_a + (r_a-s_a)*\Delta Z/2Z_b$$

Tab 1.1

1.4 ALTERNATE APPROACH TO DYNAMIC MODELLING

In practise one would like to supply new data on areas or reflection coefficients at a lower rate than the sampling frequency. Each time new area data is then supplied in presence of signals there will be disturbing transients in the partial waves. The following heuristic approach was developed as a means to get rid of these transients.

Now then, instead of considering the tube joints at the time of reflections, let us focus on a specific tube. Here the net pressure and flow can be inferred from (1.12) or (1.21), though these are strictly valid only just at the time of reflections. Let us suddenly change the area or impedance of that tube. The continuity requirement posed is that the pressure and flow remain unchanged in the section. We let some "Maxwell's demon" adjust the existing forward and backward waves in order to effect this continuity.

Since we are concerned with a single tube the space subscript is now omitted. The net pressure and flow before the area change are

pressure wave analogy | flow wave analogy

$$P_a = p_a + q_a$$
$$U_a = (p_a - q_a)/Z_a$$

$$P_a = (r_a + s_a)*Z_a \qquad (1.41)$$
$$U_a = r_a - s_a$$

After the area change the equations are all the same, except time subscript is b. Equating $P_a=P_b$ and $U_a=U_b$ then gives

$$p_a+q_a = p_b+q_b$$
$$(p_a-q_a)/Z_a = (p_b-q_b)/Z_b$$

$$(r_a+s_a)*Z_a = (r_b+s_b)*Z_b \qquad (1.42)$$
$$r_a-s_a = r_b-s_b$$

from which the adjusted waves can be solved as

$$p_b = p_a - (p_a-q_a)*\Delta A/2A_b$$
$$q_b = q_a + (p_a-q_a)*\Delta A/2A_b$$

$$r_b = r_a - (r_a+s_a)*\Delta Z/2Z_b \qquad (1.43)$$
$$s_b = s_a - (r_a+s_a)*\Delta Z/2Z_b$$

The procedure is then: whenever the area of a tube in the line analog is changed, then with (1.43) compute a new pair of forward and backward waves, for that tube only, from the existing waves and the area change. Also, before the scattering equations are excecuted of course the reflection coefficients at the ends of that tube have to be recomputed. For the scattering computations the static equations (1.16) or (1.23) are used.

Let me call this procedure: adjustment for P-U continuity.

This adjustment and the static scattering equations give a fairly complicated expression when brought together into a dynamic scattering equation. It is not clear that the procedure is computationally economic by

comparison to the other dynamic models, but the point of it was that it should not have to be used at every sample interval.

With this wave adjustment procedure the dynamic model is split up into two separate parts, one to take care of the area change, and one to take care of the scattering. This is a simplification when programming a computer. Waves are adjusted when new areas are supplied, and there is no need to save the old areas for the scattering computations eventually fortcoming.

With both partial procedures the underlying requisite was that total pressure P and net flow U were continuous in time and space. A close correspondent to this should be the dynamic scattering of Maeda that was derived from the same basic criteria. Tests following will show that the Maeda model and the F-U adjustment indeed give very similar results.

The next invention is to try a different wave adjustment, this time with a continuity requirement similar to the one posed by Strube, that of continuity in $P*A$ and U/A. (Not exactly like: we do not dispose of the space dimension inside a single tube). Properly multiplying and dividing with the area A_a in (1.41) gives us

pressure wave analogy | flow wave analogy

$$P_a*A_a = (p_a + q_a)*A_a \qquad\qquad P_a*A_a = (r_a + s_a)*\rho c \qquad\qquad (1.44)$$
$$U_a/A_a = (p_a - q_a)/\rho c \qquad\qquad U_a/A_a = (r_a - s_a)*Z_a/\rho c$$

Again, after the area change the equations are all the same, except time subscript is b. Equating $P_a*A_a=P_b*A_b$ and $U_a/A_a=U_b/A_b$ and solving for the adjusted waves having subscripts b then gives

$$p_b = p_a - (p_a+q_a)*\Delta A/2A_b \qquad\qquad r_b = r_a - (r_a-s_a)*\Delta Z/2Z_b \qquad\qquad (1.45)$$
$$q_b = q_a - (p_a+q_a)*\Delta A/2A_b \qquad\qquad s_b = s_a + (r_a-s_a)*\Delta Z/2Z_b$$

Let me for simplicity call this: adjustment for f-v continuity (force and velocity, regarding the dimensions of $P*A$ and U/A).

Shutter

Sharp squeeze

F-pattern

time ⟶
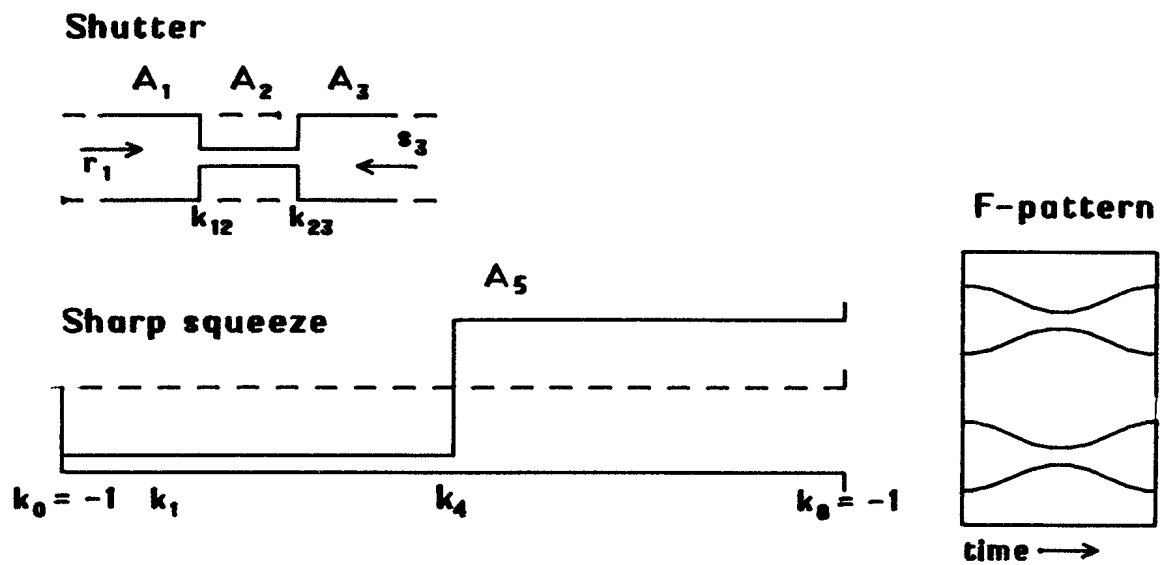
Fig 1.4. Layouts of model test situations.

a. Shutter test. Three tube sections, terminated without reflections and externally fed with pseudo DC signals. Area of middle tube is altered in steps.

b. Squeeze test. Eight section line terminated by open and short circuits is gradually perturbed into a narrow and a wide half, and back. The lossless line is pre-excited with a standing wave on one of its resonance frequencies. To the right resonance patterns in spectrogram style.

Flow, Amln .2

U
Stat
P

U
L-ts P-U
P

U
Maeda
P

U
L-ts f-v
P

U
Strube
P
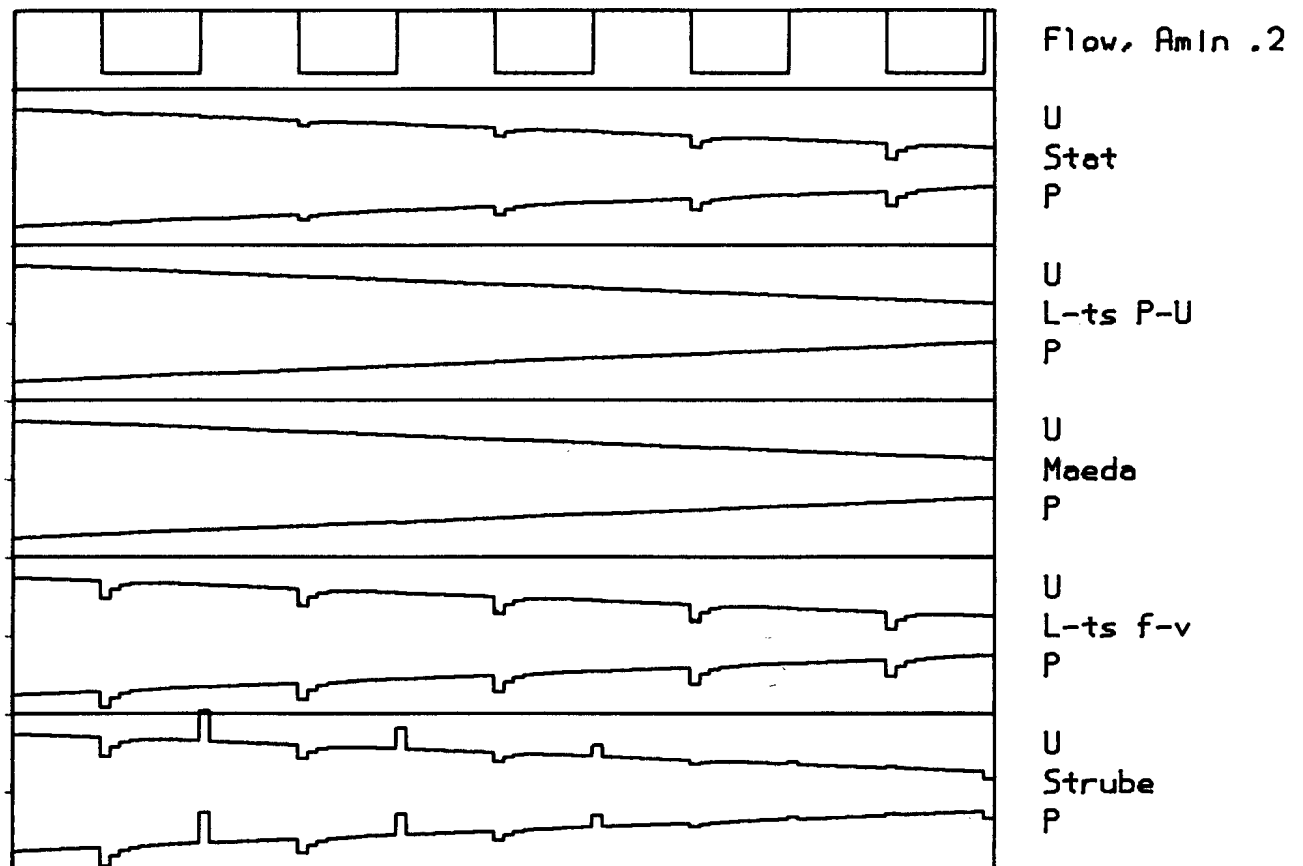
Fig 1.5. Output from shutter test of fig 1.4a. The area of the middle tube is varied stepwise as shown in the top trace. Then follow net flow and pressure in that tube for the static model, for the P-U adjustment, the Maeda model, the f-v adjustment, and the Strube model.

Fig 1.4, 1.5

1.5 TESTS ON THE SCATTERING EQUATIONS

We have now ten candidates for the scattering problem: pressure and flow analogy, and for each of those the static solution, two dynamic, and two approximately corresponding wave adjustment proposals. Needless to say they all give identical results with a line that is time invariant. But with a time varying line the models do behave differently. Also the apparent simplicity of the reflection model is indeed quite treacherous. I have tried the models in a number of different tests and will now describe three of those that give quite different results, this to illuminate the puzzling complexity of the choice. They put hard stress on the models as compared to normal speech synthesis situations, but hopefully accentuate the differences. The topic of model selection will also be revisited later in connection with simulation of the glottis.

1.51 SHUTTER TEST

The first is the 'shutter test' where a single tube section is stepwise varied and the signal is pseudo DC. The purpose is to show what happens when area is changed in large relative steps, as happens close to an area cutoff. A line made of three tubes as in fig 1.4a is simulated. The end tubes constantly have unit area and are externally excited. A constant forward wave is injected into tube 1 and a slowly varying ramp for the backward wave enters tube 3. The particular design of these inputs makes the test cover the range from zero static pressure and unit net flow up to zero flow and unit pressure. The waves leaving the end tubes are neglected, the simulation thus covers a reflection free termination at both ends of the three tube array.

Some results are shown as functions of time in figure 1.5 . The upper trace of the figure shows the area of the middle 'shutter' tube. To give clearly visible effects the shutter area is varied abruptly in large steps a number of times. The area variation is from unit area, as in the abutting end sections, to 0.2 of this. The total pressures and net flows in the same tube are derived from sums and differences as in eqs (1.21) and are shown for the static model, the Maeda and Strube models, and for the proposed wave adjustment models.

An initial remark is that for each dynamic model the resulting pressures and flows come out the same, either the pressure or the flow analogy is used, though the partial waves do not.

For the static model, when the area of the middle tube is suddenly made smaller, we see transients in the line. These transients gradually die out as the waves adapt to the new situation of areas and reflection coefficients. The time constant in the decay of the transients is actually inversely proportional to the area of tube 2. If this area is very small it will take a long time before the waves trapped inside it manage

to get out through the barriers of near total reflection at its ends. Similar transients come in the adjacent tubes (not shown). When the middle area is switched back to unit size another transient is generated in the neighbour tubes. This one lasts only one time slot, after that it disappears without reflections out through the ends of the tube array.

A specific observation in this test is that after the transients have died out the pressure and flow in the shutter are the same as in its neighbours, and also the same as they would have been without any area change. My wave adjustment scheme for P-U continuity was designed to give precisely this equilibrium situation immediately on the area change. The Maeda model, though formally different, in this particular test happens to give exactly the same result. The Strube model on the other hand gives even stronger transients than the static one, and so does the adjustment for f-v continuity (but not the same!). This does not disqualify these models as such, but it suggests them to be particularly sensitive to undersampling of the area parameter data.

The static flow wave model, the Maeda, and the continuous P-U model have the idiosyncrasy that when area is taken towards zero the net flow is not interrupted. This is the consequence of continuity of flow and is mathematically all right, a finite flow can well go though an infinite impedance. But this would make a practical system useless to model a cutoff. Later we will see that this disastrous misbehaviour can be helped once resistive losses are accounted for.

1.52 UNIFORM EXPANSION TEST

The next two tests are staged in an eight section system, terminated at both ends with a reflection coefficient of -1, that is, closed in one (the glottal) end, and open, acoustically short-circuit, in the other. Initially the line is set up with an odd number of quarter periods of equal forward and backward waves. We thus simulate a lossless resonator tube, pre-excited to operate at one, but only one of its four resonances.

Test number two is to let this pre-excited system have tubes of equal areas that uniformly expand exponentially with time. Since areas are equal all reflection coefficients are zero except at the ends. The initially set up waves will oscillate back and forth indefinitely in a standing wave pattern without attenuation. It is fairly easy to see how a physical line would behave in this situation. The energy per unit length in one of its segments is composed from the kinetic part $E_k = L*U^2/2$ and the static part $E_s = C*P^2/2$ and, since $L = \rho/A$ and $C = A/\rho c^2$ it is straightforward to find the changes dE/dA of kinetic and static energies with area. Putting both these changes to be zero, that is, assuming no energy interchange with the walls, and no change in standing wave pattern (same relation between static and kinetic energies) one immediately arrives at $dU/U = dA/2A$ and $dP/P = -dA/2A$. Ideally then with an exponentially increasing

area we would expect the flow to rise and the pressure to fall exponentially, but with half the exponent. This special result also conforms with the general theory of 'adiabatic' perturbations elaborated by Jospa (1977).

In the uniform expansion test all the dynamic models behave very closely like this but none identical to any other. I find this remarkable thinking of their difference in the other tests. For instance an area increase of 5% per sample interval, repeated 128 times (a total area growth more than 500 times) gives quite insignificant differences in amplitude and phase, to the order of a few percent.

Conversely, the static model does not work at all. In this nothing will happen to the partial waves because of the expansion. Thus in the pressure analogy the pressure is constant while the flow increases with the area, and in the flow analogy the flow is constant while the pressure is inversely proportional to area (cf eqs (1.12) and (1.21)).

1.53 SQUEEZE TEST

The third test could be called the 'squeeze test'. It uses the same lossless, pre-excited eight tube line. Initially all sections have unit area. During the test the area function is gradually perturbed to become narrow at the glottis end and wide at the mouth end, and then back to the original shape again. The four segments at the 'glottis' end are uniformly made narrower and the other four correspondingly wider so that the total volume is constant. Fig 1.4b shows the area function and corresponding formant trajectories in a schematic form. The area change with time is shaped as a biassed inverted cosine period over a time span of 256 samples. At maximal perturbation one would like to have the narrowest tube collapse to zero area. We cannot go all that way because the models break down from arithmetic reasons. Instead the minimum area at maximum perturbation was arbitrarily selected as 0.02.

In a milder form of this test, the 'smooth squeeze' the area was perturbed as eight points on a half cosine period. This gives other formant trajectories, but apart from this the results were similar.

The squeeze test illustrates a somewhat academic theorem not generally thought of in speech work. It applies to linear multi-mode resonant systems like strings, membranes, cavities, that are subject to slow, continuous changes of shape, and without external energy exchange, 'adiabatic changes'. It can be named the principle of mode isolation, a consequence of the resonance modes being orthogonal:

When a resonator, freely oscillating in a resonant mode, is slowly perturbed, the oscillating energy is confined to that same mode.
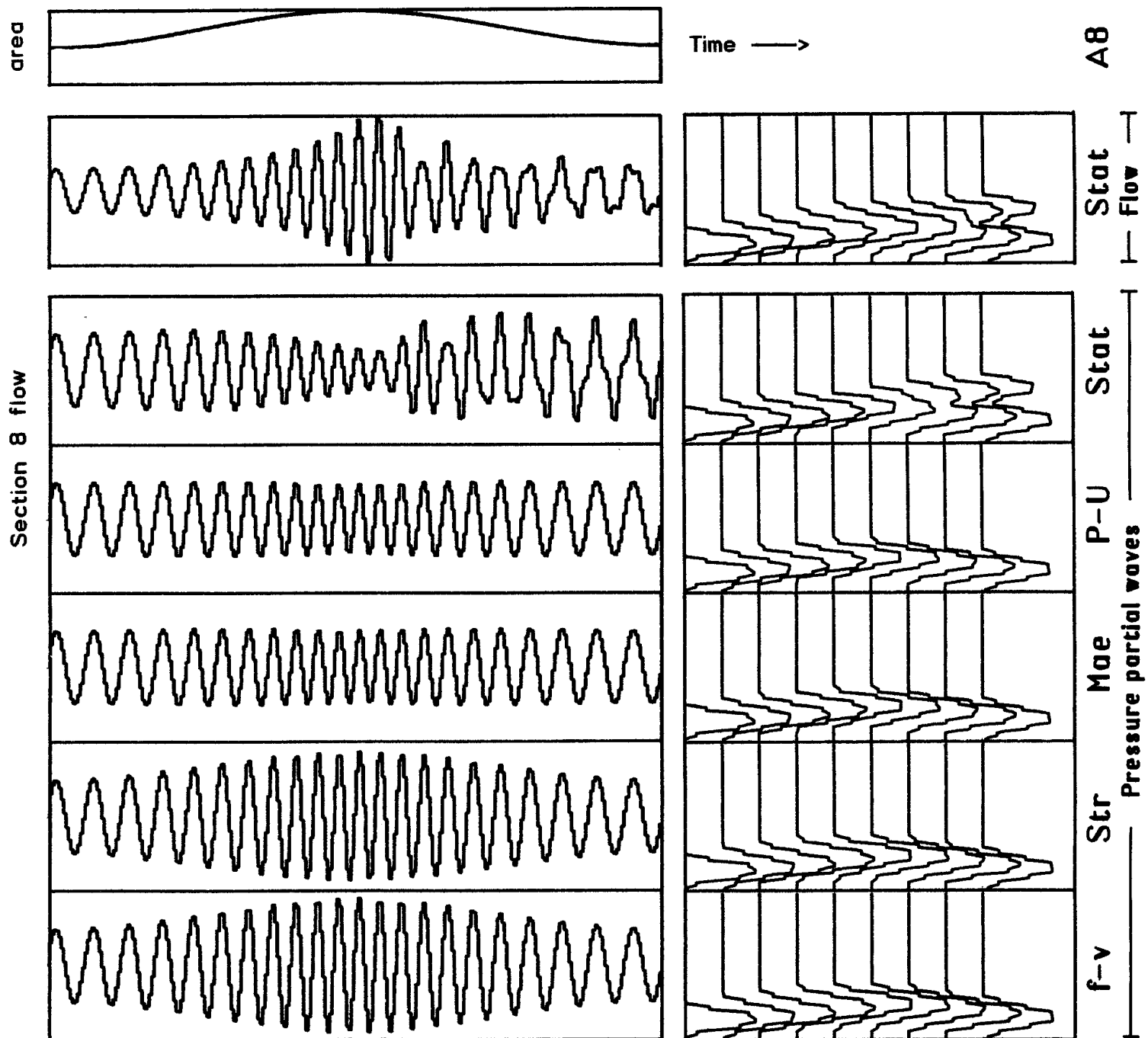
Fig 1.6. Outputs from squeeze tests of fig 1.4b.

a. Line was pre-excited with F1 only. Top trace is area of sections 4-8. Then follow net flows in section 8 as they develop in time for the different models as indicated. To the right are Fourier spectra of overlapping 48-sample intervals, using raised cosine (Hanning) window. Consecutive spectra are displaced 24 samples in time and 20 dB in level.

The static model gives very different waveforms for the flow and pressure partial wave analogies (two top waveforms), and in both cases energy is transferred to F2. The four dynamic models all keep the energy in F1, and each gives same waveform with either flow or pressure partial waves.
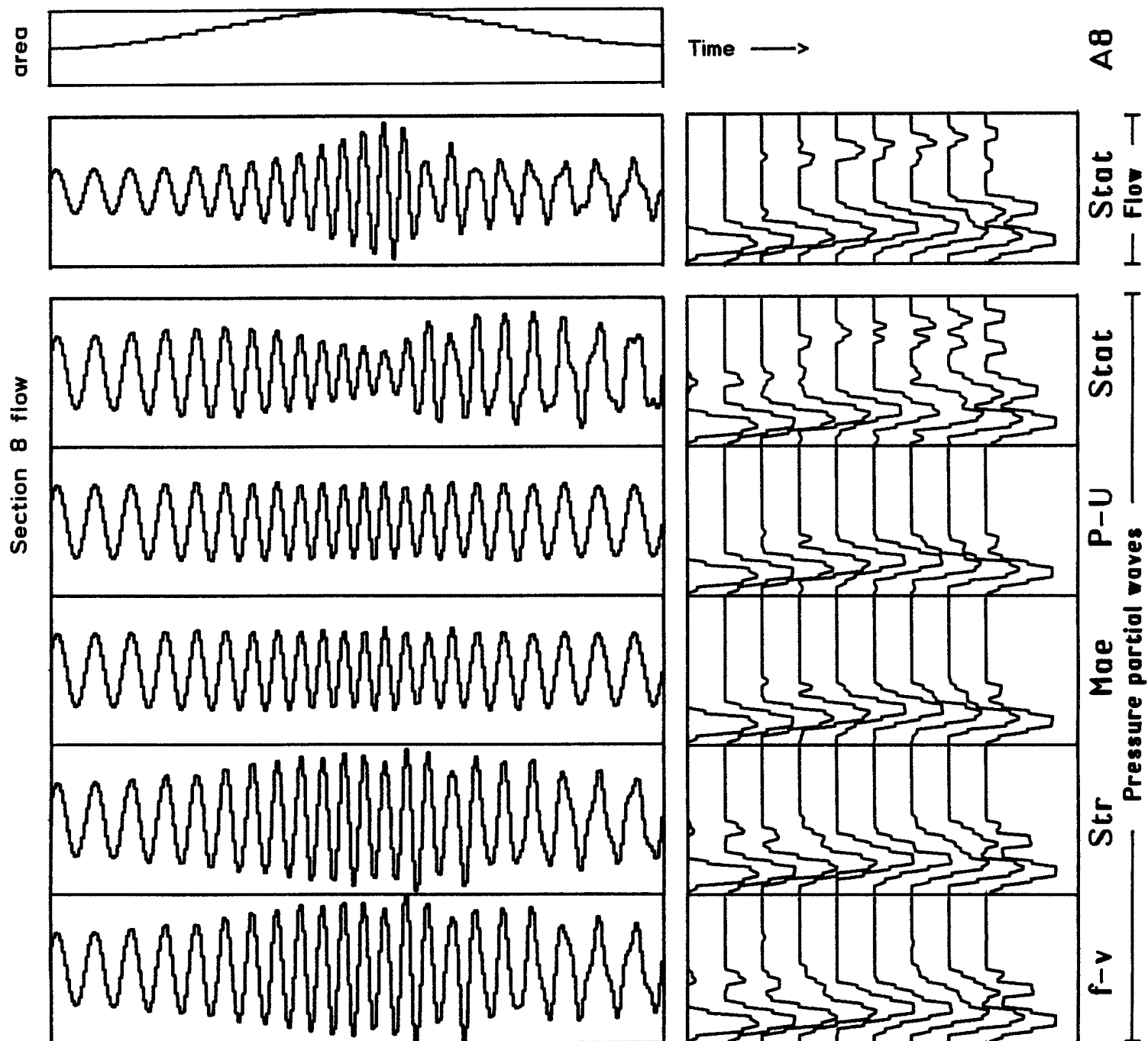
Fig 1.6a

Fig 1.6b. Same as fig 1.6a, but with area function undersampled by
a factor of 7 as seen from the notches in the area trace at top.

All models show energy leakage to higher formants, but the P-U
adjustment and the Maeda model generally less than the others.
Details vary considerably depending on the undersampling factor.

Fig 1.6b

The frequency of this mode may well change as a consequence of the perturbation, the point is that the oscillations of the mode considered do not excite other modes, lower or higher. For a mode to be isolated this way it is obviously required that the system is linear, and that perturbations are slow enough that they are insignificant within the period of one oscillation. Otherwise harmonics will be generated that will excite other modes as well. Conversely this principle could work as a test on physical accuracy in the models.

Fig 1.6 shows results from the 'squeeze' test on the static and all the dynamic models, for pressure wave and flow wave analogy. Time waveforms shown are the flows in tube 8, supplemented with successive Fourier spectra, weighted with a Hanning window. It is obvious from the spectra that all the dynamic models adhere to the mode isolation principle, and that the static model does not. Also the static model again differs greatly in the pressure and flow analogies, while the dynamic models are independent of type of analogy.

During the perturbation the first two formant frequencies approach until they differ only a small fraction of the total frequency range. Still the dynamic models manage to keep the neighbour modes 'uncontaminated' from the excited one. The corresponding holds also when the system is pre-excited to its higher modes. (In the limit of a total closure in the narrow half line the formants would have pairwise coincided, making mode isolation impossible.)

The physically reasonable situation at maximal perturbation would be that virtually all energy is squeezed out from the narrow half line into the wide one. Energy conservation then requires the mouth flow to be twice its unperturbed value, that is seen from $E{\sim}LU^2/2{=}\rho lU^2/2A$ (more precisely: integrated over the line and weighted with a similar sinusoidal amplitude function) where length l is halved and area A is doubled during the perturbation. The Strube model and the model of f-v continuity adjustment are the only ones to manage this aspect correctly. The Maeda model and the P-U adjustment give little change in the mouth flow amplitude (as opposed to the uniform expansion test), they generally are under-sensitive to the area perturbation. This appears to illustrate the importance of conserving longitudinal momentum to insure the proper longitudinal transport of energy.

Parenthetically I could mention a similar experiment that failed in mode isolation. It was done on a cascade of four second order formant resonators implemented as recursive z transform filters, a classical formant speech synthesizer, fig 1.7. The lossless resonators were artificially pre-excited for stationary oscillation on F1, and then the formant frequencies were perturbed. The output gets strong 'leakage' of signal into F2 either it is F1 or F2 that is perturbed. The reason is that the filters are derived from assumptions of a static system, so all kinds of

artifacts can and will occur when the filter coefficients are changed in presence of a signal. There is also an inherent asymmetry in the system since the signal goes in one direction only. The first resonator is isolated from what happens in the later circuits, but the opposite is not true.

This raises the philosophical conclusion that it is not feasible to develop a dynamic version of the classical cascaded formant resonator filter. It is additionally required that a feedback path is included, a feature inherent in wave filters, of which the line analog is an example.

But the cascade synthesizer has proven useful for so many years, then why so much fuss about dynamics with the line model? The main motives are:

In a formant syntesizer the spurious signals are filtered so that they are effectively masked by the wanted signal. But in a line model spurious signals can excite formants outside the wanted frequency band. The line is no lowpass filter unless explicitly engineered that way.

Relative formant frequency changes are almost always moderate. Specifically, formants never approach zero frequency, in practical synthesis work you would not let them do that, and in the real vocal tract the finite wall mass will prevent it from happening. But in a line analog relative area changes can <u>not</u> be moderate when you want to approach cutoff.
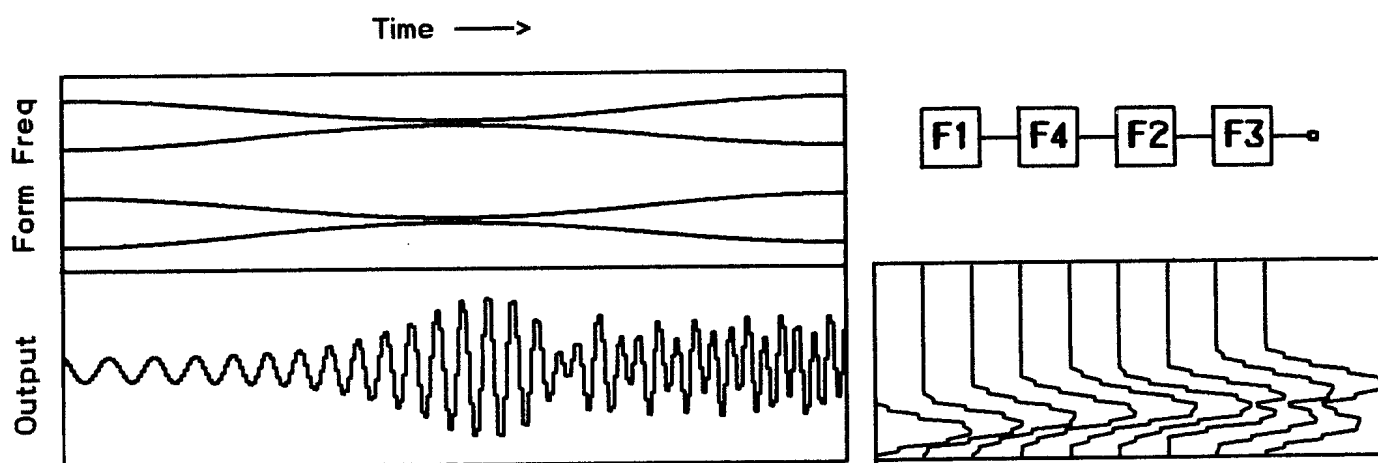
Fig 1.7. A parallel to the squeeze test, but using a classical cascade formant filter, F-pattern at top. A greater part of the energy is transferred to F2.

Fig 1.7   BASIC TOOLS   1 – 15

1.54 UNDERSAMPLING THE AREA FUNCTION

For computational economy one would like to update areas and reflection coefficients at intervals much longer than the sampling interval, to undersample the area function. The correspondent with formant and LPC synthesis, to undersample formant and bandwidth data, has always been practised successfully. Theoretically it is of course not permissible at all, but it is motivated from these data being pseudo stationary. There will however be artifacts in the synthesis results, like clicks or harshness. The classical way to minimize the artifacts applies to voiced sounds only, that is to update formant data pitch synchronously, when the oscillating energy is minimum, just before the major excitation at glottal closure. This trick is just as applicable to the line analog, but can hardly be applied when most needed, as when synthesizing consonantal transitions.

To fill in the picture fig 1.6b shows what happens in the squeeze test when areas are updated every seven sample intervals, and in correspondingly larger steps. It is obvious how the waveforms are severly distorted.

From several runs of the squeeze test with area undersampling my general conclusions, not very revolutioning, are:

The most prominent distortions and accompanying mode leakages occur when the ratio of undersampling coincides with low multiples or sub-multiples of the number of line segments; the artifacts then tend to accumulate in phase.

The Maeda model and the adjustment for continuous P-U give typically 10 dB lower mode leakage than the other models. These other, including the static, do not show very significant mutual differences in this respect.

Also it holds that the static model gives very different waveforms for the flow and the pressure partial wave analogy. Each of the dynamic models give essentially the same result with either type of analogy.

1.55 TEST CONCLUSIONS

This tends to resolve the competition between the different dynamic models:

For a very rapidly changing area, like the glottis, then area under-sampling is out of question, and the Strube model or the f-v adjustment would be appropriate. These models appear to account correctly for the dynamic effects on momentary wave amplitudes.

For the slower area changes in the vocal tract area undersampling can be used with some care, and then the Maeda model or the P-U adjustment are preferred to minimize spurious transients. A consequence will then be that the dynamic effects will be under-represented. More specifically, the changes in formant bandwidths due to dynamic area change will not be as large as they should.

Probably this is not very important, as was shown with a simple resonator example by Fant (1980). Concurrent tests by Meyer and Strube (1984) also seem to indicate the effect to be perceptually marginal. For the band-width contribution to be significant the relative area change must be fast, something that can happen only with small areas close to cutoff, the typical situation is the release of an occlusion. A balancing factor is that when the area is small the laminar and, perhaps even more, the jet losses are large. The point is, that for the bandwith term from the area change to overrule that of the resistive loss there is very little time available, up to a few milliseconds.

1.6 HANDLING THE SCATTERING EQUATIONS

Though the scattering equations are simple enough the use of them for simulation purposes offers several pitfalls to the programmer. It would be worthwhile to contemplate the detailed wave flow diagram in fig 1.8 showing the partial waves as little packages moving in space and time.

Here I have elected to mark the time scale in a time unit equal to the sample interval. But at the speed of sound the length of each tube corresponds to a travel time being only half of the sample interval. That is why the time axis is marked also in 'half units'. In the literature it happens that this 'half unit' is denoted by τ. The diagram shows all possible partial waves that can follow from the spreading of a single input pulse, injected at the start of the diagram.

The following features could be noted:

It might be possible to inject another input impulse half a sample inter-val after the original one. This new one would then spread into a similar pattern completely interlaced with the one shown. But <u>nowhere</u> would then these two patterns interact. Thus there is the theoretical possibility to handle two entirely independent wave systems simultaneously in the same line analog, but this appears to be of academic interest only.

Let us then scrap that possibility and work only with the causally con-nected waves shown. In each line in the diagram only half of the joints are involved, alternatingly the odd and the even ones. This should be taken care of in writing computer programs so that half of the computer power is not wasted on calculating zeroes. One way of doing this is to arrange the set of computations for one sample interval into two passes. The first pass handles the even numbered joints computing the odd num-bered forward waves and the even backward waves. The second pass goes through the odd joints to make the even forward and the odd backward waves.

In a specific tube, at each instant of time there is either a forward wave or a backward wave, not the two simultaneously. Compare the basic expressions (1.21) giving the net pressure and flow. These are strictly valid only at those time instants the waves meet at the joints.

A consequence is that the computations of the scattering equations can be made 'in place'. The results of one pass can be written back into their proper storage locations where the old values will no more be needed.
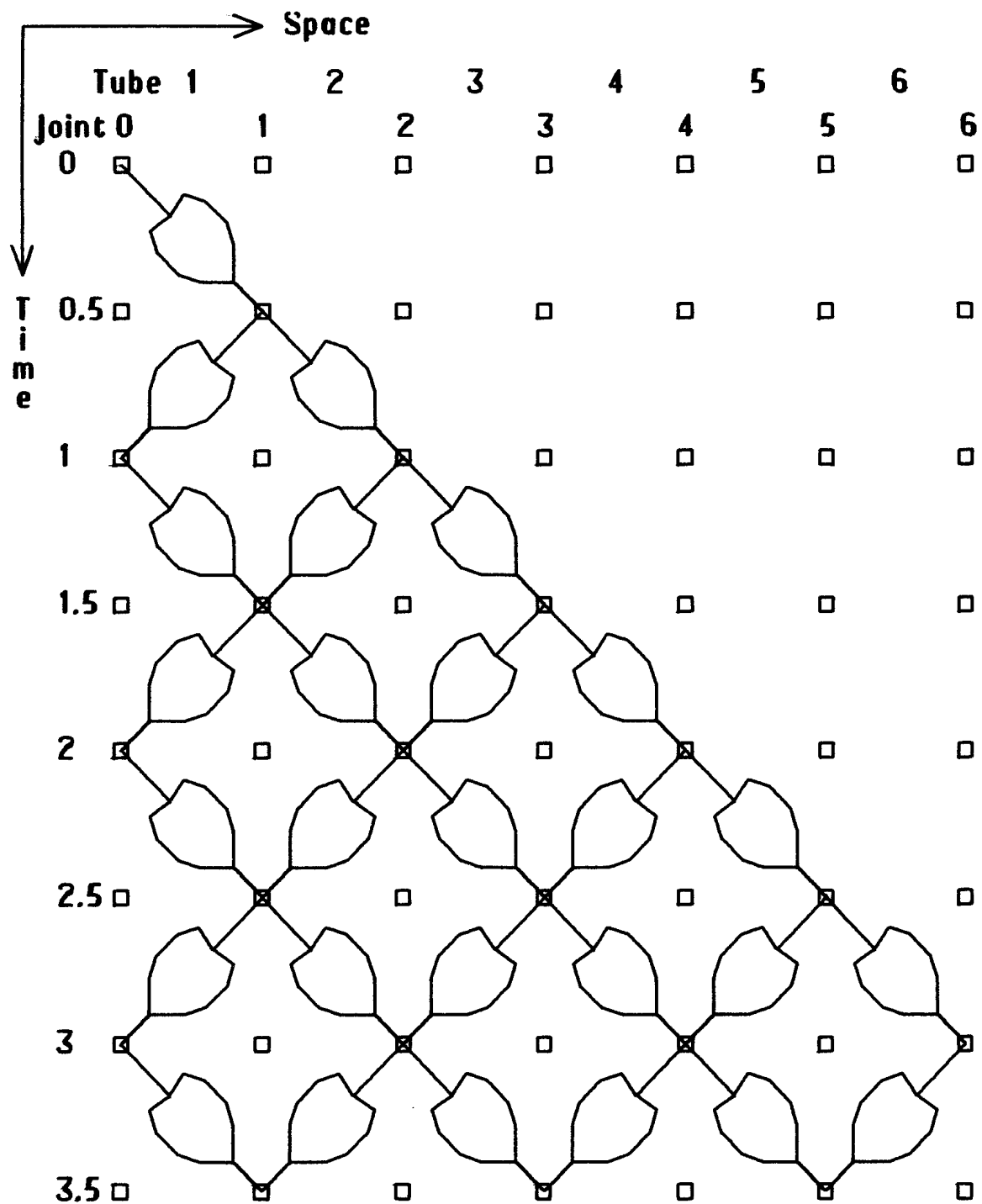
Fig 1.8. Space time diagram of the computation sequence in the line analog. Waves reach the output (here joint 6) only at integer multiples of the sampling interval.

Fig 1.8

Conceivably, area changes can be done at twice the nominal sampling rate. In practice this seems to be of no particular interest, rather a burden. Also, because of the sampling theorem it would not make sense to update areas more often than the sampling rate. Mostly one would instead like to update areas at some multiple of the sampling interval, not least in order to save computer time. Recomputation of areas and reflection coefficients generally takes more computer capacity than the scattering equations.

This has a paradoxal consequence: the dynamic scattering equations are used in only one of the sample interval passes, they will never be applied to more than every second joint. It may seem puzzling that, for instance, when changing the middle tube in the three tube shutter test, the dynamic correction is applied only to one end of that tube. But when the waves reach the other end half a sample interval later, then area conditions are static.

Similarly, when applying the novel wave adjustment scheme, only half of the forward waves and half of the backward waves need be adjusted. The others will never more enter the scattering equations.

It may be worth pointing out that neither dynamic model accounts for the pseudo DC pumping effect of an area change, that is the incremental flow from the volume change.

1.7 BRANCHING THE LINE

To simulate a nasal tract connected to the vocal tract by a velar valve we must have a way to branch the single line into two. There exist equivalent solutions to this in digital wave filter theory (Fettweis (1971), Fettweis and Meerkötter (1975)) comprising so called multi-port adaptors. To stay within limits of generality and to keep down the number of notations we however go on independently with direct reference to the paragraphs above.

Start by looking at the single joint that was treated in eq (1.23) and illustrated in fig 1.2.

Now look only at tube 1, how is the reflected wave s_1 composed? One part of it is the k_{12} fraction of the 'own' incident wave r_1. The other part of it is the $(1+k_{12})$ fraction of whatever wave arrives at the other side of the joint, here s_2.

It does not make any difference to what happens in tube 1 if the other tube is composed of two compartments that together have the area A_2. Let us then draw it as a three-way junction, fig 1.9a.

Of course you will have to make a decision on what is "forward" and what is "backward" in the different branches, and then later have that decision in mind.

Using (1.15) and the fact that tube 1 meets the area A_2+A_3 gives us

$$k_{1x} = (A_1 - (A_2+A_3)) / (A_1 + (A_2+A_3)) \qquad (1.71a)$$

The same reasoning applied when looking into tubes 2 and 3 renders

$$k_{2x} = (A_2 - (A_1+A_3)) / (A_1 + A_2 + A_3) \qquad (1.71b)$$
$$k_{3x} = (A_3 - (A_1+A_2)) / (A_1 + A_2 + A_3) \qquad (1.71c)$$

Observe that the k_{nx} here are all defined for waves going toward the joint. These may be "forward" or "backward" waves, not necessarily one kind only. In passing we note that any one of the three reflection coefficients can easily be eliminated since

$$k_{1x}+k_{2x}+k_{3x} = -1 \qquad (1.72)$$

and proceed to draw a graph of the waves and their splitting as in fig 1.9b.

a.   **b.**
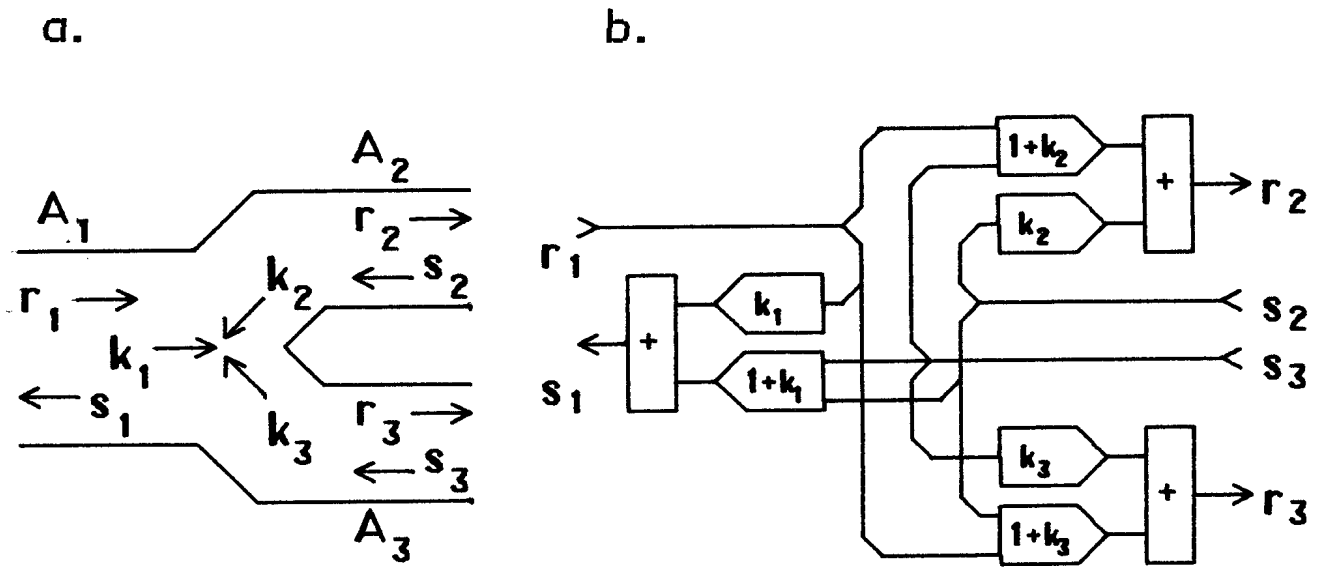
Fig 1.9. Left, layout, right, wave scattering in a three-way joint.

The equations of the three-tube joint can then be found by inspection and arranged as

$$u = r_1 + s_2 + s_3$$
$$r_2 = r_1 + s_3 + k_{2x}*u$$
$$s_1 = s_2 + s_3 + k_{1x}*u$$
$$r_3 = r_1 + s_2 + k_{3x}*u$$

(1.73)

with u being the total partial wave influx to the branching point (not the net influx which is zero).

With the pressure wave analogy it is not so simple to get the scattering correctly from a flow diagram. The safe thing is to start from a correspondent to (1.13), the total pressure in all branches is the same, and the sum of net flows is zero. These equations can eventually be solved as

$$t = k_{1x}*p_1 + k_{2x}*q_2 + k_{3x}*q_3$$
$$p_2 = p_1 + q_3 + t$$
$$q_1 = q_2 + q_3 + t$$
$$p_3 = p_1 + q_2 + t$$

(1.74)

using the k_{nx} of (1.71).

In wave filter terminology these scatterings are called three-port 'series adaptors' (1.73), and 'parallel adaptors' (1.74).

Fig 1.9   BASIC TOOLS   1 – 21

1.8 SERIES LOSS CORRECTION

The classical way to introduce losses used by Kelly-Lochbaum is to attenuate each partial wave by a small fraction each time it passes through a tube (though they appear to have done it with the backward waves only). This is also generally combined with losses at both ends of the line. The end reflection coefficients are set to a little less than unit magnitude. Then at the glottis end for instance, in each cycle there will be a small backward flow that will never enter the line again, it is lost for good. This works fine as a first approximation, but there are numerous things in the real vocal tract that can not be modelled properly this way.

As an example imagine the following case. The lips are shut to give a total reflection back into the vocal tract. All r_n and s_n inside the tract are equal and non-zero. This corresponds to the mouth being charged with a constant DC pressure but with no net flow. Should the losses be simulated by indiscriminate attenuation of the partial waves in such proportions as to give reasonable formant bandwidths, then the DC pressure would be lost in a fraction of a second. The modelled tract would appear to be made of felt rather than gas tight flesh.

Obviously we would like the line model to handle the losses in such a way that we can trace them back to the real tract and quantify them with a reasonable accuracy.

The first trick is now to augment the K-L model with series resistances in the line, not in balanced combinations of series and shunt resistances as with ideal lines. The absence of a shunt will insure an "airtight" line.

At joint n between two line segments of impedances Z_1 and Z_2 we insert a loss resistance R, fig 1.10.
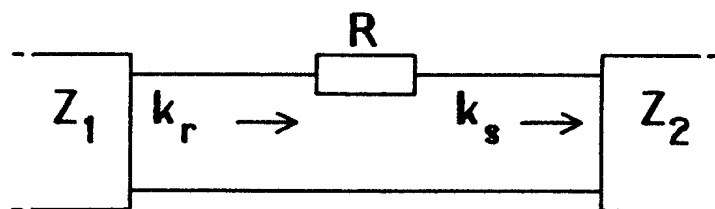
Fig 1.10. Series loss resistance between two sections.

Instead of the old reflection coefficient k_{12} we introduce the pair of a "forward reflection" coefficient k_{r12} and a "backward reflection" coef-

Fig 1.10   BASIC TOOLS   1 - 22

ficient k_{s12}, though still defined in the forward direction. In accordance with eq (1.15) we get, now omitting the '12' subscript

$$k_r = ((R+Z_2) - Z_1) / (R+Z_2+Z_1)$$
$$k_s = (Z_2 - (Z_1+R)) / (R+Z_2+Z_1)$$

(1.81)
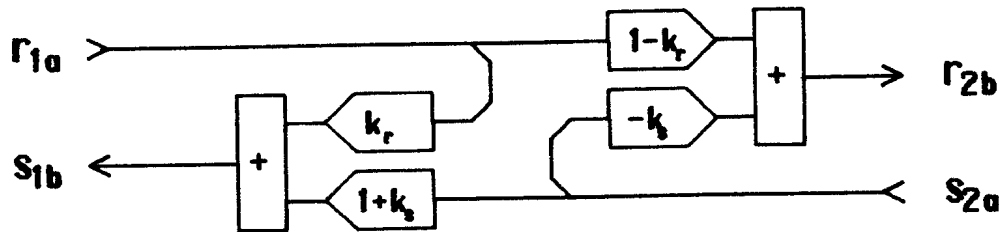
Fig 1.11. Scattering with series loss resistance.

How the waves are split is seen in the graph of fig 1.11, from which the new scattering equations can be written by inspection as

$$r_2 = (1-k_r)*r_1 - k_s*s_2$$
$$s_1 = k_r*r_1 + (1+k_s)*s_2$$

(1.82)

Now we introduce a __loss factor__ D as the ratio of R to the sum of the surrounding line impedances:

$$D = R / (Z_1 + Z_2)$$

(1.83)

This definition is in line with the classical one for the contribution to the attenuation constant by a series element in a line, compare fig 2.1. It should be noted, however, that we work here with a resistance, not a specific resistance per unit length. Inserting this and (1.15) into (1.81) gives

$$k_r = (k+D) / (1+D)$$
$$k_s = (k-D) / (1+D)$$

(1.84)

which, used in eqs (1.82), gives the desired series loss corrected scattering equations

$$r_2 = r_1 - (k*(r_1+s_2)+D*(r_1-s_2))/(1+D)$$
$$s_1 = s_2 + (k*(r_1+s_2)+D*(r_1-s_2))/(1+D)$$

(1.85)

Again, to bring out more clearly the effect of the series loss we rewrite (1.85) like the formula for the lossless case (1.23), but augmented with a series loss correction term, and get

Fig 1.11   BASIC TOOLS   1 - 23

$$r_2 = r_1 - (r_1 + s_2) * k - ((1-k) * r_1 - (1+k) * s_2) * D/(1+D) \qquad (1.86)$$
$$s_1 = s_2 + (r_1 + s_2) * k + ((1-k) * r_1 - (1+k) * s_2) * D/(1+D)$$

The loss term is sensitive to a weighted difference between the forward and backward waves, this corroborates it represents a loss due to the net flow rather than the pressure. The weighting performed by the factors $1+-k$ accounts for the area change over the joint. The loss term enters the two waves of (1.86) with equal magnitude but opposite sign. Thus the sum of the computed waves r_2 and s_1 is equal to the sum of the input waves r_1 and s_2, no flow or matter is lost at the joint. What happens with the series loss is then that the balance between the forward and backward waves is adjusted as compared to the lossless case.

The lossy scattering for the pressure wave analogy can be found as

$$p_2 = p_1 + (p_1 - q_2) * k - (p_1 - q_2) * (1+k) * D/(1+D) \qquad (1.87)$$
$$q_1 = q_2 + (p_1 - q_2) * k + (p_1 - q_2) * (1-k) * D/(1+D)$$

also this one written as the regular static lossless formula (1.16) plus a loss correction term.

1.9 SHUNT LOSS CORRECTION

Let us now in a corresponding dual manner introduce a shunt loss conductance element G into the joint. Let that element constitute the third branch in a three way joint as in fig 1.12.
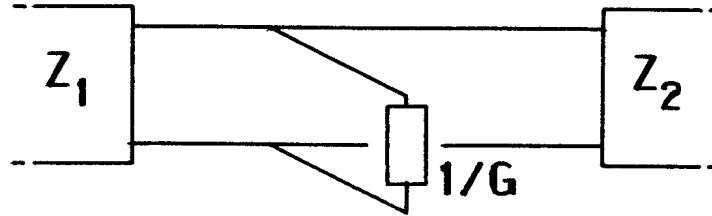
Fig 1.12. Shunt loss conductance G as a branch.

Let us also define a shunt loss coefficient E in analogy to the classical definition in fig 2.1, $E=GZ/2$ giving $G=2*E/Z=2*E*Y$. Furthermore select a norm line admittance Y such that it represents the mean admittance of the joining tubes. We then have

$$G = 2*E/Z = E/Z_1 + E/Z_2 \qquad\qquad (1.91a)$$

and using (1.11) we turn it into

$$\rho*c*G = E*(A_1+A_2) = A_3 \qquad\qquad (1.91b)$$

This represents the area A_3 in the loss branch of the three-way joint. Its magnitude is a fraction E of the sum of the joining line areas. Through this (mostly small) area energy will leak out as a forward wave r_3. This energy is lost, so there will be no backward wave s_3.

Inserting this into the expressions (1.71) for the reflection coefficients we get

$$k_{1x} = \frac{A_1-A_2-E*(A_1+A_2)}{A_1+A_2+E*(A_1+A_2)} = (k_{12}-E)/(1+E) \qquad\qquad (1.92)$$

$$k_{2x} = \ldots \qquad = -(k_{12}+E)/(1+E)$$

$$k_{3x} = \ldots \qquad = (E-1)/(1+E)$$

where k_{12} is the usual reflection coefficient for the main path as in (1.15). Going on inserting (1.92) into the scattering equations (1.73), omitting the subscript on k_{12} and remembering that $s_3=0$ we get

Fig 1.12   BASIC TOOLS 1 - 25

$$r_2 = r_1 - k*(r_1+s_2) - (1-k)*(r_1+s_2)*E/(1+E) \qquad \cdot(1.93)$$
$$s_1 = s_2 + k*(r_1+s_2) - (1+k)*(r_1+s_2)*E/(1+E)$$
$$r_3 = \qquad\qquad 2 \quad *(r_1+s_2)*E/(1+E)$$

These are put in a form to show identity with the plain two tube joint equations (1.23) augmented with a shunt loss term. The equations clearly show how the loss flow r_3 is tapped off from r_2 and s_1 in fractions balanced by k.

Again, for the pressure wave analogy, one can similarly derive a shunt loss corrected version to compare to the lossless (1.16):

$$p_2 = p_1 + (p_1-q_2)*k - (p_1*(1+k)-q_2*(1-k))*E/(1+E) \qquad (1.94)$$
$$q_1 = q_2 + (p_1-q_2)*k - (p_1*(1+k)-q_2*(1-k))*E/(1+E)$$
$$p_3 = \qquad\qquad (p_1*(1+k)-q_2*(1-k)) / (1+E)$$

The scattering equations for flow and pressure partial waves, with corrections individually for series and shunt losses, have now been outlined as (1.86), (1.87), (1.93), and (1.94). These are shown collected together in table 1.2.

The series and shunt loss corrections can be applied additively. For instance, setting D=E (as for an ideal line) such additive correction gives back the original equations (1.16) and (1.23), but uniformly scaled down by the factor (1-D)/(1+D). This then reminds of the small attenuation to the partial waves applied by Kelly and Lochbaum in their original model.

There are many alternative ways to write the scattering equations. One that shows the quartet of lossy equation pairs in terms of reflection coefficients, rescaled by the loss factor like in expression (1.84) is also given in table 1.2. From that it is obvious how keeping track of pressure and flow analogies, and series and shunt losses, is one big excercise in putting the correct signs at the correct places.

Table 1.2. Static scattering equations, separately accounting for series and shunt losses, for pressure and flow partial wave analogies.

SCATTERING WITH SERIES LOSS CORRECTION. $D = R/(Z1+Z2)$

PRESS $\quad p_2 = p_1 + (p_1-q_2)*k - (p_1-q_2)*(1+k)*D/(1+D)$ $\qquad$ (1.87)

$\qquad\quad q_1 = q_2 + (p_1-q_2)*k + (p_1-q_2)*(1-k)*D/(1+D)$

FLOW $\quad r_2 = r_1 - (r_1+s_2)*k - (r_1*(1-k)-s_2*(1+k))*D/(1+D)$ $\quad$ (1.86)

$\qquad\quad s_1 = s_2 + (r_1+s_2)*k + (r_1*(1-k)-s_2*(1+k))*D/(1+D)$

SCATTERING WITH SHUNT LOSS CORRECTION. $E = G/(1/Z1+1/Z2)$

PRESS $\quad p_2 = p_1 + (p_1-q_2)*k - (p_1*(1+k)+q_2*(1-k))*E/(1+E)$ $\quad$ (1.94)

$\qquad\quad q_1 = q_2 + (p_1-q_2)*k - (p_1*(1+k)+q_2*(1-k))*E/(1+E)$

FLOW $\quad r_2 = r_1 - (r_1+s_2)*k - (r_1+s_2)*(1-k)*E/(1+E)$ $\qquad$ (1.93)

$\qquad\quad s_1 = s_2 + (r_1+s_2)*k - (r_1+s_2)*(1+k)*E/(1+E)$

LOSSY SCATTERING, ALTERNATE FORM

	PRESS	FLOW

SERIES LOSS

$$p_2 = p_1 + \frac{k-D}{1+D}*p_1 - \frac{k-D}{1+D}*q_2 \qquad r_2 = r_1 - \frac{k+D}{1+D}*r_1 - \frac{k-D}{1+D}*s_2$$

$$q_1 = q_2 + \frac{k+D}{1+D}*p_1 - \frac{k+D}{1+D}*q_2 \qquad s_1 = s_2 + \frac{k+D}{1+D}*r_1 + \frac{k-D}{1+D}*s_2$$

SHUNT LOSS

$$p_2 = p_1 + \frac{k-E}{1+E}*p_1 - \frac{k+E}{1+E}*q_2 \qquad r_2 = r_1 - \frac{k+E}{1+E}*r_1 - \frac{k+E}{1+E}*s_2$$

$$q_1 = q_2 + \frac{k-E}{1+E}*p_1 - \frac{k+E}{1+E}*q_2 \qquad s_1 = s_2 + \frac{k-E}{1+E}*r_1 + \frac{k-E}{1+E}*s_2$$

Tab 1.2

1.10 MODELLING FREQUENCY DEPENDENCE

Viscous and heat conduction loss resistances exhibit a proportionality to the square root of frequency. One can account for that dependence doing a trick with the loss factor D (or E). Instead of having it as a constant, let us devise a suitable expression using the unit delay operator z. One such expression that is useful is for instance

$$D = D_0 * (z-b)/(z-a) \qquad\qquad (1.101)$$

where D_0 is a reference value of the loss factor that has to be determined from the actual conditions of area etc and corrected in relation to the magnitude of $(z-b)/(z-a)$ at some reference frequency.

Fig 1.13 shows the real and imaginary parts of (1.101) for two different sets of the parameters a and b. These have been selected to give a reasonable approximation to square root variation in the real part as a function of frequency. The set in fig 1.13a could be used to model viscous losses in their principal frequency range of interest between .5 and 4 kHz.

Using fig 1.13b for a shunt conductance the parameter b=1 is selected to give infinite impedance at zero frequency. This would be necessary to avoid DC losses if used for shunt loss modelling (of heat conduction loss), but of course will give a quantitatively poor approximation at low frequencies.

To install this loss model into the line analog let us take the following example covering viscous series losses. In the scattering equations (1.86) the losses are represented by the last term which can be put as

$$x = (1-k)*r_1 - (1+k)*s_2 \qquad\qquad (1.102a)$$
$$y = x * D/(1+D) \qquad\qquad (1.102b)$$

x is then the input quantity and y is the frequency dependent result to use as a correction in (1.86). Inserting (1.101) into (1.102b) and solving for y renders

$$y = (x - b*x*z^{-1})*D_0/(1+D_0) + y*z^{-1}*(a+b*D_0)/(1+D_0) \qquad (1.103)$$

Thus we can find the current value of y from the one sample interval delayed y, that is, $y*z^{-1}$, and from the current and the delayed values of x. The procedure obviously requires extra storage for the delayed x and y.

a.
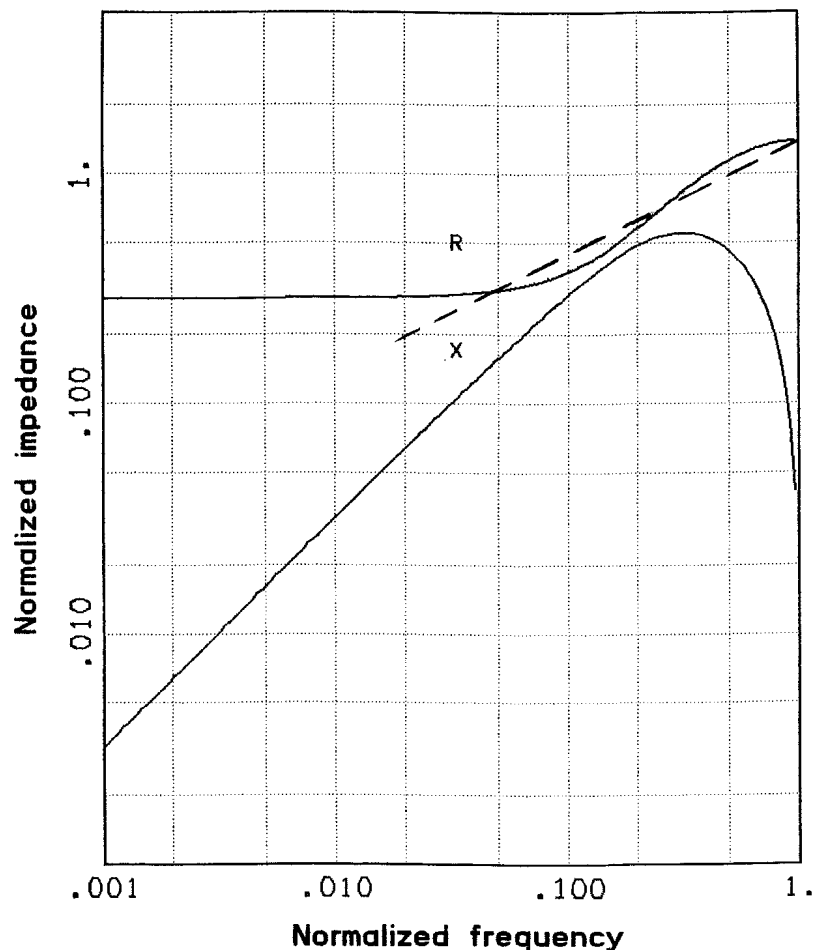
P   .300
Z   .800
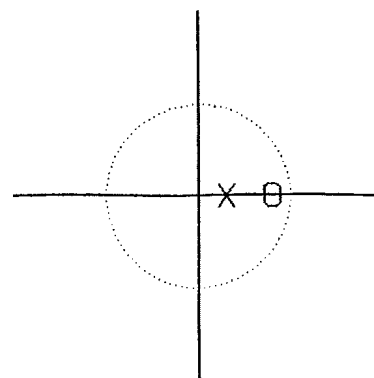
b.
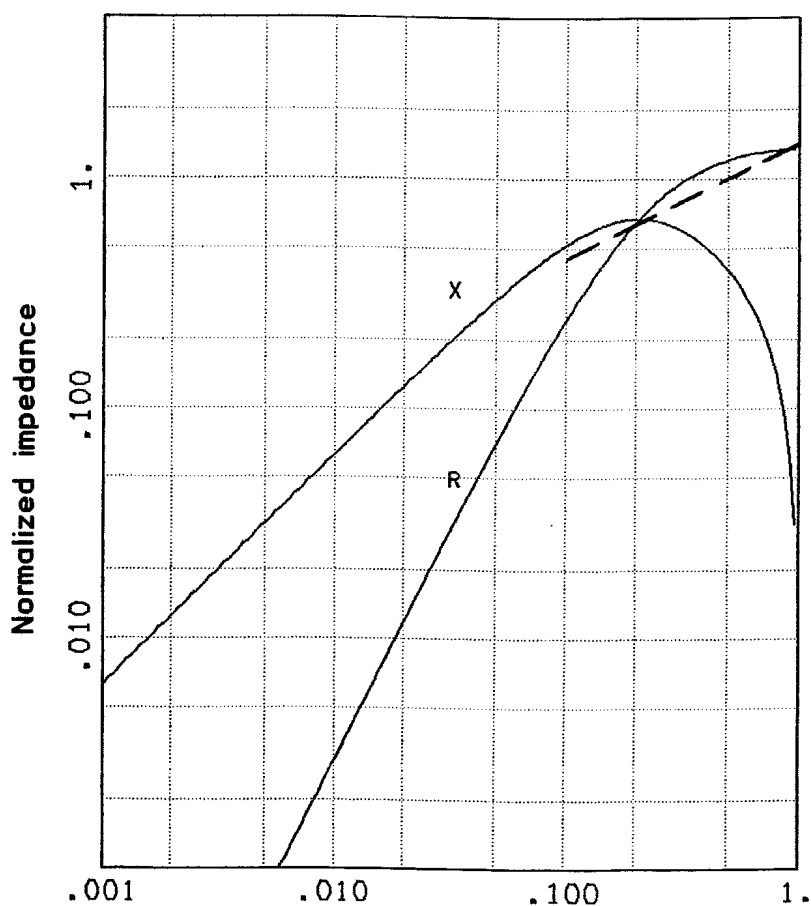
P   .500
Z   1.000
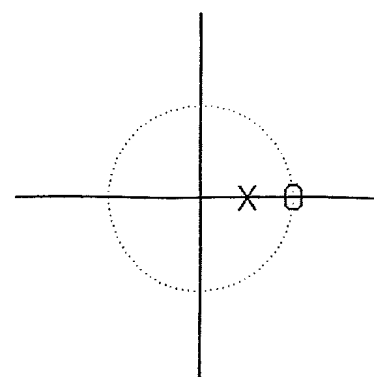
Fig 1.13.Two z-domain approximations to square root behaviour of a resistance in the frequency domain.

a. Approximation covering the upper part of the audio band.

b. Approximation with the added requirement of zero resistance or conductance at zero frequency. This alternative misbehaves for low frequencies and then becomes reactive.

1.11 INSERTION OF A PRESSURE GENERATOR

Later we will want to inject a signal into the line at some point other than its ends, especially in order to achieve noise generation at constricted passages. To that end let us briefly see how this will come out as a modification to the scattering equations.

Setting out with the continuity criteria (1.13) we insert a pressure source P_e between tubes 1 and 2 such that the total pressure undergoes a sudden jump. (1.13a) is then modified into

$$P_1 + P_e = P_2 \qquad\qquad (1.110)$$

while (1.13b,c) remain unchanged. Solving for p_{2b} and q_{1b} as before the scattering equations (1.16) will come out with additional terms as

$$p_{2b} = (1+k_{12})*p_{1a} - k_{12}*q_{2a} + (1+k_{12})*P_e/2 \qquad (1.111)$$
$$q_{1b} = k_{12}*p_{1a} + (1-k_{12})*q_{2a} - (1-k_{12})*P_e/2$$

This is easy to visualize directly. The inserted pressure P_e is split into two parts, balanced by the reflection coefficient k_{12}. One part goes on in the forward direction with a positive sign, the other goes with a negative sign in the backward direction. In case the two tubes are equal, then k_{12} is zero and the injected waves become equal in magnitude, namely half the inserted P_e.

If we instead consider the flow analogy doing exactly the same derivation from (1.22) the modified resulting correspondent to (1.23) will be

$$r_{2b} = r_{1a} - k_{12}*(r_{1a}+s_{2a}) + P_e/(Z_1+Z_2) \qquad (1.112)$$
$$s_{1b} = s_{2a} + k_{12}*(r_{1a}+s_{2a}) - P_e/(Z_1+Z_2)$$

As usual with the flow analogy this is perhaps even more tractable to the mind. The partial flow waves induced by the P_e generator are equal in magnitude and opposite in sign, conservation of matter and no hole in the wall. The magnitudes of the partial waves are as could be immediately expected from the aggregate of the pressure generator and the impedances of the two tubes.

Fig 1.13   BASIC TOOLS   1 - 28

1.12 KINETIC PRESSURE DROP

A weakness in any linear line model is that it fails to account for the pressure changes from accelerations at constrictions in the line, often referred to as the Bernoulli effect or the kinetic pressure drop. The basic reason for this is that unlike electricity for which line theory was developed, in acoustics the transported medium itself has a mass.

To account for this at each joint in our model appears to bring about an almost exorbitant complexity, so let us hold back and restrict ourselves to examine a simplified case. In reality areas gradually decrease approaching a constriction, and the kinetic drop is correspondingly gradual. The approximation now considered is to regard the drop to occur at one specific place, namely at the entry of the narrowest tube. Identify this drop with an extraneous pressure generator P_e as in (1.112). We get from the elementary expression of kinetic pressure $P = \rho * v^2 / 2$

$$P_e = \frac{\rho}{2} * \left(\left(\frac{U_1}{A_1}\right)^2 - \left(\frac{U_2}{A_2}\right)^2 \right) \tag{1.120}$$

remembering $U_1 = U_2$ because of continuity. Our approximation is now that we assume the entering particle velocity U_1/A_1 to be negligibly small beside the velocity U_2/A_2 in the constriction. Similarly we neglect A_2 beside A_1 and also Z_1 beside Z_2. Now identify the final terms in (1.112) as an extraneous correction flow U_e, and using (1.120) we approximate it as

$$U_e = P_e/(Z_1 + Z_2) \approx - U_2^2/(2 * c * A_2) \tag{1.121}$$

From (1.22) we have $U_1 = r_{1a} - s_{1b} = U_2 = r_{2b} - s_{2a}$ and inserting (1.112) we can express the net flows as

$$U_1 = U_2 = r_{1a} - s_{2a} - k_{12} * (r_{1a} + s_{2a}) + P_e/(Z_1 + Z_2) = U_o + U_e \tag{1.122}$$

the point being that they are now expressed in terms of the incident waves only. I have also identified the two auxiliary flow quantities U_o and U_e. U_o is the net flow we would have had with the original scattering equations, and U_e is the correction flow to examine now. Combining (1.121) and (1.122) makes the quadratic equation

$$U_e * 2 * c * A_2 + (U_o + U_e)^2 = 0 \tag{1.123}$$

which can be solved as

$$\frac{U_e}{U_o} = -(1+y) \begin{array}{c} + \\ (-) \end{array} \sqrt{y*(2+y)} \qquad \text{with} \qquad (1.124)$$

$$y = c*A_2/U_o = c/v_o$$

y is a velocity ratio between the speed of sound and the uncorrected particle velocity.

Fig 1.14 shows this correction ratio U_e/U_o as a function of the velocity ratio y. The horizontal scale is simultaneousiy calibrated in the velocity $v_o=U_o/A_2$ and the corresponding kinetic pressure. It is easy to see that for large y, that is, for low particle velocities the curve closely approximates the relation

$$- U_e/U_o \approx 0.5*U_o/c*A_2 = .5*v_o/c = 1/2y \qquad (1.124b)$$

This is very like (1.121) except that it uses the uncorrected flow U_o instead of the corrected flow U_2, and just indicates that provided the correction is small it does not matter which one of them we use. For small y, high particle velocity (1.124) will instead approximate $U_e=-U_o$.
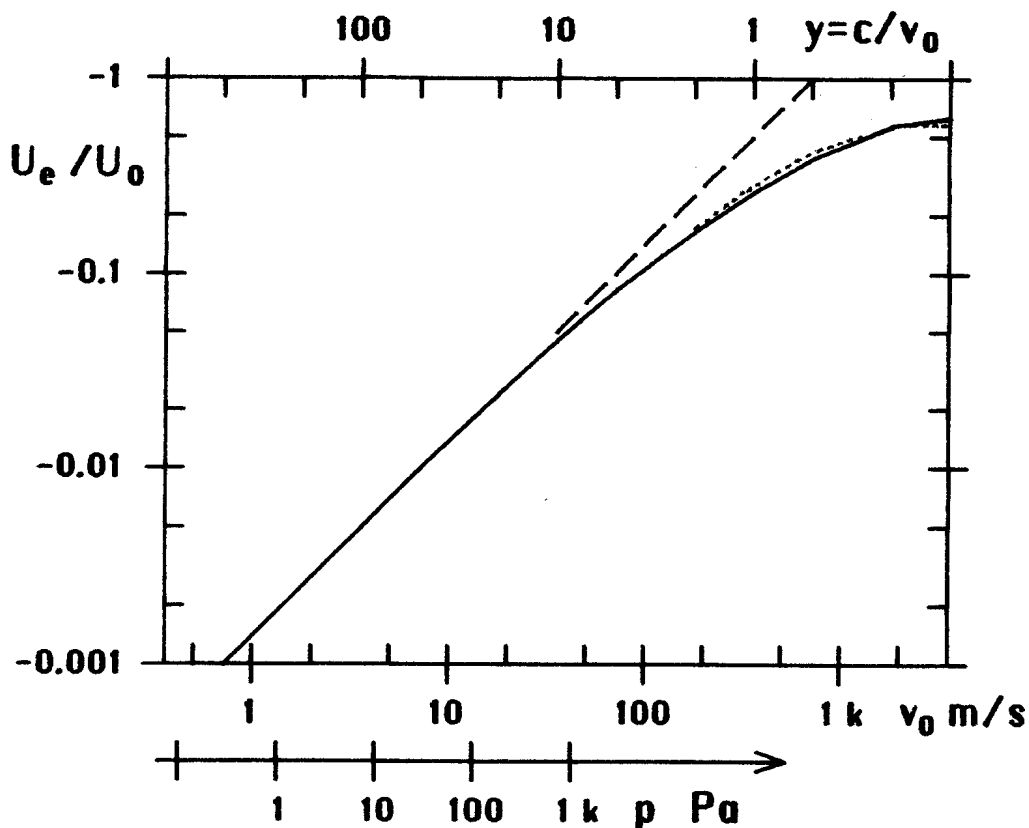
Fig 1.14. Ratio of kinetic correction flow to uncorrected flow vs. uncorrected particle velocity and kinetic pressure drop.

The expression (1.124) is unnecessarily complicated. To save a little computer time you can use the simple approximation

Fig 1.14   BASIC TOOLS   1 - 30

$$U_e/U_o \approx -1/(1.5+2|y|)$$
$$(1.125)$$

dotted in fig 1.14. The absolute sign on y is important. The correction always operates to diminish the absolute value of U_o, irrespective of its sign.

To summarize, the flow scattering equations (1.112) corrected for the kinetic pressure drop will be

$$u_{12} = k_{12}*(r_{1a}+s_{2a}) \qquad\qquad =(1.23)$$
$$U_o = r_{1a} - s_{2a} - u_{12} \qquad\qquad =(1.122)$$
$$U_e \approx -U_o|U_o|/(1.5|U_o| + 2*c*A_2) \qquad \text{from } (1.125)$$
$$r_{2b} = r_{1a} - u_{12} + U_e \qquad\qquad =(1.112)$$
$$s_{1b} = s_{2a} + u_{12} - U_e$$

In practice it is essential that the approximation to the correction as in (1.124b) is not permitted to become numerically greater than U_o, as it would never become by the 'exact' equation (1.124). Otherwise there will be an over-correction causing a change of flow direction and instability in the computations.

It might appear unnecessary to stress this since that danger limit is not reached until v_o exceeds 2c, a completely absurd velocity in speech. But, unless you tame it, this is precisely what the flow wave analogy may give you when a tube is closed down towards a small area. For proper simulation of a closure some series loss, for instance this kinetic correction, is mandatory.

A corresponding procedure with the pressure wave analogy may be to find the particle velocity in the narrow section 2 using (1.11) and (1.12b) as

$$v_2 = U_2/A_2 = (p_{2b} - q_{2a})/\rho c \qquad\qquad (1.126)$$

and, using (1.111) have it in terms of the incident waves as

$$v_2 = (p_{1a} - q_{2a} + P_e/2)*(1+k)/\rho c \qquad\qquad (1.126b)$$

In the absence of P_e correction we would have had the uncorrected velocity

$$v_o = (p_{1a} - q_{2a})*(1+k)/\rho c \qquad\qquad (1.126b)$$

The assumption that A_2 is much less than A_1 implies that k is approximatively one. Let us further define the auxiliary pressure quantity P_o as

$$P_o = 2*(p_{1a} - q_{2a}) \qquad\qquad (1.127)$$

Inserting this and k=1 into (1.126b) gives the approximate

$$v_2 \approx (P_o + P_e)/\rho c \qquad (1.128)$$

which we insert into the kinetic pressure equation to find

$$-P_e = \frac{\rho}{2}v_2{}^2 = \frac{1}{2\rho c^2} * (P_o + P_e)^2 \qquad (1.129)$$

This is solved for P_e/P_o and we arrive at exactly the same expression as (1.124) for the quotient, with

$$y = \rho c^2 / P_o = c/v_o \qquad (1.1210)$$

like before.

The kinetic correction procedure with the pressure wave analogy can thus be summarized as

$$P_o = 2*(p_{1a} - q_{2a}) \qquad\qquad = (1.127)$$
$$P_e \approx -P_o * |P_o|/(1.5|P_o| + 2\rho c^2) \qquad \text{from (1.1210), (1.125)}$$

and insert this into scattering equations (1.111). These can also be rearranged using (1.127):

$$p_{2b} = q_{2a} + (1+k)*(P_o + P_e)/2 \qquad (1.1211)$$
$$q_{1b} = p_{1a} - (1-k)*(P_o + P_e)/2$$
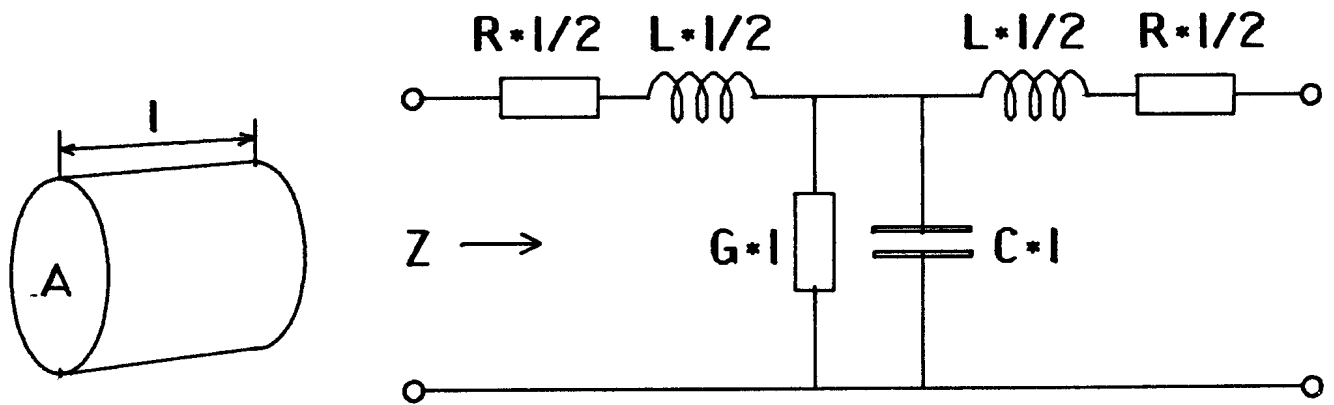
$$L = \rho/A \qquad Ns^2/m^6 \qquad \text{inductance per unit length}$$

$$C = A/(\rho * c^2) \qquad m^4/N \qquad \text{capacitance per unit length}$$

$$Z = \sqrt{L/C} = \rho * c/A \qquad Ns/m^5 \qquad \text{characteristic impedance}$$

$$D = R/2Z + GZ/2 \qquad Nep/m \qquad \text{attenuation per unit length}$$

Fig 2.1. Acoustic line segment and its analogy with key elements and properties.

Fig 2.1

2.   EVALUATION OF LOSS MECHANISMS IN THE VOCAL TRACT

Formal evaluation of distributed series and shunt losses from viscosity and wall interaction. Discrete losses from radiation, glottal resistance, jet formation. Magnitude comparison of different loss mechanisms.

There are a substantial number of different loss mechanisms having important influence on the waves in the line. They are mostly dependent of the geometry but also often of the pressure, flow, or frequency. This makes the total picture rather complex.

The following is a concise review and quantification of the losses. Most of these are well established in the literature, see for instance Fant (1960) and Flanagan (1965). Those and other results will be compiled and all put into terms of contributions to the loss factor. The aim is to bring out the numerical coefficients needed in the loss corrected scattering equations and to elucidate their relative importance.

For reference fig 2.1 shows the conventional expressions for some main elements and characteristics in the acoustic line.

R and G are series resistance and shunt conductance per unit length. The attenuation constant expresses the attenuation of the wave per unit length of the line. In line theory the expression shown is valid only if R is small compared to Z, and 1/G is large. We do not restrict ourselves to that condition since we use the formula mainly as a normalizing vehicle. This will not impair the validity of the modelling, but the special interpretation as attenuation per unit length is not granted at high losses. But this does not matter very much since the concept of attenuation as such is not particularly interesting in the speech synthesis application.

To enable a unified overview the various losses will be formulated in terms of their contributions to the attenuation constant. In the application at hand this quantity is of course anything but constant, so I prefer to use the alternate term loss factor.

All the distributed loss phenomena are treated under the assumption of a circular cross section. Mostly the losses are due to effects at the tube wall surface, the exception is the turbulence loss, and thus sensitive to the relation between the perimeter and area of the tube. A convenient way to account for other than circular cross sections is to multiply the loss coefficients by a 'shape factor'. This factor is the ratio of perimeters for the actual shape and for the circular shape at equal areas. In the vocal tract this shape factor is often of a magnitude around 2.

2.1 SERIES LOSSES

The class of series losses incorporates what we may call laminar, turbulent, and viscous losses. They have in common that they primarily are functions of the flow, not the pressure, and also of course of the area. They will have to be distributed along the line in order to model the sometimes very selective influence on formant bandwidths. The selectivity is because losses at specific places in the tract will affect the formants in relation to their individual resonance patterns.

These three loss mechanisms are basically in fact one and the same, but they manifest themselves differently depending on the circumstances. They are all due to shear in the air stream. With the laminar and viscous cases the shear is in the boundary layer just inside the tube wall where the particle velocity increases with the distance from the wall. The effective thickness of this boundary layer is inversely proportional to the square root of frequency. This is reflected in the viscous loss. At sufficiently low frequencies or small diameters the boundary layer thickness becomes comparable to the tube cross dimension and then the frequency dependence vanishes and we reach the laminar region. Is however the velocity high enough the flow is no longer stable, the boundary layer more or less vanishes and the whole passage is filled with turbulence.

2.11 LAMINAR AND TURBULENT DC FLOW LOSS

The pressure drop in a steady flow through a cylindrical pipe is generally given in the handbook literature (e g Schlichting (1951), quoted by Fant (1960)) as

$$p = \rho * l * v^2 * \lambda / (2 * d) \qquad (2.1)$$

p is the pressure drop at the speed v through the pipe of length l and diameter d. λ is a resistance coefficient depending on the Reynold's number

$$Re = v * d * \rho / \mu \qquad (2.2)$$

For laminar flow the resistance coefficient is

$$\lambda_{lam} = 64 / Re \qquad (2.3)$$

and for turbulent flow, empirically

$$\lambda_{trb} = 0.316 * Re^{-1/4} = (100 * Re)^{-1/4} \qquad (2.4)$$

Expressions (2.3) and (2.4) together imply the transition between laminar and turbulent flow to occur at Re = 1188. The formulas are valid for liquids and for gases when pressures are low enough that compression can be neglected.

Compiling (2.1)-(2.3) and the general relations $R=p/U$, $v=U/A$, $Z=\rho*c/A$, $A=\pi*d^2/4$ will render a loss factor, for the laminar flow at circular cross section

$$D_{lam} = R/2Z = 4*\pi*\mu/(\rho*c*A) \qquad\qquad (2.5)$$

numerically becoming

$$D_{lam} = 5.858*10^{-7} / A \qquad\qquad /m \qquad\qquad (2.6)$$

with A in m^2.

Similarly using (2.4) for turbulent flow I get the rather awkward expression

$$D_{trb} = \frac{1}{8*c} * (\frac{\pi^{5/2} * \mu}{200*\rho} * \frac{U^3}{A^{11/2}})^{1/4} \qquad\qquad (2.7)$$

and, with the constants inserted, U in m^3/s and A in m^2

$$D_{trb} = 1.234*10^{-5} * U^{+.75} * A^{-1.375} \qquad\qquad (2.8)$$

2.12 VISCOUS LOSS

Contrasting to these steady flow losses there is the frequency dependent viscous loss in the boundary layer at the tube wall. The loss resistance is, after Fant (1960), p 32

$$R_{vsc} = (S / A^2) * (\omega*\mu*\rho/2)^{1/2} \qquad\qquad (2.9)$$

S being the circumference of the tube. For simplicity assuming a circular section and reusing the basic relations given before (2.5) it can be put into a viscosity loss factor

$$D_{vsc} = \frac{R_{vsc}}{2Z} = \frac{\pi}{c} * (\frac{f*\mu}{\rho*A})^{1/2} \qquad\qquad (2.10)$$

and numerically, with f in Hz and A in m^2

$$D_{vsc} = 3.626*10^{-5} * (f/A)^{1/2} \qquad\qquad (2.11)$$

We observe that the only loss factor yet mentioned that is really well behaved is the one for laminar flow, it is independent of both frequency and flow and represents a pure resistance. The turbulent loss factor is flow dependent, as is the consequence of a non-linear resistance, and the

viscous loss factor depends of frequency.

To give a hint of their relative importance the laminar, turbulent, and viscous loss factors are shown as composite diagrams in fig 2.2 and 2.3. Here the loss factor is given as a function of flow and tube area. For the viscous loss factor a frequency of 1 kHz is assumed. The plots show the largest of the three rather than their sum. This causes the lines of equal loss to exhibit distinct breaks instead of smooth bends. The plot is divided into three regions dominiated in turn by each of the loss types. For the flow range of interest in speech the viscous and turbulent losses are the most prominent ones. The laminar losses rule the actions only close to complete cutoff.

The fact that laminar losses are negligible in vowels does not mean that they can be let aside in the model, namely if it shall be used also for articulations and sounds involving area cutoff.

2.13 JET LOSS

This section is more elaborated since the phenomenon is relatively little mentioned in the speech modelling literature.

When an air stream meets a more or less sharp constriction, then static pressure power is converted into kinetic power obeying the classical relation $p = \rho * v^2/2$. (Compare Fant (1960), pp 269-274, discussing the glottal flow. Many times the formula is augmented with a near unity constant (k) to account for shape variations and viscosity effects). Though this pressure drop in principle does not mean a loss of power it can be rewritten in our formulation of the loss coefficient using the ratio $p/vA = p/U$ for the 'resistance'. As shown by Fant a simple differentiation leads to a differential resistance of twice this value. We would then have $R = 2p/vA$ and putting it into a loss factor would then give

$$D_{jet}*l = (2p/vA)/2Z = v/2c = U/2cA \qquad (2.12)$$

The notation D*l rather than D is to take into account that this loss effect is not a distributed one. The loss resistance is a discrete one at a specific place, independent of the tube length l. The notation D*l then will keep the dimension of D to be 1/m.

Let us compare this formula to findings of Ingard and Ising (1967) investigating the impedance of a circular orifice in a thin plate. At low sound levels this impedance behaves the classical way as the radiation impedance of a baffled piston. At high levels however the situation changes completely as the resistive part will dominate and take the specific value $\rho * v$ where v is the peak particle velocity. The critical velocity of this takeover was around 1 m/s, independent of orifice area. They also noted that the takeover coincided with the formation of jets of
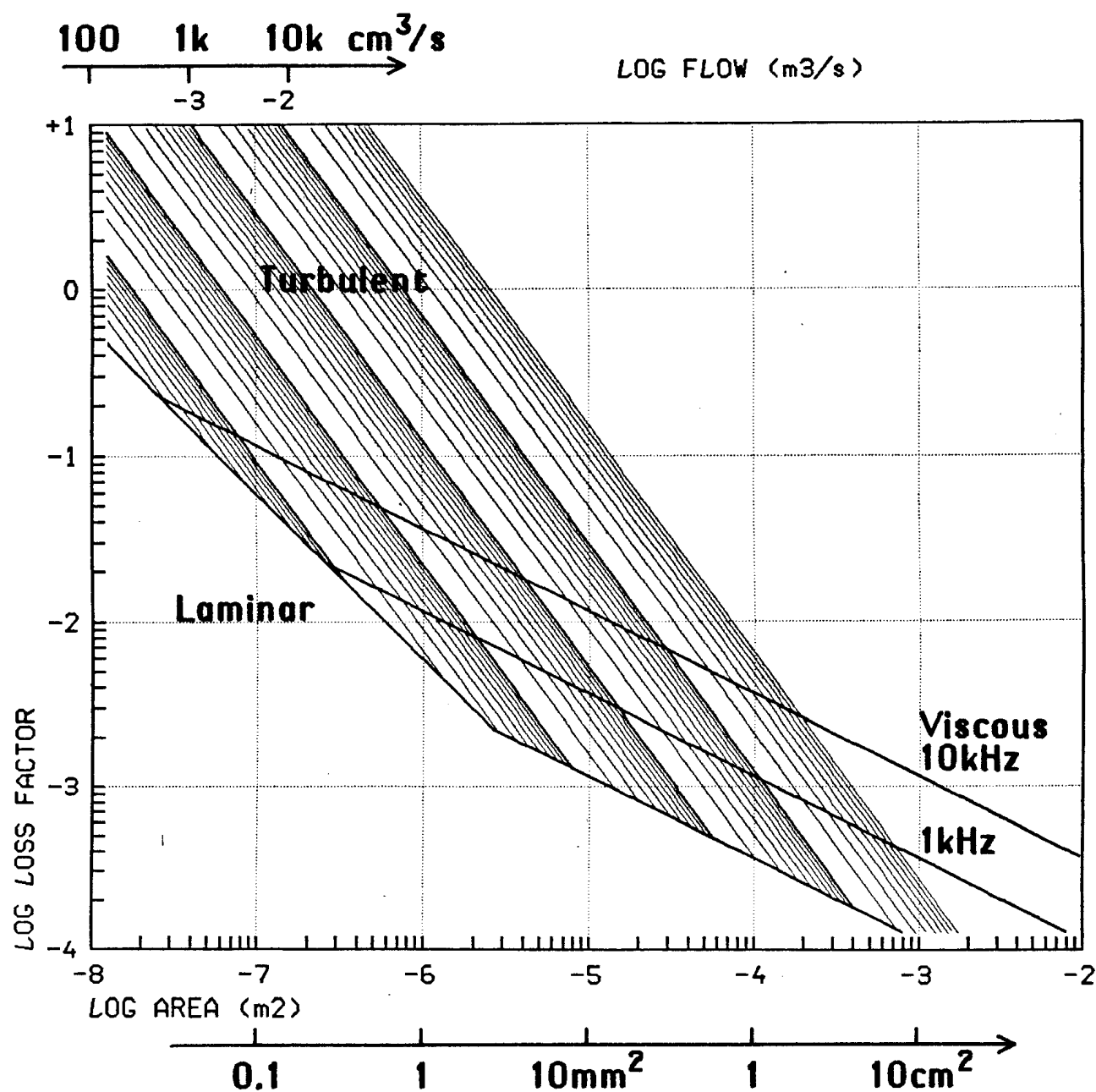
Fig 2.2. Series loss factor per unit length, R/2Z, as a function of area, with frequency and flow as parameters. The loss factor is shown by its logarithm to the base 10, and its dimension is 1/m. Regions controlled by viscous, laminar and turbulent loss are indicated. The turbulent and viscous losses are additionally dependent on flow and frequency, respectively, as shown by the sets of trajectories.

Fig 2.2
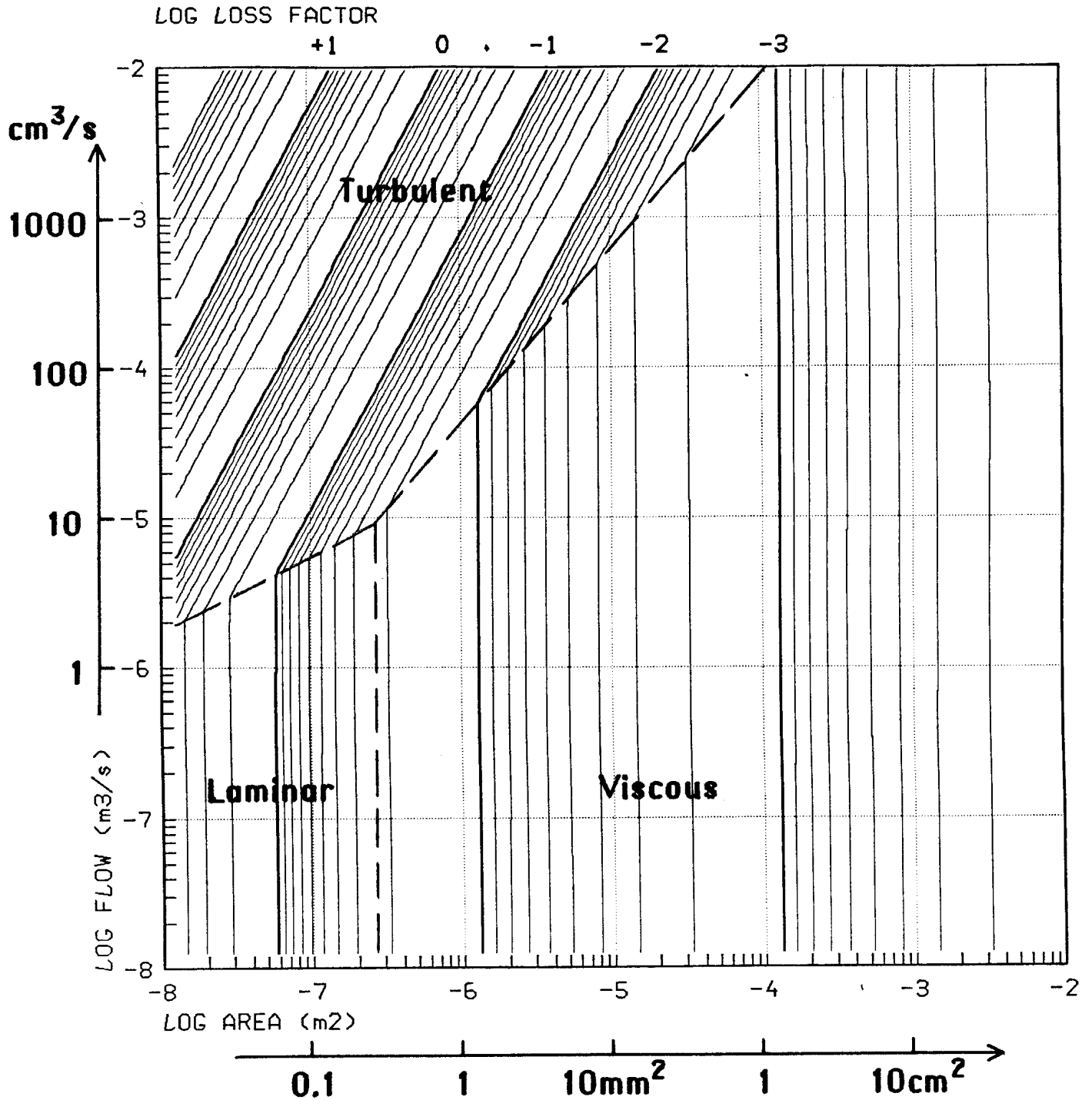
Fig 2.3. Contours of equal series loss factor per unit length, R/2Z, as function of area A and flow U, assuming 1 kHz for the viscous loss. Regions where laminar, turbulent, and viscous losses dominate are indicated. Scales show logarithms to the base of 10. This figure is a different projection of the same information as in fig 2.2.

Fig 2.3

air into the lower pressure side instead of symmetrical air movements at low and high pressure sides of the orifice. Additionally this specific resistance was found to be the same when the sound pressure was lowered and superimposed on a continuous flow of comparable velocity v.

This specific resistance on the area A will give the acoustic resistance $\rho * v/A$, and dividing by twice the characteristic line impedance let us form it into our loss factor. This does indeed give eq (2.12), let us call it the 'jet loss'. With the numerical value of c inserted it becomes

$$D_{jet} * l = 1.429 * 10^{-3} * U/A \qquad (dimensionless) \qquad (2.14)$$

Using the jet loss derived for glottis or a sharp constriction will need further justification when applied to the vocal tract line since this will mostly have rather gradual constrictions. When air is forced through a passage with smooth narrowing and expansion the kinetic energy may be recovered as static energy on the expansion side, and then we have the Bernoulli underpressure in the passage, but no physical loss of energy. If however the air stream releases from the wall in the expansion phase and forms a jet, if we get 'flow separation', then much of the wave energy will indeed be lost. Part of it will be locally converted into heat when the jet turbulence is damped out by the viscosity, part of it will radiate away in the form of noise.

The question is then, does the air release from the wall or not? It is pointed out in fluid dynamics handbooks (e g Goldstein (1938, 1965), p57) that this can not be resolved from the flow and the wall shape only. It is also necessary to know whether the downstream pressure field has a gradient capable of balancing the centrifugal force that will otherwise cause a release. To resolve this in detail would be tremendously complex and will not be attempted.

Instead, as a rough quantitative example, consider an air stream of velocity $v = U/A_1$ following a curved surface of radius r, see fig 2.4. The centrifugal force will then give a radial pressure gradient $dp/dr = \rho * v^2/r$. The air enters a section of area A_2 having a standing wave of frequency f and the pressure could then approximately be $p(x) = p_0 * \cos(2 * \pi * f * x/c + \varphi)$ where φ is some unknown phase and $p_0 = U/Z_2 = U * \rho * c/A_2$ is the pressure from the same flow driving the impedance of the larger area section. Differentiating $p(x)$ to give an axial pressure gradient and equating its amplitude to the radial pressure gradient would then give a rough idea of a critical radius $r = U * A_2/(2 * \pi * f * A_1^2)$. With $A_1 = .1$ cm^2, $A_2 = 5$ cm^2, $U = .25$ l/s, $f = 1$ kHz we get $r = 20$ cm. This rather considerable value should then indicate the smallest r where a separation can still be prevented should all conditions like phase and gradient directions be maximally favourable. Since r comes out so large it seems safe to assume that jet formation should be the normal state of affairs at almost any significant constriction at flows representative for speech. Further support may be that with this .25 l/s DC

flow the critical particle velocity 1 m/s is reached already with an area as large as 2.5 cm².

It is important to remember that this loss is not distributed, but comes from a discrete resistor at the appropriate place. It can be argued whether this place is at the entry or exit of the constriction, but this is probably not very distinctive. The pressure drop occurs at the entry but the actual loss is at the exit where the kinetic energy fails to be reconverted into static.
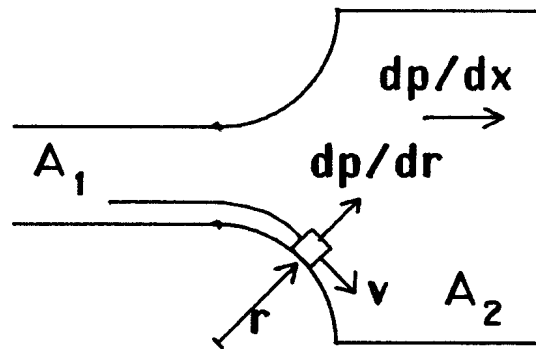
Fig 2.4. Simplified conditions for flow separation.

Fig 2.4

LOSSES   2 - 6

2.2 SHUNT LOSSES

We now consider losses that should be represented by a properly connected shunt element, this because the loss mechanism is sensitive to the pressure in the line rather than the flow. In section 1 I used E for the shunt loss factor to distinguish it from the series loss factor D. In the following I prefer to equivalently use D only, but augmented with a subscript for the mechanism involved. This is natural since the series and shunt losses appear as additive terms in the total 'attenuation constant' of fig 2.1. Also remember we work now with a distributed loss factor, i e loss per unit length. Before use in the scattering equations the loss factors must be multiplied by the appropriate line segment length.

2.21 HEAT CONDUCTION LOSS

The temperature variations induced by the sound pressure are to some extent lost by conduction into the walls. Ingard (1953) gives the specific surface conductance as (cf Fant (1969), p 33)

$$G_{o,heat} = \frac{\eta-1}{\rho*c^2} * (\frac{\lambda*\omega}{2*\rho*C_p})^{1/2} \tag{2.15}$$

where η is the adiabatic constant, λ is the thermal conductivity, and C_p is the heat capacitivity. To obtain the loss conductance G per unit length the specific conductance is to be multiplied by the circumference of the tube section. For a circular section this is $\pi*d=2*(\pi*A)^{1/2}$. Then we can assemble an expression for the heat conduction loss factor as

$$D_{heat} = GZ/2 = \frac{\pi*(\eta-1)}{c} * (\frac{\lambda}{\rho*C_p} * \frac{f}{A})^{1/2} \tag{2.16}$$

Inserting numerical values of the constants gives

$$D_{heat} = 1.610*10^{-5} * (f/A)^{1/2} \tag{2.17}$$

which also matches the observation by Fant (1960) that it is smaller than the viscous loss factor by a constant factor of 2.2.

2.22 LOSSES INTO THE VOCAL TRACT WALL

Next we look at the direct radiation from the air into the tube wall. The wall has the specific wave impedance ρ_w*c_w using the density and the velocity of sound in the wall material. The specific conductance is the inverted value of this, and multiplying by the tube circumference gives

the loss conductance per unit length. From that the loss factor for the wall absorption will be

$$D_{wab} = (\pi/A)^{1/2} * (\rho*c)/(\rho_w*c_w)$$

(2.18)

This assumes infinite thickness of the walls. The effect is a companion to, but should not be confused with the effect of thin wall vibrations treated in the next paragraph. Assuming that the wall material has essentially the same properties as body temperature water of 1% salinity, $\rho_w = 1000$ kg/m^3, $c_w = 1500$ m/s, will give the numerical value

$$D_{wab} = 4.71*10^{-4} * A^{-1/2}$$

(2.19)

It is not known to me that this loss term is mentioned in the literature. The term absorption loss will have to be taken with some reservation. Much of the power absorbed may well be reradiated at the outside of the wall.

2.23 LOSS FROM WALL VIBRATIONS

The losses caused by sound induced vibrations in the vocal cavity walls are widely recognized though quantitative data on the wall are relatively scarce. For the audio frequency loss evaluation let us use some resonable values, namely a specific surface resistance $R_o = 12*10^3$ Ns/m^3 in series with a specific surface inductance $L_o = 20$ kg/m^2 (which implies a wall thickness around 2 cm). The figures represent a gross average of data from Ishizaka et al (1975) and Fant et al (1976), further reviewed in table 4.1.

For audio frequencies the inductive part dominates widely over the resistive in the series equivalent. To have the loss represented by a pure shunt conductance we can make the transformation $G_o = R_o/(\omega*L_o)^2$. To give the conductance per unit length it is multiplied with the tube circumference. Assuming a circular cross section with circumference $\pi*d$ we get $D_{wal} = \pi*d*G_o*\rho*c/(2*A)$ which is rearranged into

$$D_{wal} = \frac{R_o*\rho*c}{4*L_o^2*\pi^{3/2}} * \frac{1}{f^2 * A^{1/2}}$$

(2.20)

and with numerical values inserted

$$D_{wal} = 537 * f^{-2} * A^{-1/2}$$

(2.21)

Comparing this to the wall absorption we find them to be equal, with the wall material assumptions done, at the frequency 1070 Hz. Above that frequency then the absorption formula for D_{wab} should be used instead of D_{wal}.

Fig 2.5 shows the dominant shunt loss factors as functions of frequency and area. It is seen that the wall vibration loss dominates at low frequencies and heat conduction loss at high frequencies. The transition is in the neighbourhood of 1 kHz. The wall absorption is always smaller, it is just below the others at the minimum around 1 kHz and thus will contribute only marginally, if at all.

It can be argued that the vibrating wall surface is independent of articulation. For that case the effective circumference in establishing the loss conductance is constant. Assuming S=10 cm would then give the alternate vibration loss formula

$$D_{was} = 15.2 * f^{-2} * A^{-1} \tag{2.22}$$

This would give a different slope of the contours of equal loss of fig 2.5, but not a too big difference in values at larger areas. The wall loss model of section 4 assumes the total wall surface of the vocal tract to be constant.

2.3 LOSS FROM LIP RADIATION

To a first approximation the termination at the lips can be represented with a classical series R-L combination representing one side of a baffled circular piston. For the resistance we take the notation used by Fant (1960) p 35

$$R_{rad} = \rho * \omega^2 * K_s / (4 * \pi * c) \tag{2.22}$$

K_s is a frequency dependent radiation resistance factor that accounts for the more complicated effects from baffling. K_s has a relatively narrow range of variation from 1 at zero frequency to 2 with an infinite baffle. With large areas and high frequencies K_s will drop below 1, but the value 1.4 can represent a gross average for normal conditions.

Combining this radiation resistance with the characteristic line impedance Z=ρ*c/A for area A gives a radiation loss factor

$$D_{rad} * l = R/2Z = \pi * f^2 * A * K_s / 2 * c^2 \tag{2.23}$$

and numerically, using $K_s = 1.4$

$$D_{rad} * l = 1.795 * 10^{-5} * f^2 * A \tag{2.24}$$

Like the jet loss it is a discrete loss, not a distributed parameter, and for that reason we put it as D*l instead of D to keep the dimension correct.
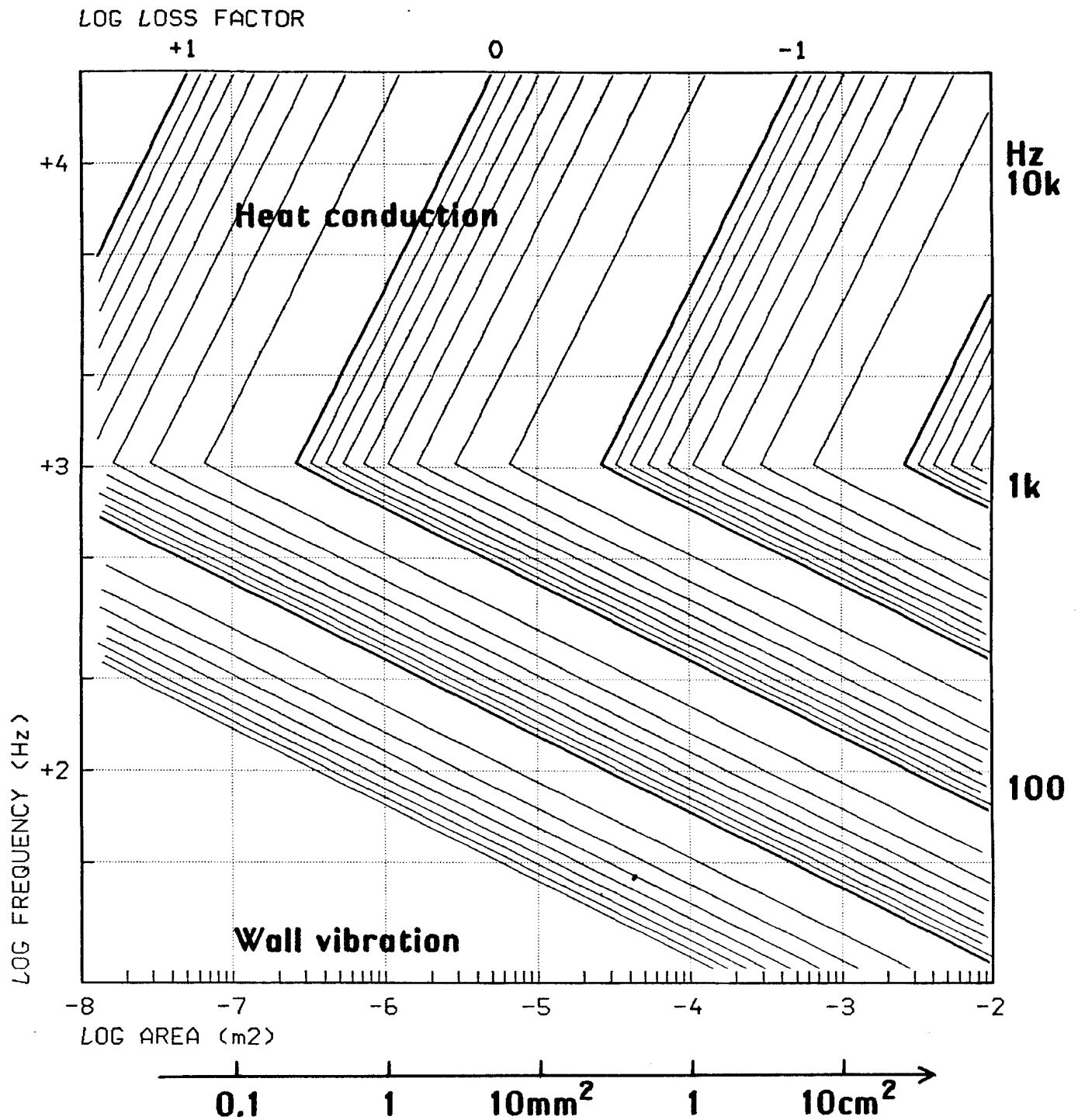
LOG LOSS FACTOR

Fig 2.5. Contours of equal shunt loss factor per unit length,
GZ/2, as function of area and frequency. Regions where wall
vibration and heat conduction losses dominate are indicated.
Scales show logarithms to the base of 10.

Fig 2.5

The figure here, and also the coming one for the glottis, must not be used for more than indications that they pertain to important effects. Just because of this there is no doubt it has to be incorporated into the model, and that the modelling has to be done with greater care than in the simple consideration above.

2.4 LOSS IN THE GLOTTIS

The glottal passage will cause substantial losses to the waves in the vocal tract. The closer analysis of this belongs in a treatment of the excitation rather than the transmission line, but it would be good to have a rough indication of the magnitude of an equivalent terminating shunt resistance. For the present let us just neglect the complications from time variation and assume a transglottal pressure of $p_g = 1$ kPa and a mean glottal flow $U_g = .25$ l/s. The mean flow resistance is then $p_g/U_g = 4*10^6$ Ns/m5 and, as discussed under the jet loss, the differential resistance twice this value. Now regard this resistance as a shunt conductance across the ideally open-circuit line of impedance Z and area A to give a loss factor

$$D_{glt}*l = ZG/2 = \rho*c/(A*2*2*p_g/U_g) \tag{2.25}$$

Inserting $p_g = 10$ cm $H_2O = 1$ kPa, $U_g = .25$ l/s $= .25*10^{-3}$ m3/s, and $A = 8$ cm2 $= 8*10^{-4}$ m2 renders a reference value of

$$D_{glt}*l = 0.031$$

In itself it is interesting to note that the glottal damping decreases with transglottal pressure and increases with glottal flow.

The glottis loss conductance will have a great influence because it is in principle always located at a pressure maximum for all formant frequencies, even if the relative magnitudes of these maxima are strongly dependent on the particular area function.

2.5 COMPARISON OF LOSSES

For the comparison we rewrite the loss factors into logarithmic form to make the numbers less awkward to compare. To care for the distributed losses we include a line segment length l.

$$\log(D*l) = \log(D_0) + X_a*\log(A) + X_u*\log(U) + X_f*\log(f) + X_l*\log(l)$$
(2.26)

X_a, X_u, and X_f are then the exponents for area, flow, and frequency in the different formulae. X_l is just a binary switch to indicate the distributed losses. Values for all the loss mechanisms considered are compiled in table 2.1.

The log(Dl) column shows results assuming the following arbitrary reference values

$$A = 8 \text{ cm}^2 = 8*10^{-4} \text{ m}^2 \qquad \log(A) = -3.097$$
$$U = 0.5 \text{ l/s} = 5*10^{-4} \text{ m}^3/s \qquad \log(U) = -3.301$$
$$f = 1000 \text{ Hz} \qquad \log(f) = +3.000$$
$$l = 17.5 \text{ cm} = 1.75*10^{-1} \text{ m} \qquad \log(l) = -0.757$$

The final column Dc/π is a measure of the contribution to formant bandwidth for each of the loss factors, assuming the same uniform tube. Of course this tube will have no formant at that particular frequency, but the numbers are useful as a basis for scaling as shown in fig 2.7. In this the figures of the Dc/π column marked with * have been multiplied by a factor of 2. For the discrete losses this is to account for their action at pressure and flow maxima where the RMS value is twice the space average. For the surface dependent distributed losses it is to account for a area shape different from a circular shape. Furthermore the dimensionless loss factors for the discrete losses have been divided by the length to give figures with comparable meaning.

The uniform tube can serve as a general reference, but the influence of many of the loss factors is clerly negligible for this case. But in most speech sounds the area will have minima much smaller than the average, and with correspondigly higher losses. The influence of these losses on the formant bandwidths is critically dependent on the place of the loss in relation to the standing wave pattern for that particular formant with that particular area function. Thus the series losses will be null at the flow nodes as will the shunt losses at the pressure nodes. For practical purposes however the bandwidth contributions in the uniform tube as in fig 2.6 and 2.7 can be regarded as minimum values. The exception is the bandwidth contribution from radiation loss which is area dependent in a way that is inverse to what are all the other loss factors.

Table 2.1. Logarithms of loss factors and exponents of dependences of area, flow, frequency, and length, all in SI units.

Discrete:	$\log(D_0 * l)$	X_a	X_u	X_f	X_l	$\log(DI)$	Dc/π
	<0>	<m²>	<m³/s>	<Hz>		<0>	<Hz>
Lip rad	−4.746	+1	.	+2	.	−1.84	9.2 *
Glottis	−1.5	.	.	.	.	−1.5	20 *
Jet	−2.845	−1	+1	.	.	−3.05	0.6

Distributed:	$\log(D_0)$	X_a	X_u	X_f	X_l	$\log(DI)$	Dc/π
	</m>	<m²>	<m³/s>	<Hz>	<m>	<0>	<Hz>
Wall vibr	+2.730	−0.5	.	−2	1	−2.48	2.1 *
Wall abs	−3.327	−0.5	.	.	1	−2.54	1.8 *
Heat cond	−4.793	−0.5	.	+0.5	1	−2.50	2.0 *
Turbulent	−4.909	−1.375	+0.75	.	1	−3.87	.1
Laminar	−6.232	−1	.	.	1	−3.89	.1 *
Viscous	−4.441	−0.5	.	+0.5	1	−2.15	4.5 *

To find the resulting formant bandwidths analytically from all the loss components and the area function is a rather complex operation. One of the main attractions of a lossy line model is of course that this is performed implicitly and the synthesized signals will get proper formant bandwidths.

Figs 2.6 and 2.7 is an attempt to give a general overview of the different loss effects. It shows the logarithms of dominating loss factors as functions of area and frequency, for negligible flow (that is, less than 0.01 l/s) and for a flow of .5 l/s. To make it possible to compare the distributed and discrete losses the distributed ones are computed for the reference length 17.5 cm, as in the DI column of table 2.1. For this reason the magnitude indications differ from those of figures 2.3 and 2.5. We must also keep in mind that the diagram is valid only for uniform tubes. It is however interesting to see that the general loss minimum is walled in at typical F1, F2 frequencies and typical areas. One could speculate on nature having optimized the absolute size of the speech organs.

Tab 2.1   LOSSES 2 − 12

Fig 7.8 may show all these loss terms are not vain. It is a diagnostic printout from a simulation in a dynamically controlled line, showing the spatial distribution of series losses as functions of time. A coarse gray scale shows magnitudes of the loss factors D*l with one step for each power of ten. The black level is a maximum where the loss resistance approximates the line impedance so that Dl~1. The point of maximal constriction in the vocal tract is easy to follow. Plots like this support my view that the jet loss is worthwhile to include in a model, but the turbulent loss is more doubtful. There is no doubt of the importance of viscous and laminar losses.
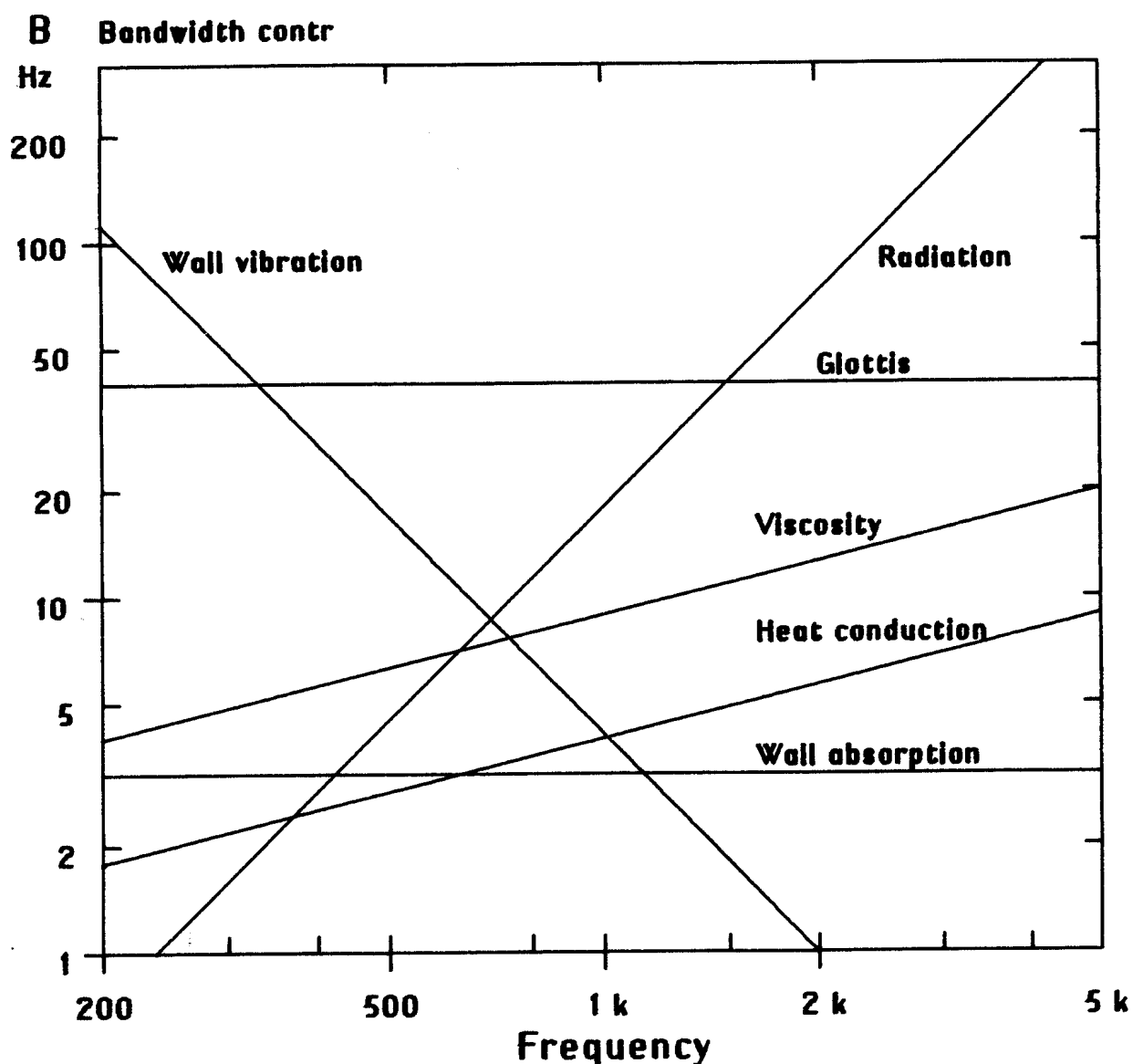
Fig 2.7. Contributions to formant bandwidths as functions of frequency for a uniform 8 cm2 tube of length 17.5 cm, shape factor 2. This diagram represents a section of fig 2.6 at area 8 cm2.

Fig 2.7
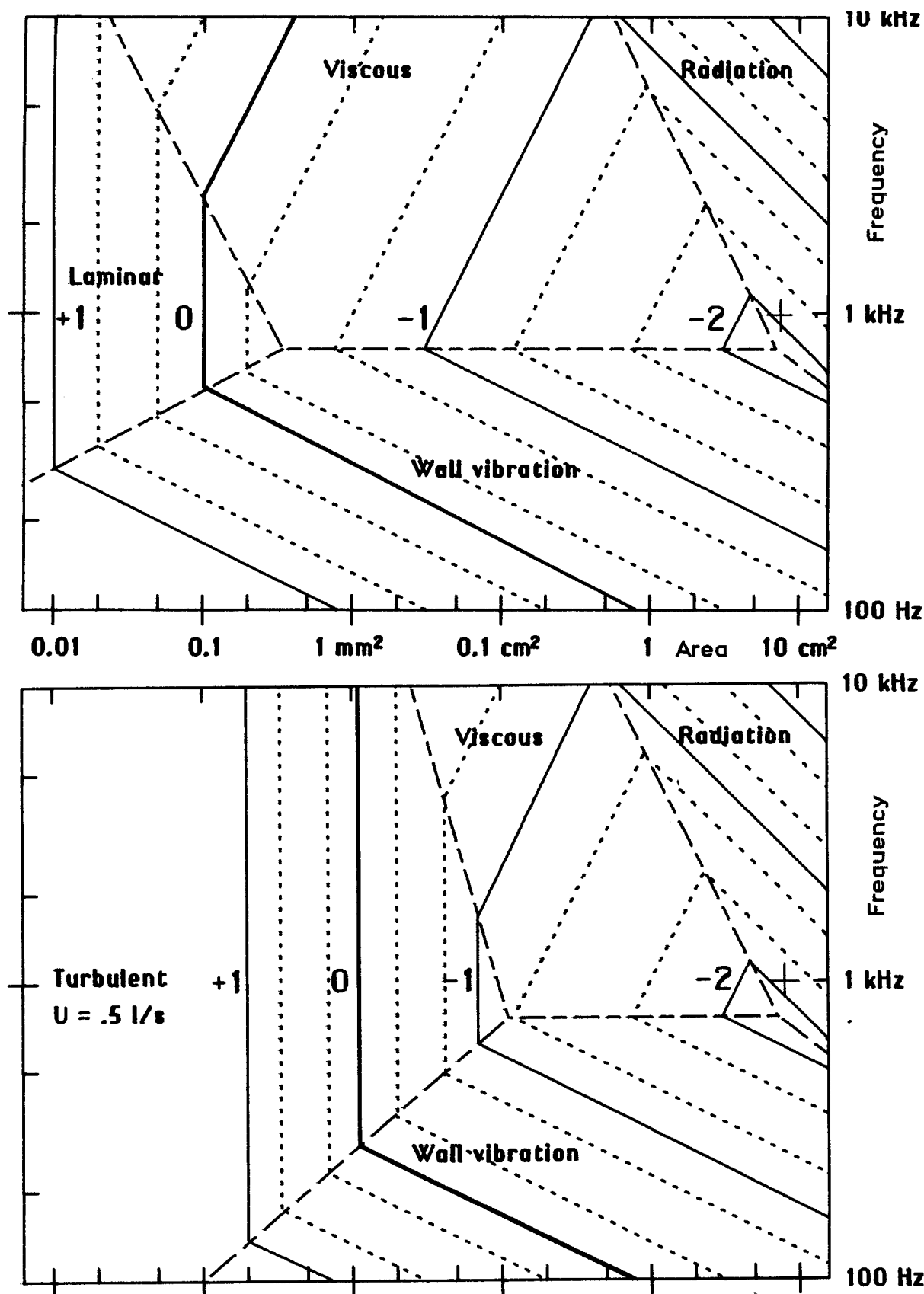
Fig 2.6. Comparison of different loss mechanisms as functions of area, frequency and flow, assuming a uniform tube of length 17.5 cm and with shape factor 2. The reference point of the DI column in table 2.1 is marked. Regions where different loss mechanisms dominate are indicated and can be seen as a set of intersecting oblique planes. The minimum near 10 cm^2 and 1 kHz is additionally covered by losses from glottis. Contours show equal values of loss factor times length in logarithms to the base 10. They are also contours of equal formant bandwidth, the -1 contour corresponds to B=127 Hz, and -2 to B=13 Hz.

Top: low DC flow, less than .01 l/s. Bottom: DC flow .5 l/s.

Fig 2.6
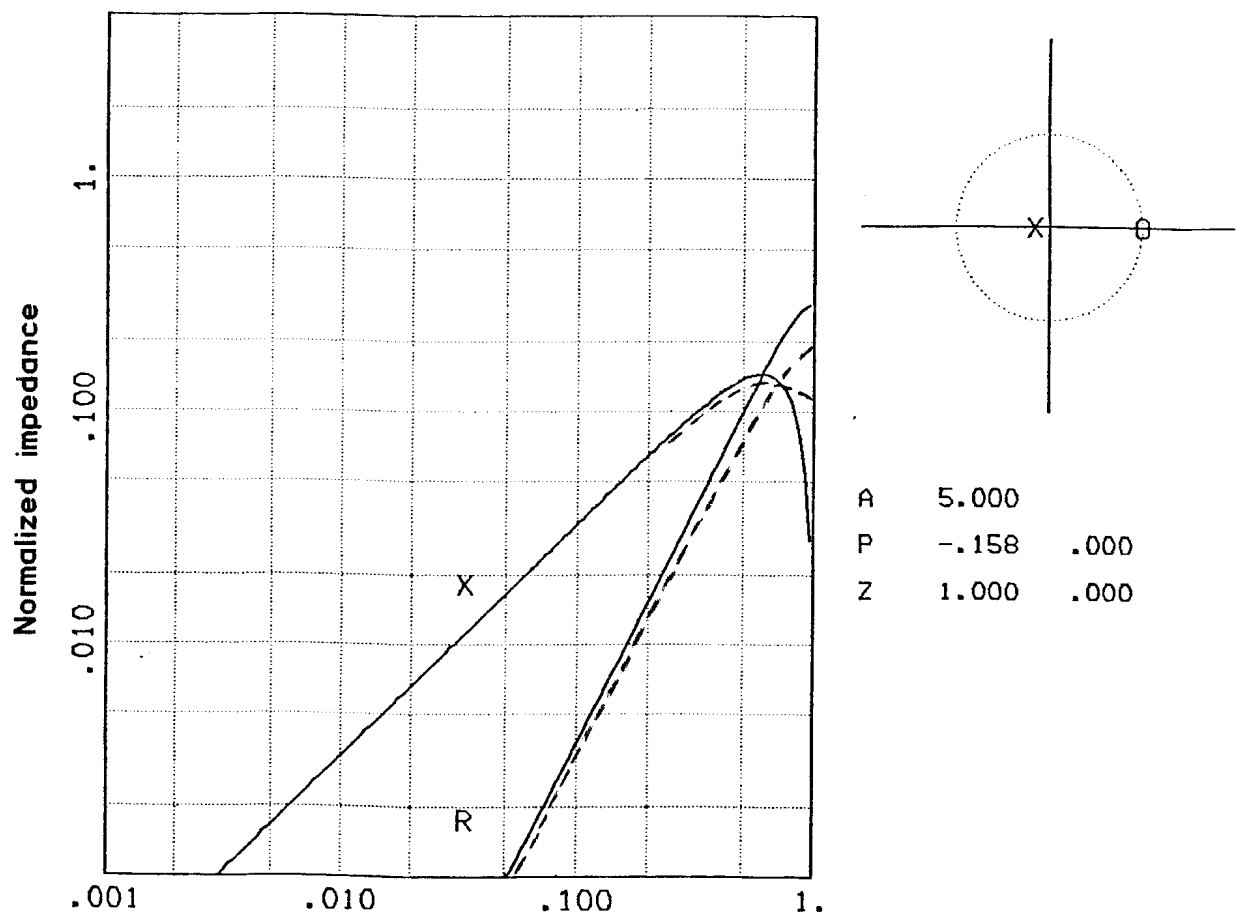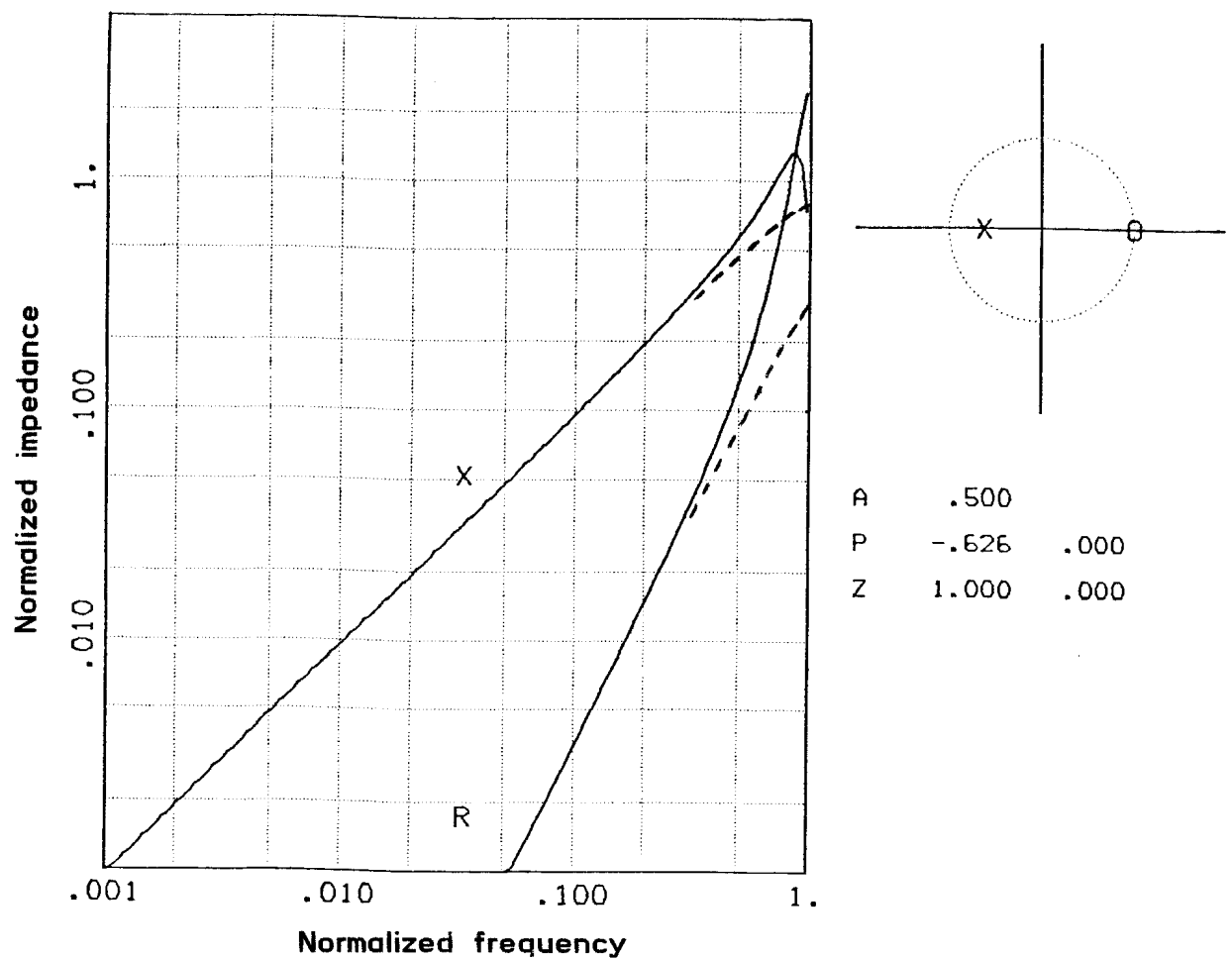
Fig 3.1. Normalized model of radiation impedance, $Z_{rad3}*A/\rho ca$, in z domain and as resistance and reactance vs normalized frequency, $\omega\tau/\pi$. Theoretical values for a piston in sphere, assuming 16 kHz sampling is shown with dashed lines. Diagrams apply to areas .5 (top) and 5 cm2.

Fig 3.1

3. LIP RADIATION MODEL

z domain lip radiation impedance model of Laine. Interfacing the model to the reflection line analog.

The acoustic impedance of the lip opening can be represented by a resistance R_{rad} and a reactance X_{rad} as

$$R_{rad} = \rho * \omega^2 * K_s / (4 * \pi * c)$$
$$X_{rad} = 4 * \rho * \omega * L_s / (3 * \pi^2 * r) \tag{3.1}$$

(cf section 2.3) where r is the radius of the opening. K_s is the radiation resistance factor of Fant (1960) and L_s a corresponding reactance factor. We repeat the general observations that at low frequencies the acoustic resistance is independent of the area but proportional to the square of frequency. The reactance is that of an inductance, dependent of the area. Usually this is put in terms of an end correction, the geometrical tube length is apparently incremented by a certain fraction of its radius. The effects of baffling and the deviating behaviour at higher frequencies is accounted for by the correction factors K_s and L_s. In the frequency and area range of interest they take values not far from unity. The underlying theory, see for instance Morse and Ingard (1968), renders very complex expressions for these correction factors, including Bessel functions.

Laine (1982) exploited that the radiation impedance of a piston in a sphere could be well approximated by sine and cosine functions and proceded to suggest a number of models of the radiation impedance in the z domain. The simplest of these is

$$Z_{rad1} = \rho * c * C1 * (z-1)/z = \rho * c * C1 * (1-z^{-1}) \qquad Ns/m^5 \tag{3.2}$$

where C1 is a parameter to be optimized in relation to the sampling frequency. At 16 kHz he gives it as 633 $/m^2$.

In the frequency domain this corresponds to

$$Z_{rad1} = \rho * c * C1 * (1 - \cos\omega\tau + j * \sin\omega\tau) \tag{3.3}$$

This model handles the radiation resistance well, but a shortcoming is that, since the area does not enter the formula, the variation of the end correction with the area is not accounted for. Instead the reactance will be that of a fixed inductance.

To have one more degree of freedom we look at Laine's pole-zero model

$$Z_{rad3} = \frac{\rho c}{A} * a * \frac{z-1}{z-b} = \frac{\rho c}{A} * a * \frac{1-z^{-1}}{1-b*z^{-1}} \qquad (3.4)$$

By mean square error matching to the impedance of a piston baffled in a sphere Laine found expressions for the coefficients

$$a = .0779 + .2373*\sqrt{A} \qquad (3.5a)$$
$$b = -.8430 + .3062*\sqrt{A} \qquad (3.5b)$$

with A in cm^2, valid for areas in the range A=.5 to 6 cm^2, and the radius of the sphere being 9 cm. The coefficient 'b' determines the location of the z domain pole along the real axis and gives a means to influence the ratio of resistance to reactance. The coefficient 'a' operates as a magnitude scale factor. For the special case with b=0 the singularities coincide with those of the simpler model of (3.2).

For reference the Z_{rad3} model is shown graphically in fig 3.1 for two different areas. The plot shows Z resolved into real and imaginary parts as a function of normalized frequency, that is, following the unit circle in the z domain. The log scales conceal the fact that the resistance and reactance basically have sinusoidal shapes.

To incorporate this z domain impedance into the model I form a reflection coefficient at the lip end. That is where the last tube with area A(N)=A meets the radiation impedance. Using (1.11), (1.15), and (3.4) the reflection coefficient will be

$$k = (Z_{rad3} - \rho*c/A) / (Z_{rad3} + \rho*c/A) =$$
$$= ((a-1)*z - (a-b)) / ((a+1)*z - (a+b)) \qquad (3.6)$$

Unlike the conventional reflection coefficients we have had so far this one is now a function incorporating the delay operator $z^{-1}=1/z$. It is also a function of the area A, indirectly from its influence on the a and b coefficients.

Here at the end joint of the line analog the scattering equations (1.23) luckily will simplify, because we have no backward wave s(N+1). In other words, what has been radiated away from the line end as r(N+1) is lost forever as far as the line is concerned.

Inserting (3.6) into (1.23) with s(N+1) = 0 after some manipulation gives the flow radiated from the lips. To simplify the notation, use subscript 1 for section N and 2 for section N+1 and get

$$r_2 = (1-k)*r_1 \quad \text{giving}$$
$$r_2 = ((a+b)*r_2*z^{-1} + 2*r_1 - 2b*r_1*z^{-1}) \,/\, (a+1) \qquad (3.7a)$$

Similarly with (1.23) the flow reflected back into the line is

$$s_1 = k*r_1 \quad \text{giving}$$
$$s_1 = ((a+b)*s_1*z^{-1} + (a-1)*r_1 + (b-a)*r_1*z^{-1}) \,/\, (a+1)$$
$$\qquad (3.7b)$$

These equations replace the usual scattering equations here at the end of the tube array. The input varible r_1 is the result of the scattering computation at the joint between the two last tubes. Special extra storage is required to save the delayed input r_1*z^{-1} from the previous sample time. r_2 and s_1 are not changed by the normal scattering equations so they can be left in place and used in the next time slot as delayed values.

This way now a frequency dependent element using the z transform technique has been interfaced to cooperate with the reflection model. In doing this kind of operation we use the sample interval τ and the corresponding delay operator z^{-1}. It is wise not to confuse this with the travel time through one tube section which is only one half of this time.

The radiation impedance has several important effects in the model. It will give the appropriate frequency and magnitude shaping of the exit flow wave. Through its reactive part it will account for the lip area dependent end correction that in turn will affect the formant frequencies. Its resistive part will influence the formant bandwidths. Finally it determines the reference point of zero pressure because of the missing backward wave from outside space.

A weakness in the line analog with a constant number of equal length sections is that it cannot simulate a varying total length other than by varying the sampling frequency, something that is rather intricate. In connection with the end correction, the inductive part of the radiation impedance, there is however an additional possibility of minor adjustments of apparent total length.

τ/Ke²   **2Ke**   **Kτ**

-2K
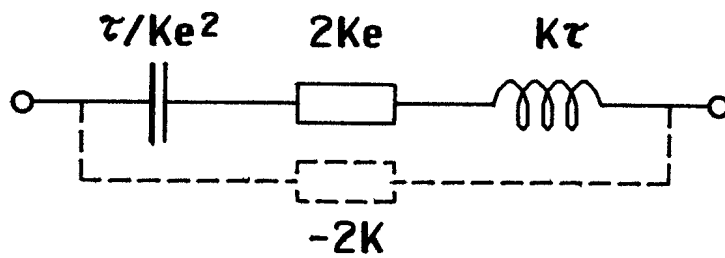
Fig 4.1. Wall impedance z domain model resolved into discrete elements.

Pole   **1.00**
Dual zero 0.98

-X   X   -R

R

Normalized impedance

.001   .01   .1   1

Normalized frequency
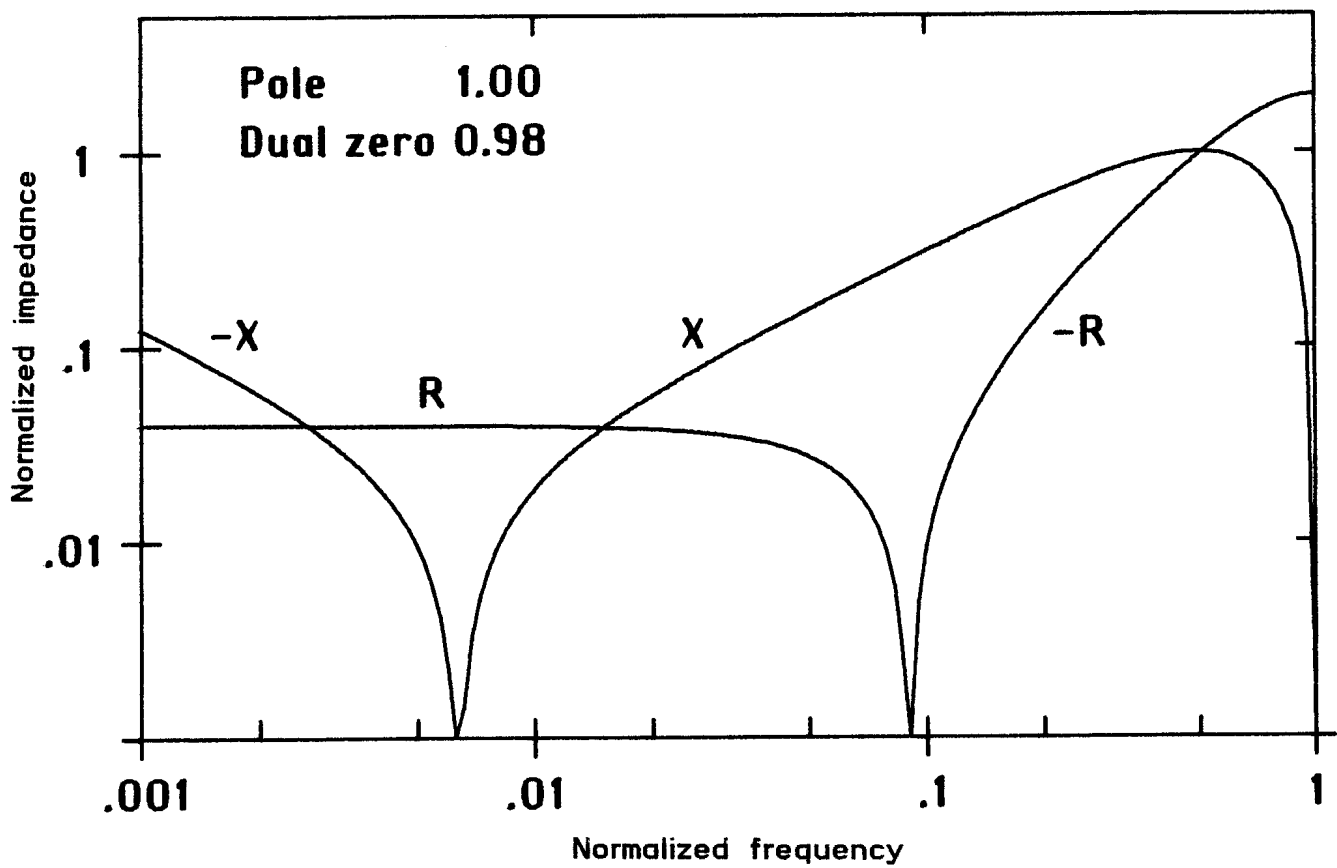
Fig 4.2. Resistance and reactance of the normalized wall impedance model, Z_W/K, as function of normalized frequency,

Fig 4.1, 4.2

4. WALL INTERACTION MODEL

*z domain modelling of the wall impedance and its installation
into the line. Discussion of quantitative data.*

The vocal tract walls are displaced by the internal waves. One consequence pertinent to audio frequencies will be that the resonance frequencies of the tract are elevated. This is related to the comparatively large acoustic mass of the walls effectively shunting the mass of the air in the tract. Another aspect is that the resistive part of the wall impedance will introduce a loss of energy. Since the walls are mass controlled in the audio range the resistance will have its main influence at low frequencies as we saw with the wall vibration loss. The flow into the walls is partly absorbed in the tissue, partly reradiated through the skin into the free air surrounding the speaker.

This study neglects pumping from area variations, but it comes natural to include a related thing in the wall model, its function as an air reservoir in case of a downstream closure: under static pressure the wall compliance will store a certain volume of air. At very low frequencies the wall impedance model should then be capacitive.

The most accurate would be to install one wall element at each tube. Fair approximations can be expected with a smaller number, perhaps one in the mouth region for the cheeks and one at the glottis end, or even the latter one alone. The motive is that the influence is restricted to low frequencies where the pressure will be relatively homogeneously distributed along the line. Generally, there will always be a pressure maximum at the glottis and a node at the mouth opening. The greater effect from a shunt element is thus to be expected at the glottis end. However the combination of a middle constriction with small lip opening may justify also a front shunt element for the cheeks.

From these points it is reasonable to discuss a series resonant circuit as a special element to connect as a shunt to the line analog. The inductor and the resistor will then represent the wall mass and loss. The reactance of the capacitor for the wall compliance is generally negligible at audio frequencies, but is there to model the pseudo-static air reservoir.

Let us select a normalized resonance frequency

$$e = \omega_w * \tau \qquad \text{(dimensionless)} \qquad (4.1)$$

where ω_w is the angular resonance frequency of the wall mass and compliance. This frequency is rather low, the order of 50 Hz.

In the z domain the resonance manifests as a conjugate pair of zeroes located on an arc with the small radius e around the point z=1. For simplicity I arbitrarily assume a Q value of 0.5 so that the zeroes coincide on the real axis at 1-e. For DC, that is z=1, the impedance is infinite and we have a pole. A z domain representation of the impedance can then be put as

$$Z_w = K * (z - (1-e))^2 / (z - 1) \qquad (4.2)$$

where K is an impedance magnitude scale factor (in Ns/m^5). Working backwards into the frequency domain let us see whatever elements this model will represent. At low frequencies we can approximate z on the unit circle as

$$z = cos\omega\tau + jsin\omega\tau \approx 1 + j\omega\tau \qquad (4.3)$$

which, inserted into (4.2) will give

$$Z_{w,low} = K * (j\omega\tau + e)^2 / j\omega\tau \qquad (4.4)$$

It is now simple to examine $Z_{w,low}$ for the three cases that $\omega\tau$ is much smaller, equal, and much greater than e. From that we identify the impedance elements as

$$C_w = \tau/K*e^2 \qquad\qquad 8*10^{-9} \ m^5/N \qquad (4.5)$$
$$R_w = 2*K*e \qquad\qquad .8*10^6 \ Ns/m^5$$
$$L_w = K*\tau \qquad\qquad 1.2*10^3 \ Ns^2/m^5$$

where some reasonable values for the entire vocal tract are also listed. At moderate and high frequencies neglect e in (4.2) to get

$$Z_{w,high} = K*(z-1) \qquad (4.6)$$

Fig 4.2 shows the frequency dependence of (4.2) resolved into resistance and reactance. At frequencies approaching half the sampling frequency the 'inductance' begins to misbehave, it shifts into a high negative resistance (this can be avoided by including another pole at z=-1 in the model). But here we are well above the resonance frequency, the impedance is high enough compared to the line impedance to be of no practical consequence to the model.

Now that the elements of the wall model are established as shown in fig 4.1 proceed to connect it close to the glottis end of the line model. The connection can be done in a similar way as we did with the radiation impedance, but using the shunt loss formula (1.93). But we know that the wall impedance is at all frequencies considerably higher than the line impedance, so computations can be simplified by using the following approximate procedure: The wall impedance is fed from the pressure in tube 1, which is

$$p_1 = (r_1 + s_1) * Z_1 \qquad\qquad (4.7)$$

and that pressure generator has the internal impedance Z_1. We can with no great error neglect this beside Z_w to obtain the flow into the wall as

$$U_w = p_1 / Z_w \qquad\qquad (4.8)$$

Inserting Z_w of (4.2) U_w is solved for as

$$U_w = (p_1 * z^{-1} - p_1 * z^{-2})/K + (1-e)*(2*U_w*z^{-1} - (1-e)*U_w*z^{-2})$$
$$(4.9)$$

needing some extra storage locations for the delayed quantities. The coefficients K and e are found from (4.1) and (4.5). The specific values with (4.5) will at 16 kHz sampling give $\omega_w/2\pi = 50$ Hz, e=0.02, and $K = 20 * 10^6$ Ns/m^5.

The number e is rather small since the wall resonance is much lower than the sampling frequency, and this makes necessary a fairly high accuracy in the coefficients for the delayed U_w. It is definitely not permitted to neglect the e^2 term in the coefficient for U_w*z^{-2}. If you do that the zeroes will separate and one of them cancels the pole at z=1, thus the resonator behaviour will degenerate into a first order system.

Now that U_w is known, we go back to the line and "steal" that amount of flow from r_1 and s_2 like the loss flow in (1.93). Since the wall flow is a small low frequency signal you can skip the detail of rationing with regard to k_{12} which anyway is small at the joint next to the glottis.

Tab 4.1. QUANTITATIVE DATA ON THE VOCAL TRACT WALL

Ishizaka, French, and Flanagan (1975) made direct measurements of the wall impedance and gave values of mass, resistance, and stiffness, separately for the relaxed and tensed cheek, and for the neck for a male subject. All these are given per unit area, so application of those data requires knowledge of the applicable wall area. The frequency f_0 of zero wall reactance was given.

Using hypothetical effective areas 80 cm^2 for each the neck and cheek I obtain from their data

	f_o Hz	L Ns^2/m^5	R Ns/m^5	$1/C$ N/m^5	Q
Cheek, relaxed	32	2.6 E3	1.0 E6	1.1 E8	.52
Cheek, tensed	60	1.9 E3	1.3 E6	2.7 E8	.55
Neck	72	3.0 E3	2.9 E6	6.1 E8	.47
leading to					
Total, relaxed	40	1.4 E3	.7 E6	.9 E8	.50
Total, tensed	66	1.1 E3	.9 E6	1.9 E8	.51

Q is computed here from f_0, L, and R. (The Ishizaka et al data on tensed cheek stiffness is probably victim of a proof error. Value must be $213*10^3$ instead of $33.3*10^3$ dyne/cm*cm^2 to be consistent with f_0.)

Fant, Nord, and Branderud (1976) showed the exterior wall vibrations to be concentrated around the mouth and on the neck. They estimated the total wall impedance by observing the resonance of the closed vocal tract. With the subjects operating the tongue it could be divided into the contributions from cheeks and neck. Their method does not give any measure of the wall compliance. From Fant et al I get the following values of L and R, recomputed into SI units

	f_x Hz	total L Ns^2/m^5	R Ns/m^5	cheek L1	R1	neck L2	R2
5 males	38	1.1E3	.53E6	3.8E3	.75E6	1.5E3	1.8E6
7 females	47	1.4E3	.83E6	3.2E3	1.9E6	2.4E3	1.5E6

where f_x is computed here as wall resonance frequency under the assumption of $Q=1/2$ with the given L and R.

As discussed by Wakita and Fant (1978) there is an unsettled discrepancy between these sets of data as to the relative distribution between mouth and pharynx of the wall inductance, but the agreement as a whole is acceptable.

Tab 4.1

4.1 TUBE BEND, HIGH FREQUENCY ABSORBER

In real spoken vowels you observe very little energy above, say 5 kHz, which has a number of partial explanations: The glottal source may be deficient at higher frequencies. Heat conduction and viscous losses are relatively high. The detailed shape of the area function near glottis (also including sinus piriformis) can have a pronounced effect on spectral balance. Effects of cross resonances and bends. The latter are in no way accounted for when modelling the tract as a one-dimensional line.

If the line is fed with an input representing the glottal flow it is easy to remove the high frequencies by pre-filtering them out from this excitation. But when glottis itself is made part of the line it not effective to lowpass filter the controlling glottis area time function. Due to the non-linearities, especially near the important closure of the glottis, high frequency components will be generated anyway. Furthermore, since we want to inject high frequency noise downstream the line to model fricative sounds, we should definitely not filter the output of the line to obtain the lowpass effect on vowels. The vowel lowpass effect is thus a strategic problem with the line analog. In classical formant synthesis the problem is handily avoided by using separate branches for vowel and fricative synthesis.
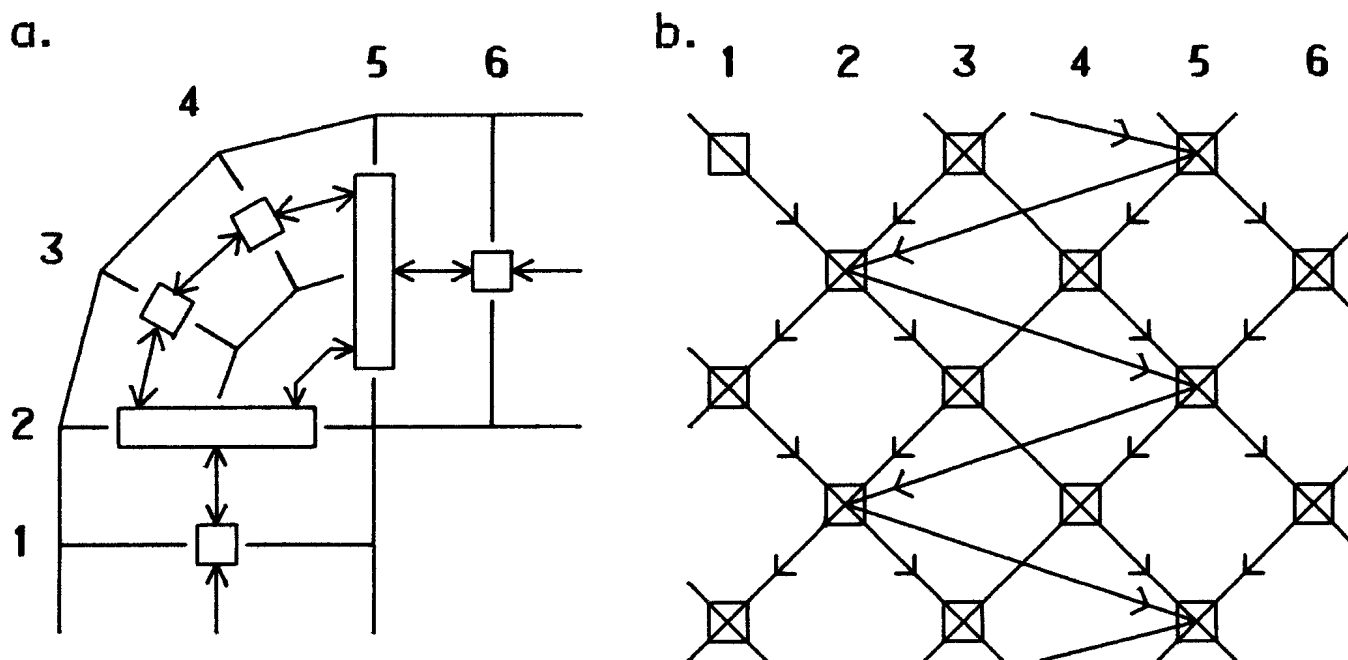

Fig 4.3. Simple approximation to a sharp bend in a line of abutting tube segments. Physical layout and partial wave flow diagram.

Fig 4.3   WALL INTERACTION   4 - 4

An attempt to simulate a bend was made with the following, admittedly rather crude method. Imagine an angle bend in the line of tube segments as in fig 4.3a. The wave leaving the last section before the bend can be seen as only partially entering the next section. The remainder takes a short cut around the corner and goes directly into the tube section two steps away. Disregarding all refinements like unequal areas at the bend this shortcut can be simulated as shown in the flow diagram of fig 4.3b. In essence then half of the appropriate waves are added to the regular incident waves two sections away, and only the remaining half is taken the regular way. It should be obvious from the flow diagram why the shortcut has to extend over two sections rather than one.

Fig 4.4b shows the effect of this trick on the response from a uniform tube. We do get high frequency attenuation. But also the formants are shifted towards higher frequencies. This is due to the shortcut diminishing the effective length of the tract model. Also the resonances get increasing bandwidths toward higher frequencies, the length of the resonator becomes poorly defined due to the shortcut. All are rather drastic effects that can upset the accuracy of the model. It is clear that this method is dangerous to apply and it has not been furhter pursued.

Another attempt to accentuate the high frequency level drop was to connect a high frequency absorber near the glottis end. It consisted of a shunting series resonant circuit in form of a z domain complex zero pair. It was similar to the wall vibration loss element, but having its resonance at 6.5 kHz with Q=5. Because the frequencies where it is active are high it was interfaced to the line using (1.93) and the basic method of section 1.10. It is critical to do this properly to avoid instabilities with effects contrary to those desired. The result for the uniform tube is seen in fig 4.4c.
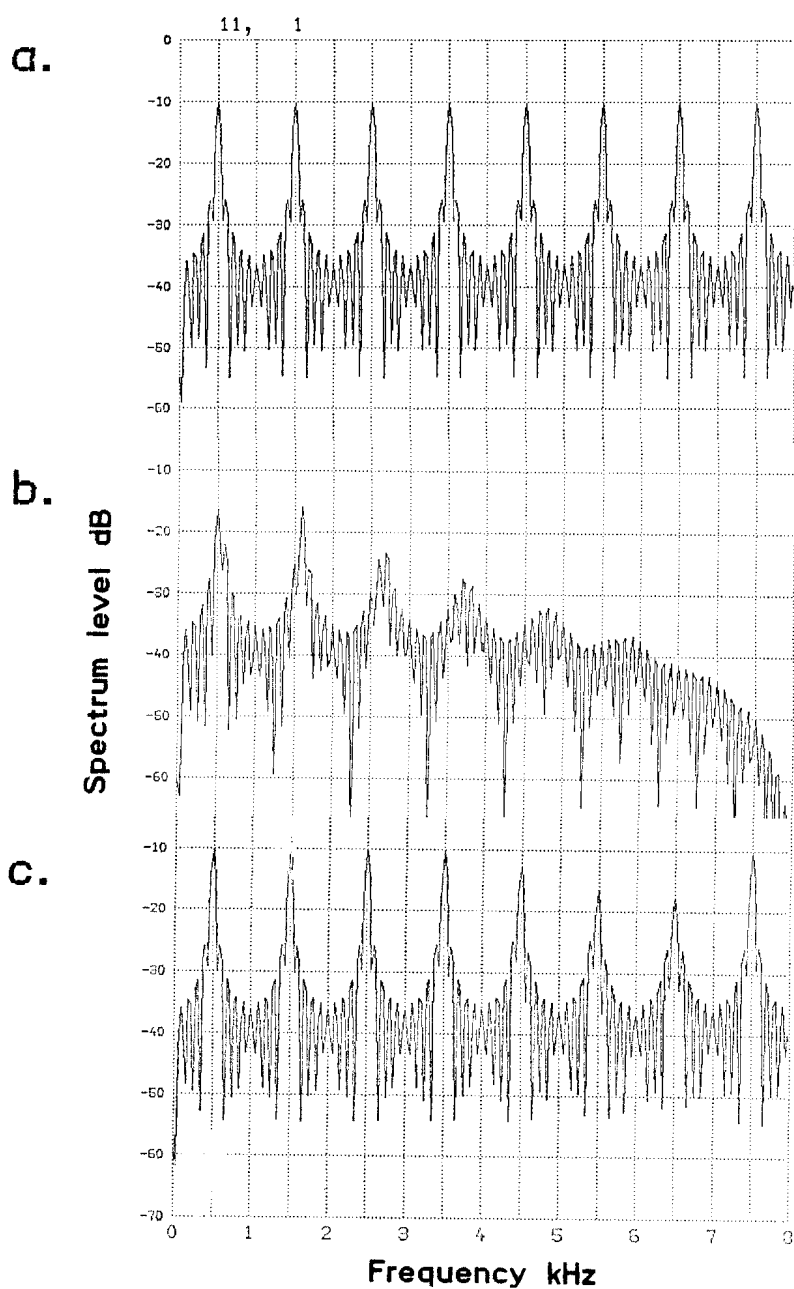
Fig 4.4. Spectra from attempts at vowel bandwidth limiting.

a. Reference response of periodic impulse excited uniform tube with ideal constant resistive losses.

b. Effect from simple simulation of a bend as in fig 4.3.

c. Effect from high frequency absorber near glottis.

Fig 4.4
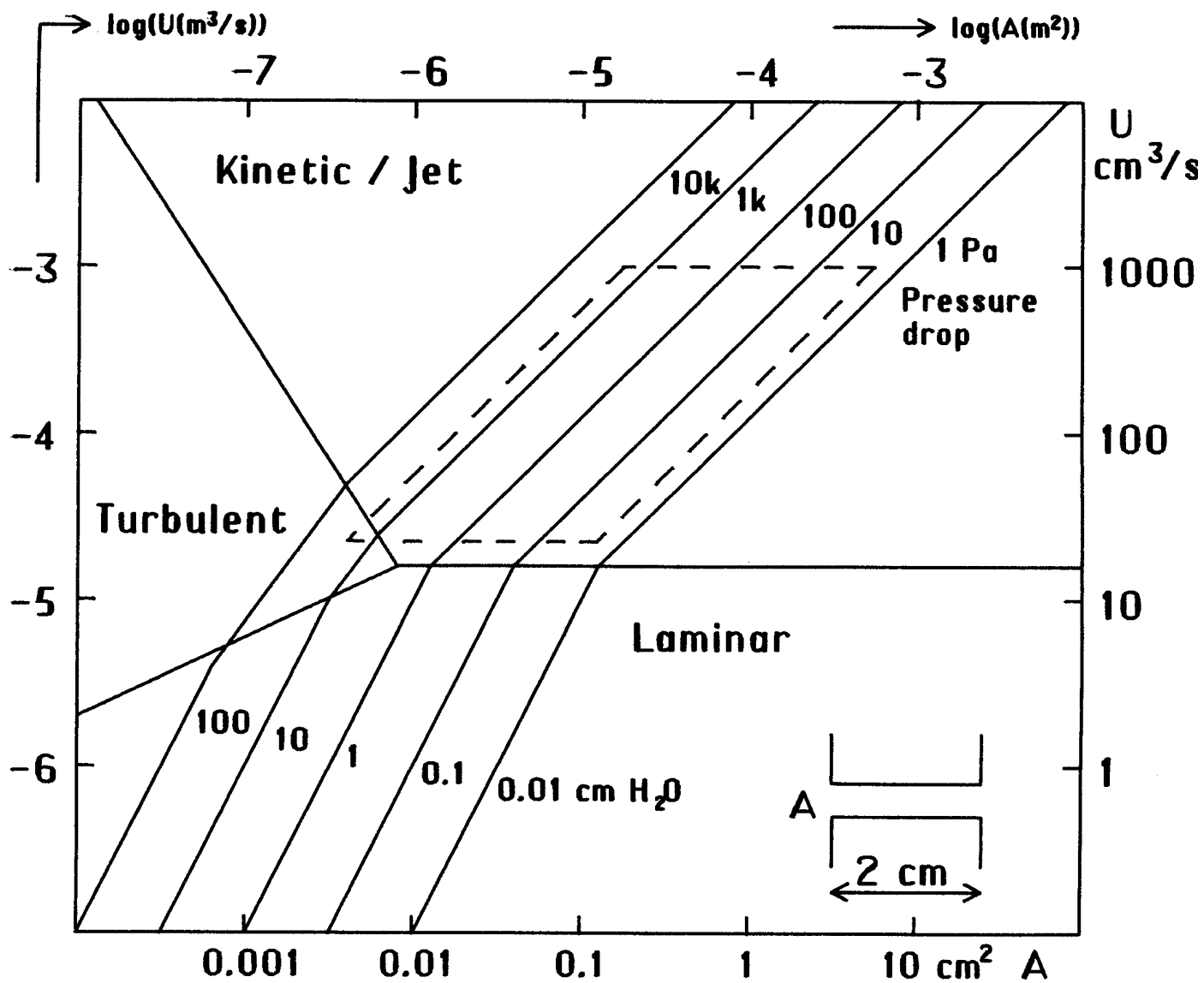
Fig 5.1. Computed flow vs area and pressure drop in a circular tube of length 2 cm, showing regions dominated by kinetic drop, laminar and turbulent resistance. The region of interest to speech noise generation is indicated and plotted otherwise in figs 5.2.
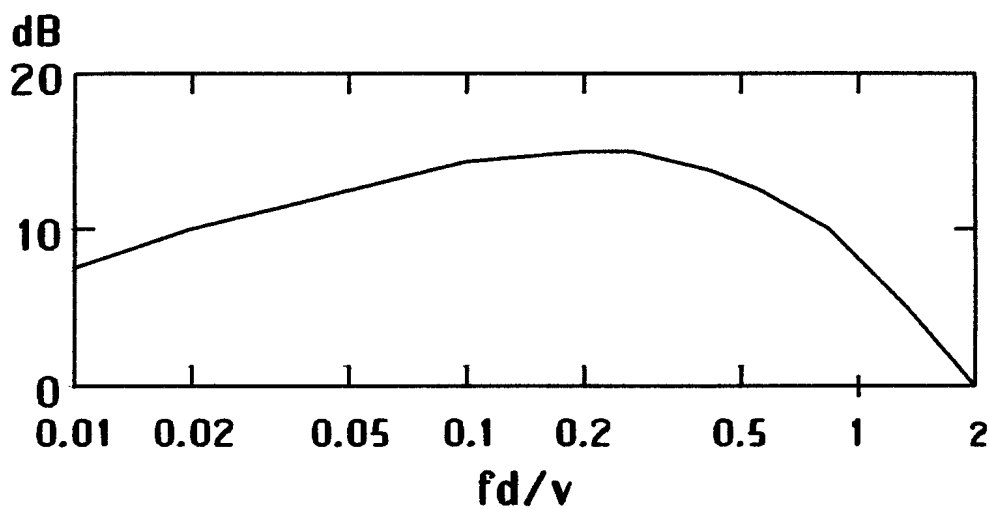
Fig 5.3. Normalized noise spectrum, after Stevens (1971)

5. NOISE GENERATION

Expressions for the noise generated at constrictions. Discussion and experiments. Practical implementation.

The problem of quantifying the noise source at constrictions in the voice organs is an elusive one treated in many classical works, notably Meyer-F ,ler (1953), van den Berg et al (1957), Heinz (1957), Fant (1 ͡ ju), Stevens (1971), Flanagan et al (1975), and the theory still lets much to be found. Here some elementary relations will be illustrated and then some practical measures to incorporate automatic noise genera- uon in the line model are outlined.

5.1 THEORETICAL CONSIDERATIONS

 ͡ ͡ ͡ a circular constriction of area A and length l. By excerting a pressure drop P_o across it we induce a flow $U_o = v_o * A$. As we saw in section 2.13 the Bernoulli equation links the pressure and flow when we neglect effects from shape and friction. From that we can assume a 'DC resistance'

$$R_o = P_o/U_o = \rho * v_o/(2*A) = \rho * U_o/(2*A^2) \qquad (5.101)$$

and a differential resistance $R_d = dP_o/dU_o = 2*R_o$. This was brought out in section 2.13 as a factor in the 'jet loss' coefficient. If we formulate it in terms of a loss coefficient $D = R/2Z$ the corresponding (partial) pressure drop will be

$$P_o = R*l*U_o = 2*Z*D*l*U_o = 2*\rho*c*D*l*U_o/A \qquad (5.102)$$

There will also be additional terms of pressure drop due to viscosity and turbulence. To visualize the dependence of differential pressure from area and flow we use the results from the loss investigation. From these D coefficients and eq (5.102) fig 5.1 was drawn. It accounts for the kinetic ('jet') drop and the laminar and turbulent resistances, remembering that R and D are per unit length with the distributed losses. The frequency dependent viscous resistance is not applicable since we now consider a DC flow. For the same reason the smaller R_o is used for the jet loss rather than R_d. In fig 5.1 a constriction length of 2 cm is used, an arbitrary but representative value. Once a length is given we can combine and compare the discrete and distributed losses in a single diagram. As was done in the earlier loss coefficient diagrams this is divided into different regions in which each one of the three in turn gives the dominant contribution. The magnitudes of this diagram essentially compare to those in a corresponding plot with van den Berg et al (1957).

The major comment to make to fig 5.1 is that for all pressures and flows typical with noise generation in speech the kinetic drop is by far the dominant one. The drop due to laminar flow resistance dominates only at very small flows, close to a cutoff passage. Its dominance here is in itself important for the function of a model, but not for the problem of noise generation. I take special notice that the drop due to turbulence in the passage is small, and interpret this as an indication that the major part of the noise is <u>not</u> generated in the passage but rather <u>downstream</u> the constriction. The rationale is that the noise power generated should be related to the mechanism by which the DC supply power is used up. Further support in this direction is that it takes time and distance for a laminar flow in a tube to break up into turbulence. Thus the flow in the short constricted passage may well be greatly laminar despite a high Reynold's number.

The relation between pressure, flow, and area is plotted a different way in fig 5.2. The axes contain pressure and flow in the range of interest for noise generation in speech, entirely dominated by the kinetic drop. This latter fact permits convenient display of other quantities on the axes. The pressure axis is thus directly marked also in particle velocity using the Bernoulli equation. Different littera of fig 5.2 show lines of constant area, lines of constant supplied power, and lines of constant Reynold's number.

The particular 2 cm constriction length of fig 5.1 was used just to establish the range where the kinetic drop dominates. Once this is the case the length does not affect the calibration of figures 5.2.

I want to make a special comment on not having a shape factor 'k' in the Bernoulli equation. This is elsewhere widely used to treat deviations from the ideal case. Omitting it makes it possible to insert the multiple variables on the abscissa of figs 5.2. When necessary you can instead account for shape effects by working with an effective area, more or less differing from, and mostly smaller than the geometric area ('vena contracta'). It is then just as easy to use a ratio of effective area to geometric. Anyway, neither this ratio, nor 'k', is a constant when flow is varied, not even with a constant shape.

The constriction has an inductance $L=\rho*l/A$. Let us further define a critical frequency f_c, at which the corresponding reactance ωL equals the differential resistance $2R_o$ as in (5.101)

$$f_c = R/(2*\pi*L) = v_o/(2*\pi*l) \qquad\qquad (5.103)$$

also put on a supplementary scale along the pressure axis together with v_o. This frequency is directly related to the particle transit time through the passage, l/v_o.

The air rushing through the constriction will now give rise to noise, let this be represented by a pressure source p_n.

Our main problem is to find a practical estimate of its amplitude. For that purpose we may refer to the classical finding of Meyer-Eppler (1953), quoted by Fant (1960), Flanagan (1965), also used by Flanagan et al (1975), saying that the noise pressure generated follows the general proportionality law

$$p_n \sim Re^2 - Re_c^2, \quad \text{when } Re > Re_c, \text{ else } 0 \qquad (5.104)$$

The noise pressure should be proportional to the Reynold's number squared, but biased off by some critical Reynold's number Re_c. Below Re_c no noise is produced. For plastic models Re_c has been established around 1800, somewhat lower for real speech sounds. I rewrite the proportionality into the form

$$p_n \sim (x - 1) \qquad \text{where } x = (Re/Re_c)^2 > 1 \qquad (5.105)$$

Using the definition of Re, eq (2.2), and the relation between area and diameter we can put

$$Re^2 = \frac{4*\rho^2}{\pi*\mu^2} * \frac{U_o^2}{A} \qquad (5.106)$$

We now approach a formulation that is applicable to our model. At its constrictions we can monitor the quantity U_o^2/A which then is proportional to Re^2 and form the quotient x

$$x = \left(\frac{Re}{Re_c}\right)^2 = \frac{4}{\pi} * \left(\frac{\rho}{\mu*Re_c}\right)^2 * \frac{U_o^2}{A} \qquad (5.107)$$

If this is less than unity we conclude that no noise is being generated and the following steps are skipped. But if greater we evaluate $x-1$. For high values of Re the noise source pressure will approach proportionality to x and hence to U_o^2/A.

It is interesting to compare this noise pressure proportionality to that of the static pressure drop which we found from the Bernoulli equation to be $P_o \sim (U_o/A)^2$. This would then lead to

$$p_n \sim (1-1/x)*P_o*A \qquad (5.108)$$

or with high Re, plainly $p_n \sim P_o*A$. This relation is further discussed by Fant (1960), where he also points out that the shape of the area has an effect on the noise generation efficiency. In short a circular area

is more efficient than a wide slit. The contours of equal p_n after eq (5.108), at 10 dB intervals, is shown in fig 5.2d.

When we for simplicity assume the internal impedance of the noise pressure generator to equal the resistance R_d we find the corresponding no-load noise flow generated to be

$$u_n = p_n/R_d \sim \frac{U_o^2}{A} * \frac{A*A}{\rho*U_o} \sim U_o*A \qquad (5.109)$$

This is just to show that we can just as well work with the proportionality u_n/U_o as with p_n/P_o, naturally they follow similar laws. Repeating, but this time with the inductive reactance ωL mentioned before (5.103) will again give the same proportionality, but this time augmented with the factor f_c/f. This may be of significance provided the noise source is located inside the constriction. It may justify special spectral slope shaping of the noise working with flows in the analog.

The spectral shape of the noise generated is an important property. Stevens (1971) illustrates the noise to have a broad peak at the frequency

$$f_n = 0.2 * v_o/d \qquad (5.110)$$

which is also shown with contours in fig 5.2c. In noise consonants f_n is typically 2-3 kHz except for aspirated sounds where it may lower down to 500 Hz. The basic spectrum shape after Stevens is shown in fig 5.3. The characteristic dimension d relates to the area over which the noise pressure excerts its force. Depending on circumstances in the geometry it may, or may just as well not, equal the constriction diameter.

Let us put the Re^2 proportionality in contrast to other findings or the same Stevens (1971) paper. Drawing from different works in aerodynamics he arrives at a basic proportionality

$$p_n \sim P_o * \sqrt{P_o} * A \qquad (5.111)$$

Fig 5.2e shows contours of equal p_n using this formula. Though this is a somewhat different proportionality, Stevens found it compatible to the Meyer-Eppler data.

When discussing the noise source in the vocal tract things are still more complicated. In the first place the noise is spectrum shaped by vocal tract filter function, in the second the load on the pressure source, and hence the power eventually reaching the output, depends on the tract shape. These factors are properly accounted for by the line model as such. But still the amplitude of the source itself is influenced by the tract shape. For instance Stevens notes that for the

quite reasonable assumption of an area function uniformly scaled with the constriction area would lead to $p_n \sim P_o$. This is an extreme deviation from the Re^2 proportionality and would give vertical equal noise contours in fig 5.2.

Finally it is necessary to state what we mean by noise pressure p_n having frequency range in mind. On one hand it may be the pressure pertaining to the total frequency spectrum generated, whichever this may be. This is the background for the proportionalities given by Stevens. On the other it may be specified within a fixed frequency band, for instance the one we would cover in a speech synthesizer. Here the noise generator should most preferably represent a spectrum density level rather than a total power. The Meyer–Eppler data is of the latter type, they were obtained from real speakers and thus include the vocal tract filtering.

Conversion between these two meanings of p_n requires knowledge of the spectrum shape. Trying this conversion with the spectrum of fig 5.3 is a bit complicated, but does at least in part explain the difference in slope between the equal noise level contours of figs 5.2d and 5.2e.

This is not the place to argue which formula is more or less right. Paying regard to the many different factors that influence the magnitude and shape of the noise tends to make a practical system overly complex. Also there is little hard data in the fricative noise literature as to what happens in the marginal regions of noise generation, those with small area (low flow) and large area (low pressure drop). To serve as a guide to a working compromise a classical nozzle-blowing experiment was repeated, with an attempt to approach these limits. Circumstances, measurements, and results are shown in tables 5.1 and 5.2.

In a first experiment air was blown into free space. This is convenient to establish a source spectrum shape, but gives experimental trouble in validating a critical Re_c because of the rather low noise levels obtained. The second experiment included a downstream cavity for a 4 kHz resonance to improve measurement range at this frequency. Moreover an obstacle was placed in the jet which increased the level up to as much as 20 dB. This again should be an argument that the natural speech noise source is located downstream the constriction. With an intraoral constriction the air jet will more likely than not strike a wall. This will increase the area subject to the fluctuating noise forces. The consequence will be louder noise and a shifted spectrum as per eq (5.110). Equal noise contours at 4 kHz from this experiment are shown in fig 5.2g. The obstacle having constant dimensions was naturally less effective with a large nozzle diameter. This shows up as a marked bend in the contours at large areas. As could be feared the critical Re_c is a function of the constriction area.

a.

b.

Fig 5.2. Flow vs pressure drop or particle velocity in a circular tube when the kinetic drop is the dominating term. For the pressure and flow range shown this implies a length less than 2 cm, cf fig 5.1. The scale of f_c applies to this length. For greater lengths the diagrams are distorted at low flows and high pressures (lower right corner).

a. Contours of equal effective area.

b. Contours of equal Reynold's number.

c.

DC power
1mW   10mW   100

100Hz /200   500   1kHz   2k   5k
Spectr peak

Uo cm3/s

Po N/m2

vo m/s

fc Hz

d.

0dB
−10dB
−20
−30
−40

200 mm2
100 mm2
50 mm2
20 mm2
10 mm2
5 mm2
2 mm2
1 mm2

$(1-1/X)P_0 A$

Uo cm3/s

Po N/m2

*Fig 5.2c-d. Flow vs pressure drop or particle velocity in a
circular tube when the kinetic drop is the dominating term.*

*c. Contours of equal supplied power and of equal noise spectrum
peak frequency.*

*d. Contours of equal noise level according to the Meyer-Eppler
rule, eq 5.108.*

Fig 5.2c, d

0dB

-10

-20

-30

-40

-50

$A * P_0^{3/2}$

200 mm2
100 mm2
50 mm2
20 mm2
10 mm2
5 mm2
2 mm2
1 mm2

Uo cm3/s

1000
500
200
100
50
20

Po N/m2

2   5   10   20   50   100   200   500   1000   2000

vo m/s

2   5   10   20   50

fc Hz

20   50   100   200

-10

0dB

-20

-00

-40

-30

$(1-1/X) * P_0$

200 mm2
100 mm2
50 mm2
20 mm2
10 mm2
5 mm2
2 mm2
1 mm2

Uo cm3/s

1000
500
200
100
50
20

Po N/m2

2   5   10   20   50   100   200   500   1000   2000

Fig 5.2e-f. Flow vs pressure drop or particle velocity in a circular tube when the kinetic drop is the dominating term.

e. Contours of equal noise level according to the Stevens rule, eq 5.111.

f. Contours of equal noise level in the model adopted, eq 5.21. The dashed load line applies to a DC source with open circuit pressure 8 cm H_2O and .5 l/s short circuit flow.

Fig 5.2e, f

g.

Uo cm3/s

200 mm2

100 mm2

Re=1800

50 mm2

20 mm2

Tab 5.2

10 mm2

5 mm2

2 mm2

1 mm2

Limit

40dB

50dB

60dB

70

Po N/m2

h.

10   17   100 mm

Hole diam d
Tube diam 20

2 mm
strip

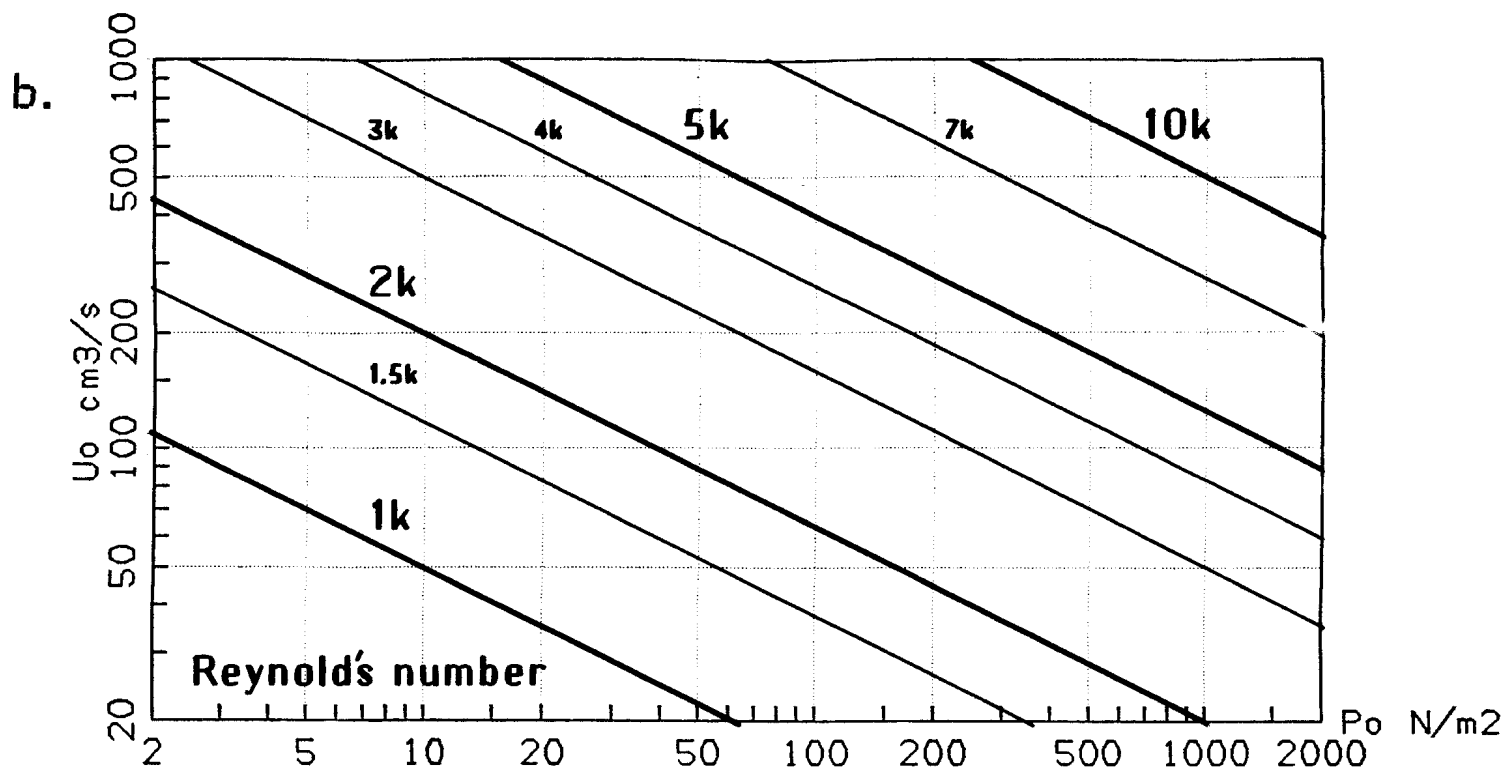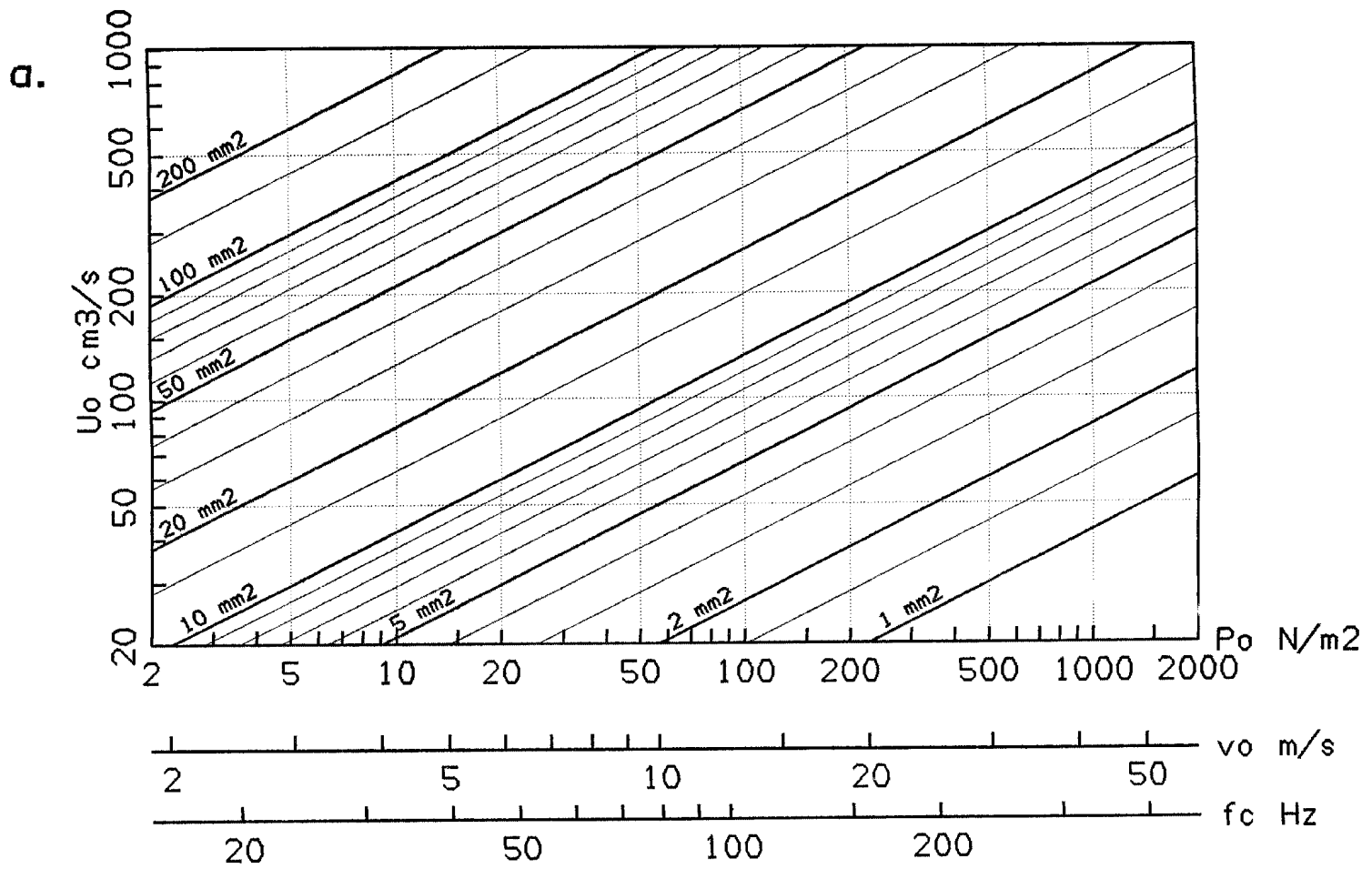U, P
Air DC supply
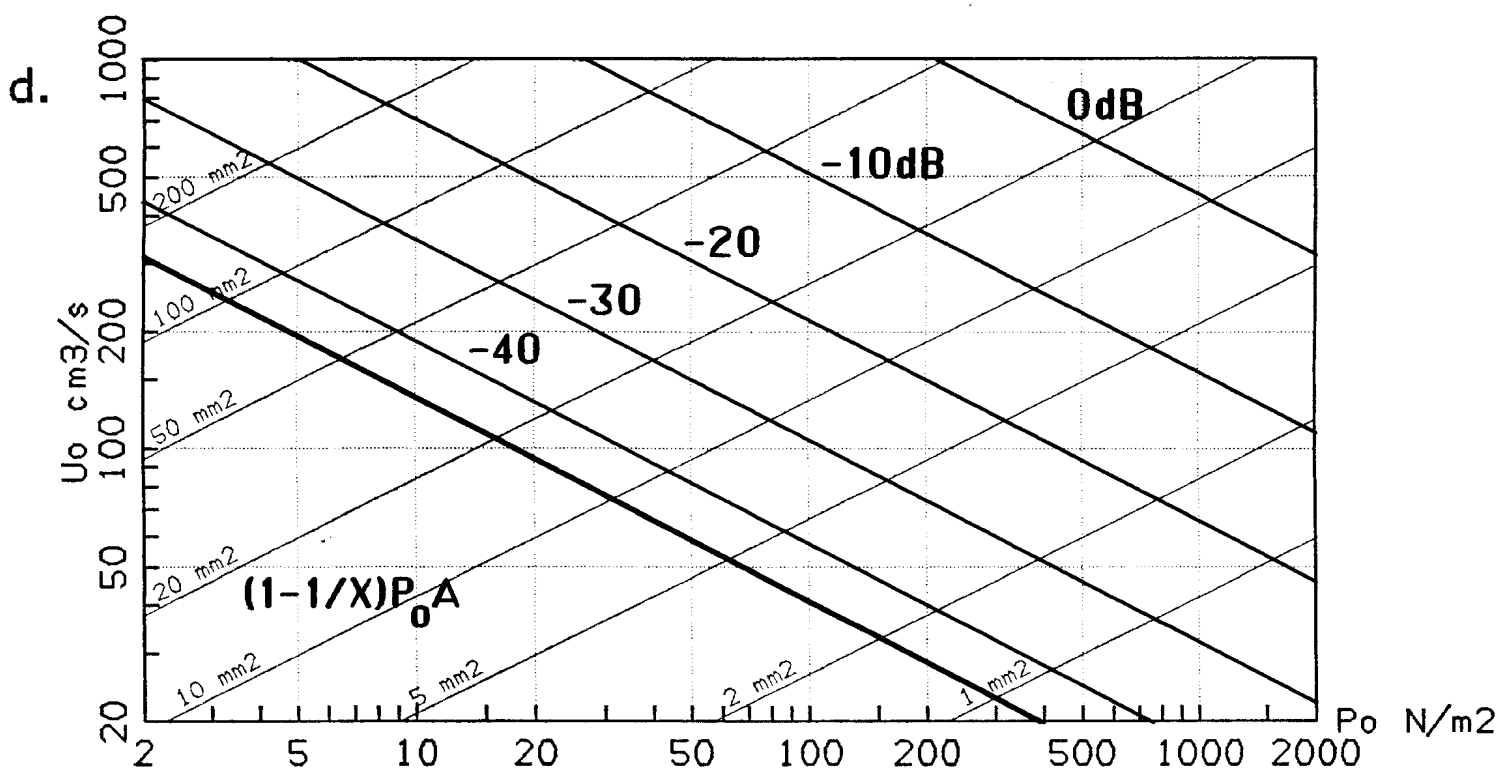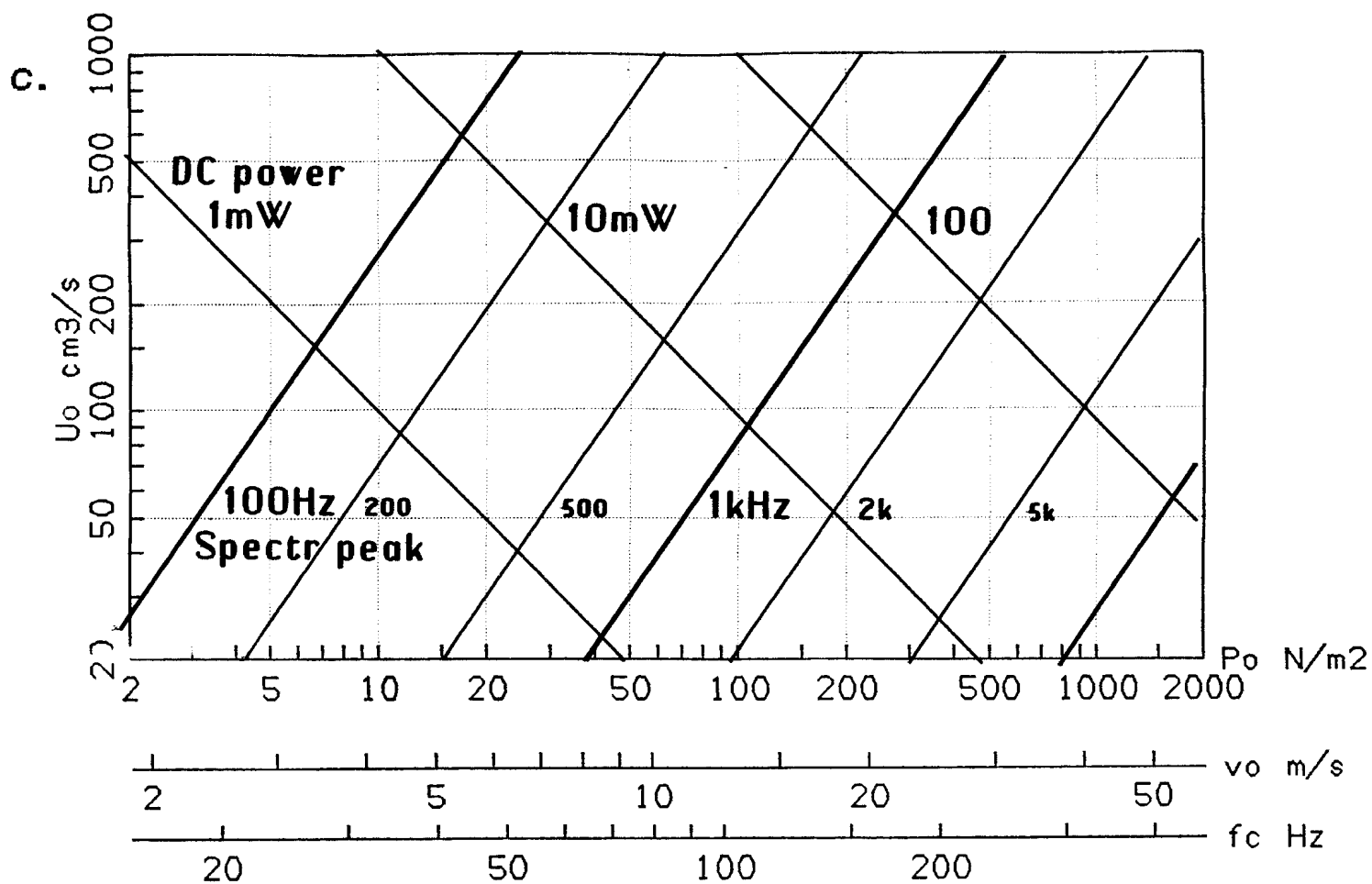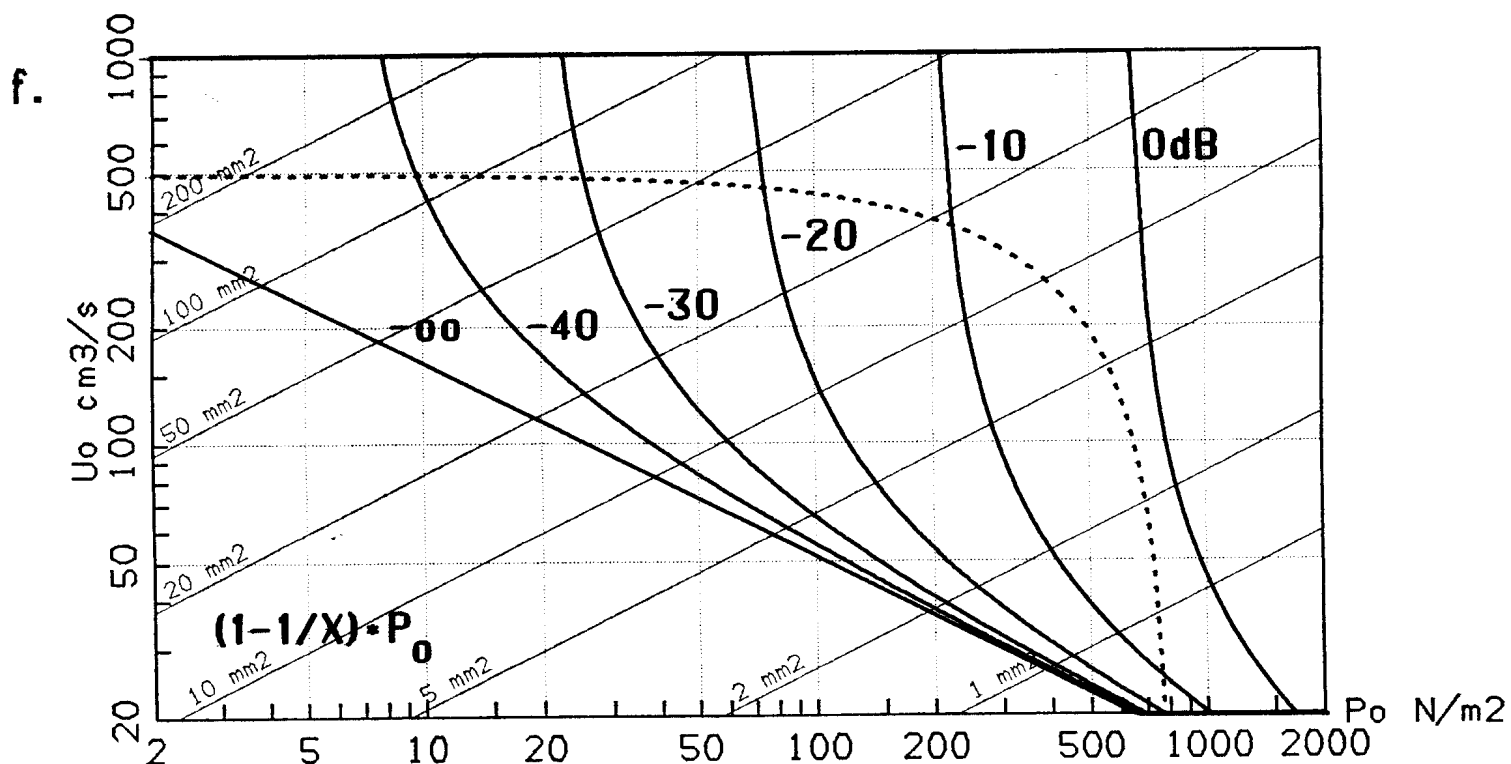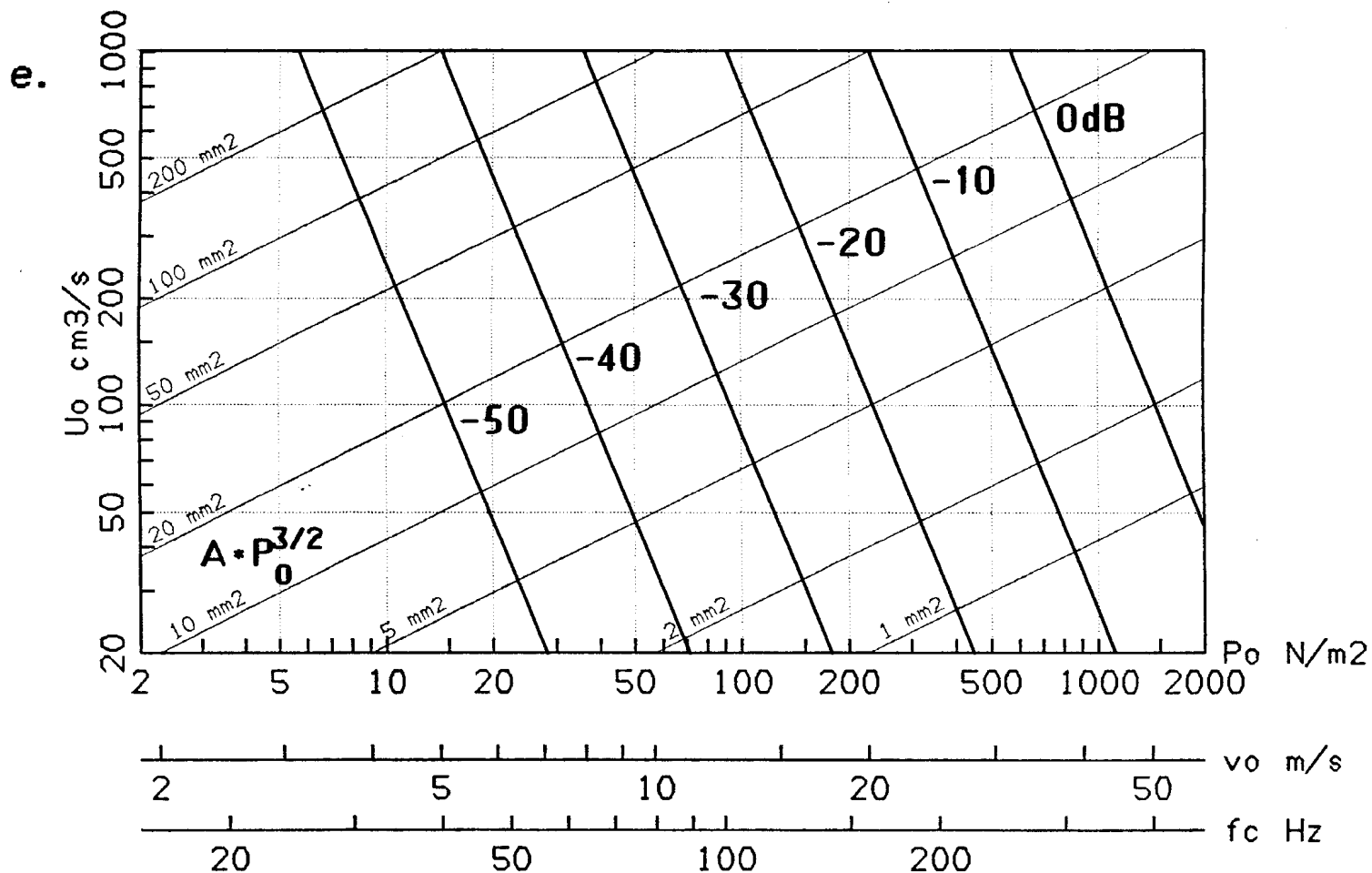
100 mm

100 mm

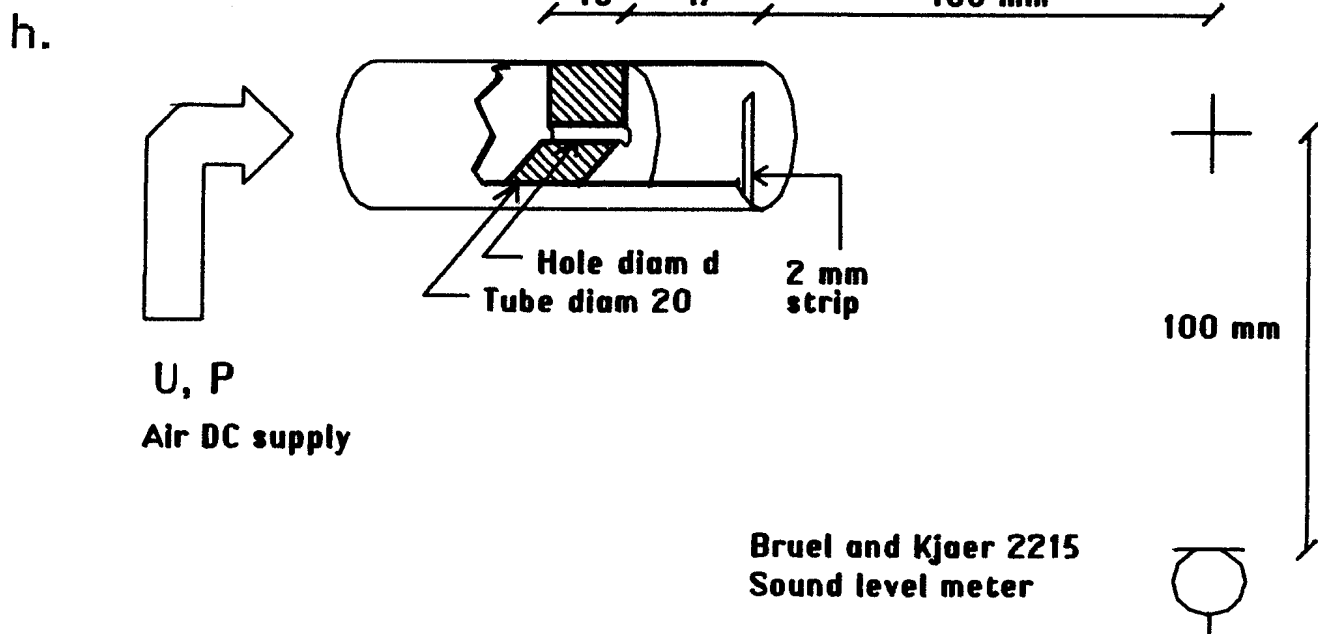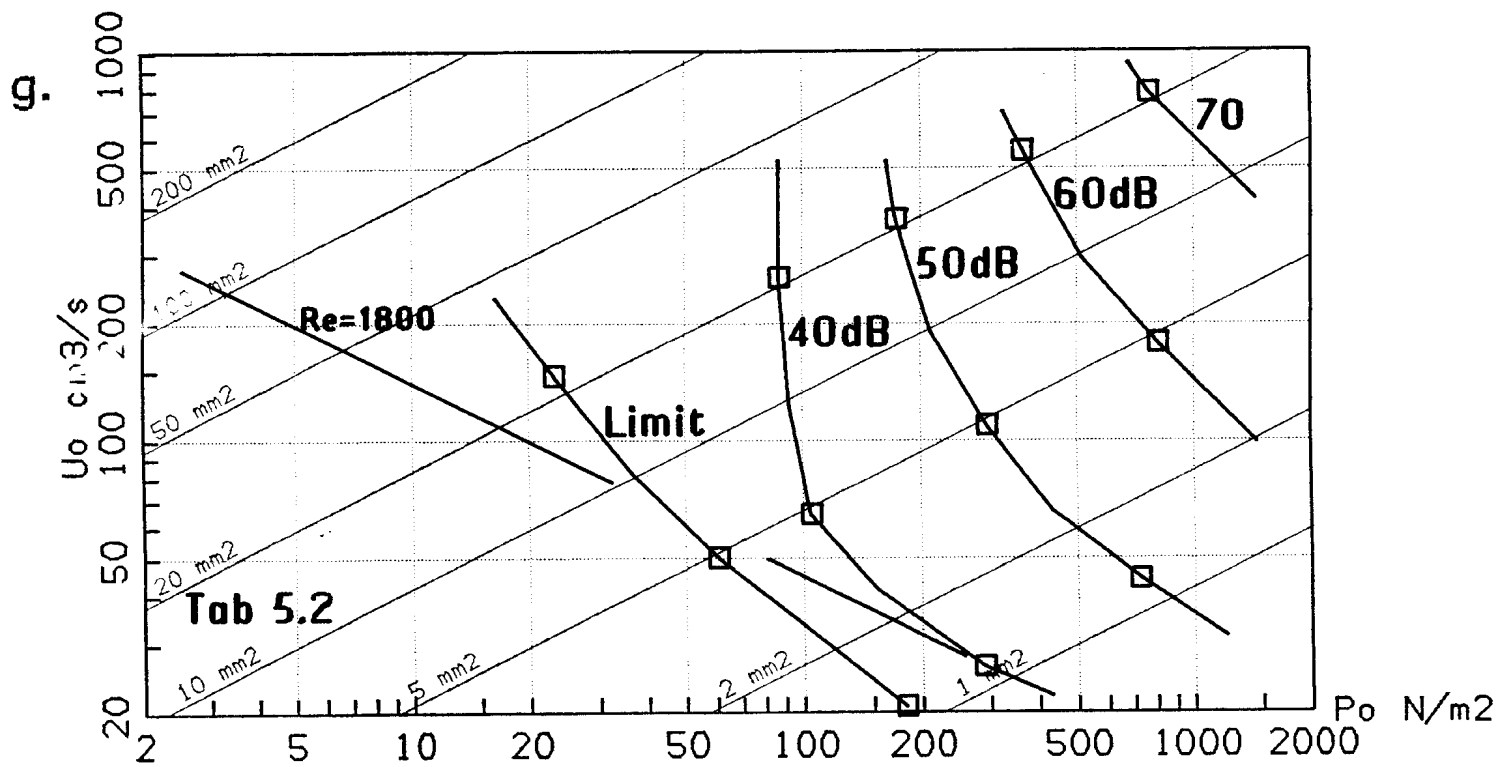Bruel and Kjaer 2215
Sound level meter

Fig 5.2g. Flow vs pressure drop or particle velocity in a circular tube when the kinetic drop is the dominating term.

g. Contours of equal noise level from the experiment of fig 5.2h and table 5.2.

Fig 5.2h. Layout of nozzle blowing experiment of table 5.2.

Fig 5.2g, h

Table 5.1.

Measurement of noise generated by air streaming out from cylindrical hole with sharp edges of axial length 10 mm and diameter d. Hole is at end of 20 mm diam feeder tube in anechoic chamber. Supply from variable transformer feeding reversed 'Nilfisk' vacuum cleaner and appr 6m of diam 40-25mm tubing.

Supply flow measured by 'VEB Pruefgeraete' airflow meter 200-2200 l/h (56-611 ml/s). Flow U is shown recomputed into ml/sec, at instances estimated.

Supply pressure measured with water manometer, given as P in mm H_2O.

SPL measured with 'A' weighting and octave analyses around 2, 4, and 8 kHz using 'Bruel & Kjaer' Prec sound level meter type 2215. Microphu.. perpendicular to, 100 mm away from exhaust axis, and 100 mm downstream.

First line is background noise.

d	P	U	LA	L2	L4	L8	v	de	Ae	fn	'm
mm	mmV	ml/s	dB	dB	dB	dB	m/s	mm	mm^2	kHz	dB
0	200	0	20	11	12	14	-	-	-		
1.4	38	31e	25.0	14.5	15.0	18.0	25.4	1.24a	1.21a	4.1	1⸮ ⸮
1.4	113	53	36.5	21.0	22.0	29.0	43.8	1.24	1.21	7.1	25.8
1.4	306	78	40.5	24.0	28.0	34.5	72.0	1.17	1.08	12.3	31.1
2.0	29	56	25.5	17.0	18.5	19.5	22.2	1.79	2.52	2.5	18.8
2.0	74	89	36.0	21.2	22.8	29.2	35.4	1.79	2.51	4.0	2o.2
2.0	187	136	43.5	26.2	29.0	37.2	56.3	1.75	2.42	6.4	33.3
2.0	292	167	47.0	29.0	33.2	41.0	70.4	1.74	2.37	8.1	37.0
2.8	8	56	-	-	-	-	11.6	2.48	4.83	0.9	-
2.8	26	100	31.5	22.5	22.8	23.8	21.0	2.46	4.76	1.7	23.3
2.8	57	144	41.0	23.5	26.2	33.7	31.1	2.43	4.63	2.6	30.1
2.8	109	194	50.0	28.0	31.2	41.5	43.0	2.40	4.51	3.6	36.6
2.8	286	319	52.0	35.0	39.0	47.5	69.6	2.42	4.58	5.8	43.3
3.9	9	131	26.0	18.5	16.5	-	12.4	3.67	10.56	0.7	15.4
3.9	22	200	33.0	24.5	26.5	25.3	19.3	3.63	10.36	1.1	25.5
3.9	52	306	43.0	32.5	36.0	37.5	29.7	3.62	10.30	1.6	36.4
3.9	102	422	50.2	39.0	43.0	46.3	41.6	3.60	10.14	2.3	44.3
3.9	222	622	58.2	45.8	49.0	55.5	61.3	3.60	10.15	3.4	52.3
3.9	315	742e	62.6	48.5	52.0	59.0	73.1	3.60a	10.15a	4.1	55.5
5.6	2	136	25	-	-	-	5.8	5.46	23.45	0.2	-
5.6	12	306	37.0	31.0	27.0	20.5	14.3	5.22	21.40	0.5	23.9
5.6	34	506	48.0	42.0	39.5	35.5	24.0	5.20	21.08	0.9	37.6
5.6	93	1043e	58.5	52.5	52.0	49.5	39.7	5.20a	21.08a	1.5	50.6

a indicates assumed value
e indicates values estimated from measurement and assumption

Velocity from strict kinetic drop: $v=\sqrt{2*P/\rho}=\sqrt{2*P/1.18}$
Effective area from flow and velocity: $A_e=U/v$
Effective diameter: $de=\sqrt{4*U/\pi*v}$
Outside flow meter range, from assumed eff area and veloc: $U_e=A_e*v$
Noise peak frequency after Stevens: $f_n=.2*v/d_e$
Mean level in speech band: $Lm=(L2+2*L4+4*L8)/7$

Tab 5.1

Table 5.2

Measurement of noise generated by air streaming out from cylindrical hole with sharp edges of axial length 10 mm and diameter d. Hole is at bottom of 17 mm deep cavity in 20 mm diam feeder tube in anechoic chamber. Supply from variable transformer feeding reversed 'Nilfisk' vacuum cleaner and appr 6m of diam 40-25mm tubing. Jet crossed by 2 mm wide thin plate strip 15 mm downstream nozzle. This obstacle increases sound level considerably. Cavity gives formant resonance around 4 kHz.

Supply flow measured with 'VEB Pruefgeraete' airflow meter 200-2200 l/h (56-611 ml/c). Flow U is shown recomputed into ml/sec. Supply pressure meas⸱⸱ ⸱ with water manometer, given as P in mm H_2O.

ɔPL measured with octave analysis around 4 kHz with 'Bruel & Kjaer' Prec sound level meter type 2215. Microphone perpendicular to, 100 mm away from exhaust axis, and 100 mm downstream.

First line is background noise.

d	P	U	L4	v	A_e
mm	mmW	ml/s	dB	m/s	mm^2
0	200	0	11	–	–
1.4	19	21e	40.0?	17.0	1.18a
1.4	37	30e	42.5	25.0	1.18a
1.4	67	40e	48.5	33.7	1.18a
1.4	150	58	55.5	50.4	1.15
1.4*	150	58	38.0	50.4	1.15
1.4	330	83	61.5	74.8	1.11
2.8	11	72	42.0?	13.7	5.26
2.8	17	83	45.0	17.0	4.89
2.8	48	117	54.0	28.5	4.11
2.8	89	181	62.0	38.8	4.66
2.8*	89	181	42.0	38.8	4.66
2.8	292	319	72.0	70.4	4.53
5.6	2e	150	13.0	6.4e	23.40a
5.6	5e	208	24.0	8.9e	23.40a
5.6	7e	258	36.0	11.0e	23.40a
5.6	12	333	45.5	14.3	23.29
5.6	25	472	55.0	20.6	22.91
5.6	66	768e	65.5	33.4	23.00a
5.6	125	1058e	76.5	46.0	23.00a
5.6	209	1369e	83.5	59.5	23.00a
5.6	58	722e	66.0	31.4	23.00a
5.6*	58	722e	57.0	31.4	23.00a

a indicates assumed value
e indicates values estimated from measurement and assumption
* indicates the obstacle was removed to show drop in sound level.
? indicates unstable, intermittent noise, at times with whistles

Velocity from strict kinetic drop: $v = \sqrt{2 \ast P/\rho} = \sqrt{2 \ast P/1.18}$
Effective area from flow and velocity: $A_e = U/v$
Effective diameter: $d_e = \sqrt{4 \ast U/\pi \ast v}$
Outside flow meter range, from assumed eff area and veloc: Uest=Ae*v
Low pressures estim from assumed eff area and flow: $P = (\rho/2) \ast (U/A_e)^2$

Tab 5.2

5.2 PRACTICAL IMPLEMENTATION

Having so many candidates for a noise level formula the following compromise has been selected with support from practical tests in the line model. The feature of the $1-1/x$ factor of (5.108) involving a critical Reynold's number is kept, the value $Re_c=1800$ seems to be adequate. A strict proportionality between p_n and P_0 tends to give a somewhat disturbing background of noise also at low Re. It would appear reasonable there should be no noise at all unless a critical Re_c is reached.

The final factor of area in (5.108), (5.109) has been dropped so that for large Re we simply have the proportionality $p_n \sim P_0$, but we never reach these Re with representative speech. This selection is not to say that the original formula is bad. But with the stiff control of the model practised so far there will easily be a pronounced "pinch" effect in producing fricatives when the area reaches too small a minimum neur the middle of the sound. Dropping the area releives this effect to some extent. The same can otherwise be reached with a more elaborate control allowing the intraoral pressure to widen the constriction passage.

To illustrate my proportionality formula used:

$$p_n \sim P_0 * (1-1/x) \qquad\qquad (5.21)$$

fig 5.2f shows the noise pressure p_n on an arbitrary dB scale in the plane of pressure drop P_0 versus flow U_0. The proposed relation appears to merge the desirable features of the theoretical formulas. Also, consider the load line from a DC pressure generator (assume 800 Pa = 8 cm H_2O), with an internal resistance to give a limited short circuit flow (assume 500 cm3/s), dashed in fig 5.2g. This will give an essentially constant noise level over a considerable range of areas, there is thus no need for particular precision in the 'articulation'.

The automatic noise generation procedure can be performed as follows First the narrowest tube downstream glottis in the line analog is located. It is hardly meaningful to inject noise from more than one supraglottal source at a time. Each sample interval the net flow U is evaluated here, and an average value is established. This averaging is to find the pseudo-DC flow rather than the momentary value. This is necessary to prevent instability caused by the noise subsequently injected. It is convenient to perform the averaging as a simple first-order lowpass filtering in the z domain. Thus from the input value U we can obtain the average

$$U_0 := (U - U_0 * z^{-1})/N_{av} + U_0 * z^{-1} \qquad\qquad (5.22)$$

where $U_0 * z^{-1}$ is the value saved from the previous sample interval, and N_{av} is the averaging time constant counted in sample intervals. For

example $N_{av}=40$ at 16 kHz sampling would give a time constant $T=40/16000$ s=2.5 ms corresponding to a LP cutoff at $1/(2*\pi*T)=64$ Hz.

Using U_0, the area A and a proportionality constant including Re_c we then find the value of x as in (5.107). If this is greater than unity we go on to form the gain coefficient $1-1/x$. Finally we compute a momentary noise flow value

$$u_n = U_0*(1-1/x)*c_n*K_n \tag{5.23}$$

where K_n is a proportionality constant telling the conversion efficiency from DC flow into noise flow. Empirically its order of magnitude is around 0.01. c_n is a random number different for every sample interval. The c_n signal should be of approximately unit average magnitude, zero average and properly spectrum shaped. For practical purposes it may suffice with simple differentiation from a sequence of square distributed random numbers, but more elaborate noise prefiltering can of course be done.

The final step is to inject the noise into the line, and that is done simply by adding u_n to the existing partial waves (compare eq (1.112)). Two notes have to be made now about into which tube to inject the noise. One considers if to take the narrow one by which we computed u_n or some tube further downstream. The other is to watch for the correct pair of forward and backward waves. As can be understood from section 1.6 only one half of those pairs will be used for the scattering computations forthcoming.

The exact location of the noise source has consequences for the resulting output spectrum of the line analog. This topic was discussed by Fant (1960) and has a continued interest, see e g Shadle (1983). With the source inside the constriction it will couple strongly to the upstream cavity at the resonance of the narrow tube. This causes zeroes in the spectrum even if these are easily concealed by low Q values due to the jet loss (the f_c scale in figs 5.2 has a connection to the bandwidth of the zeroes). A downstream source will give a more efficient excitation of the front cavity resonances.
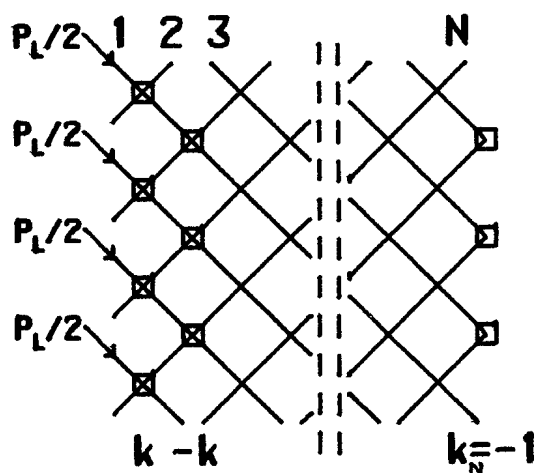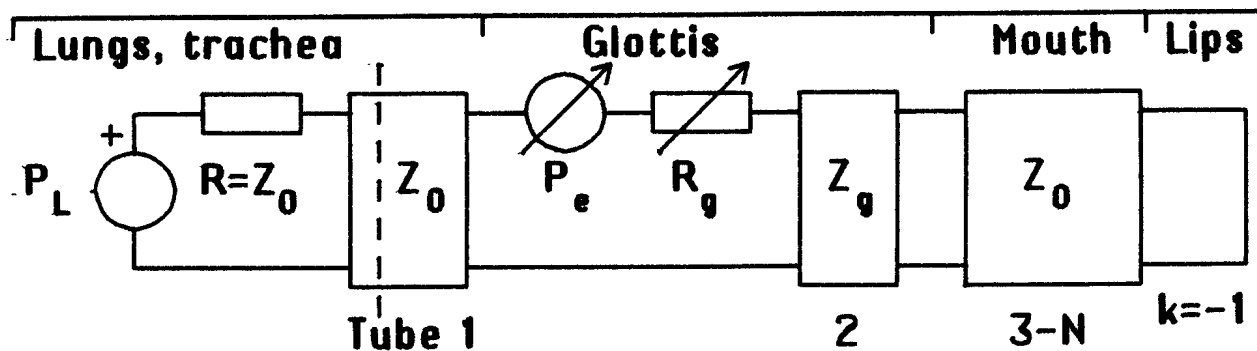
Fig 6.1. Network representation and signal flow diagram for the demonstration program of table 6.1. The system is supplied with a forward wave $P_L/2$. In case of total reflection at a closed glottis the backward wave will be equal and add up to a feeding static pressure P_L. Since the backward wave is neglected the lung termination is reflection free, hence the generator to the left of the dashed line is matched to the line impedance Z_0.

Fig 6.2. Simulation of glottis using an f-v continuity model, including laminar and jet resistance. From bottom to top: glottal area, glottal flow, glottal particle velocity, and intraoral pressure, with full scale values shown. Results are visually identical with either pressure or flow wave analogy.

The little frames to the right are explained with fig 6.4.

6. INTEGRATING GLOTTIS WITH THE LINE ANALOG

Simulation of glottal flow modulation applying nonlinear pressure drop correction or kinetic resistance. Function of the dynamic models.

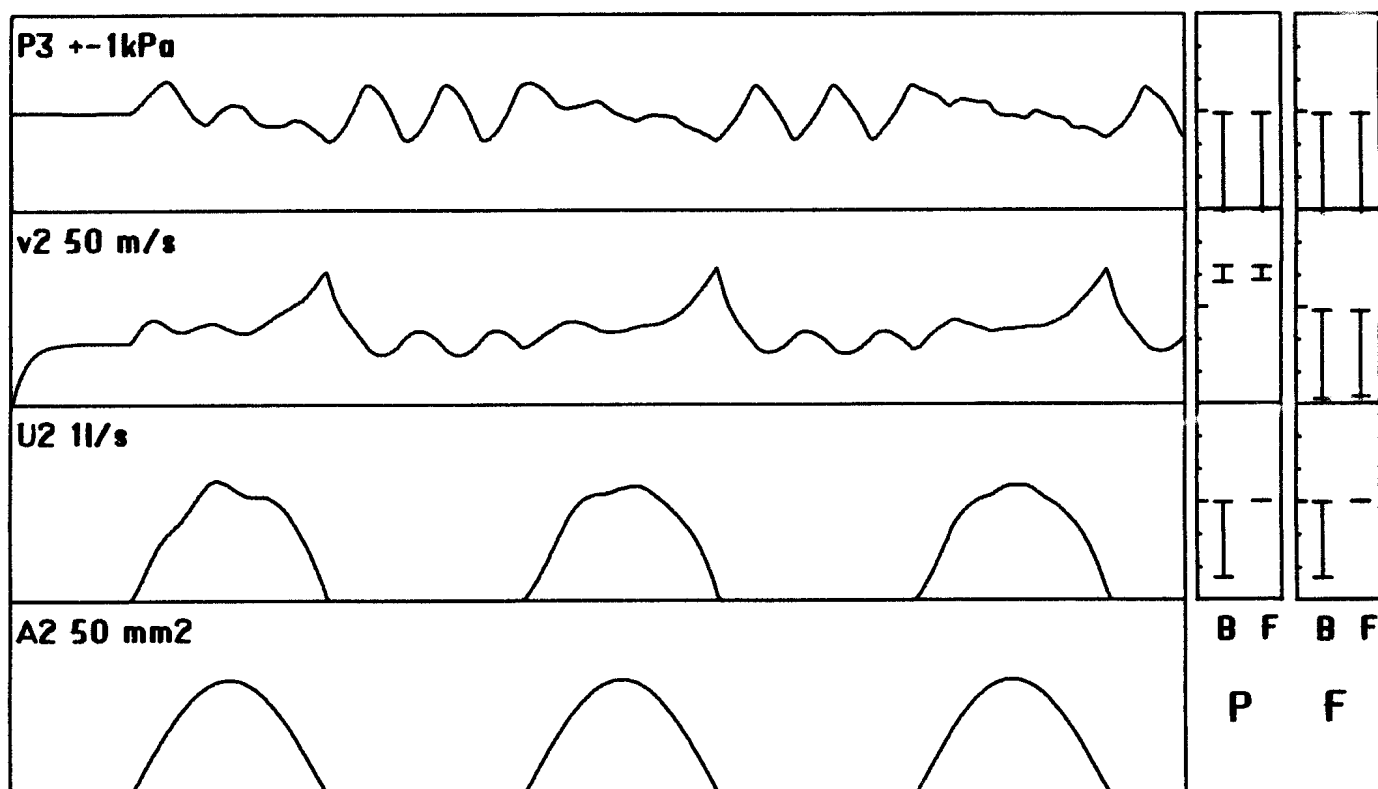Provided necessary corrections are included it is possible to treat the glottis as an integral part of the line analog. This has also been practised concurrently by Titze (1983). Advantages are obvious that we can then supply the line with a one-dimensional pulmonic pressure, and even _tter, continue the line upstream to include a model of the __uchea. The model will automatically account for the interaction of oral resonances on the glottal flow.

Titze simulated a parametric glottal model as part of a pressure partial wave analog. Seen as a component in the line his glottal model is a thin wall inserted between two tubes, and with a small hole having the variable glottis area. Accounting for the kinetic drop he derives a special ᵔᵔᵃᵗtering expression for this glottal joint. Though underlying theory is the same his result comes out formally different from the one here because of the differing physical layout.

It is fully practical to model also the mechanical properties of the glottis, but no attempts in that direction will be made here. Instead the glottis area is regarded as an input forced on the line model.

The method I have used is to let one specific tube in the line represent ᵗʰᵃ glottis, the reason is for simplicity of program structure: the glottis is just one tube like any other in the line. However the changes in glottal area are fast, so special recomputations of reflection coefficients and loss factors have to be made each sample interval for this tube.

For the ardent reader table 6.1 lists a commented Basic program, intended for experimentation on a personal computer. It simulates such a glottal valve section in a uniform tube, terminated by a total reflection at the lip end and without reflection toward the lungs, fig 6.1. The glottal area function of time is just a half sinusoid. To avoid numerical trouble the classical trick is used, area is limited downwards to a very small value (0.1 mm^2) rather than exactly zero. The program as such is a bit complicated with numerous conditional statements. This is to have it handle any of the many reflection models treated in section 1. It also includes the kinetic correction, laminar resistance, and jet loss resistance. The inclusion or omission of those features is chosen by the operator on setting switches at the start of the program. This program, augmented with plotting instructions, has produced the figures of this section.

a.
P3 +-1kPa

v2 50 m/s

U2 1l/s

B F
P

b.
P3 +-5kPa

v2 250 m/s

U2 5l/s

A2 50 mm2

B F

F
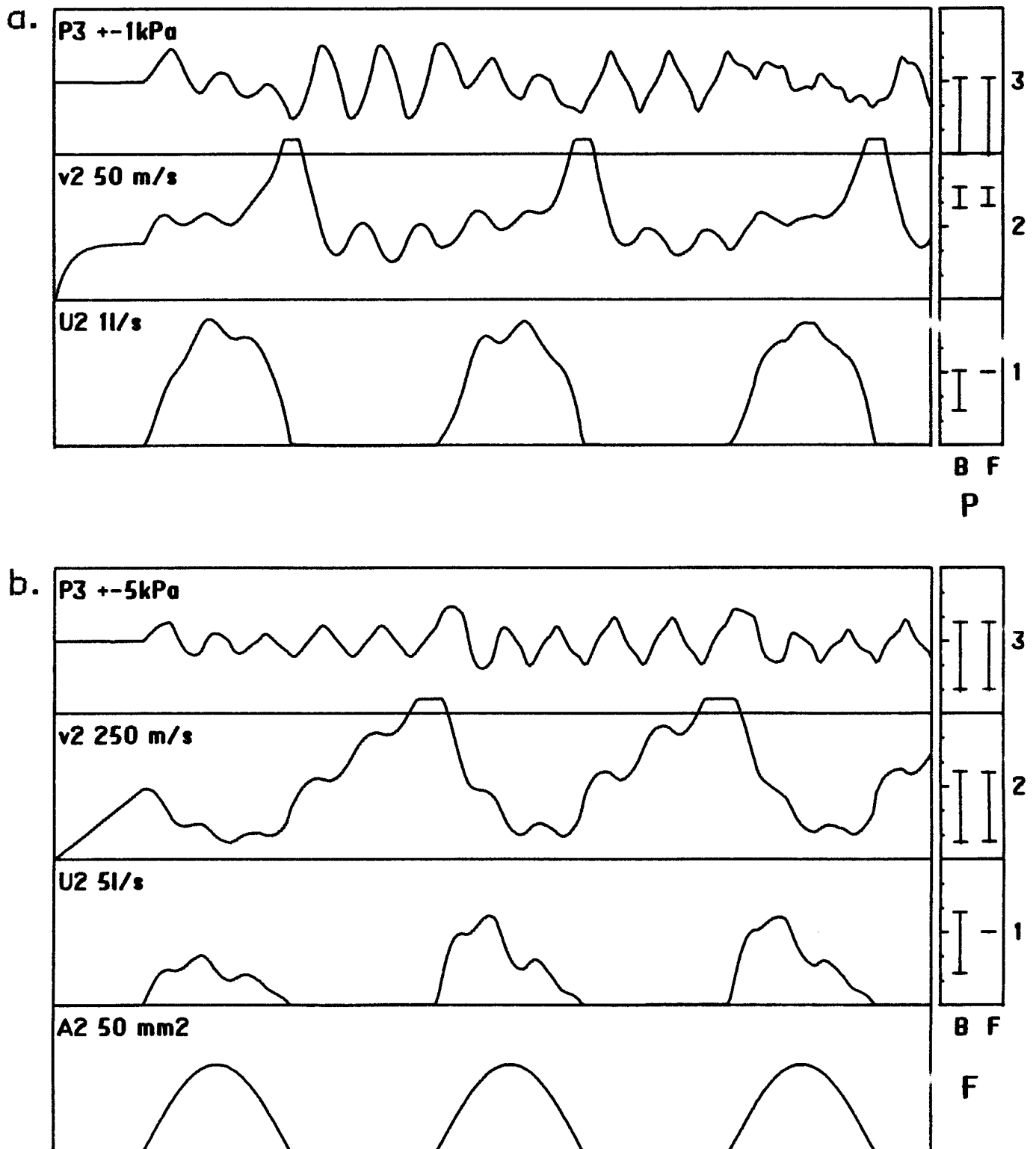
Fig 6.3. Simulation of glottis using an f-v continuity model.

a. With laminar, but excluding kinetic correction or jet resistance. Glottal velocity reaches unrealistic values, and so does the interaction ripple on glottal flow.

b. Excluding all losses (scales are reduced). Glottal velocity and flow avalanche, still the intraoral pressure shape does not look unreasonable.

In this degenerate example we can see the magnitude of the backward wave in tube 1 reaching rather large values, there are temporary negative subglottal pressures here!

Fig 6.3

Fig 6.2 shows a representative 'correct' result. From bottom to top it displays the glottal area, the glottal flow and particle velocity, and the intraoral pressure immediately downstream glottis. The particle velocity is not too interesting in itself but is good for diagnostics on how the model works. Initially this velocity is zero in the quiescent system but soon reaches an equilibrium determined by the resistance and the kinetic drop. During the open glottis phase the velocity and flow fluctuate due to pressure variations at both sides of glottis. Here the kinetic drop is very important to make the flow settle to realistic values. Take fig 6.3a with only laminar resistance to see what happens with _ it. It is essential not only because of magnitude but also because of its nonlinearity. This nonlinearity will influence the relative magnitudes of peak flow and ripple flow. Without any loss components the velocity and flow will avalanche as in fig 6.3b having reduced scales. Looking just at their shape may be misleading.

The glottal closure is the really critical point. Here the slope of glottal flow is a major determinant of spectrum level in the speech signal generated. I find the various reflection models can be put in two groups regarding their ability to cope in this application.

One group that works: the f-v adjustment and the Strube model, using either pressure or flow waves, and, amazingly, the static pressure wave model. The differences between their results are too minute to show up in the plots though the methods to get there are different enough.

The fact that the static pressure wave model happens to behave just as well as the proper dynamic ones is a consequence of the specific area conditions at glottis. Look for instance at the dynamic correction term in the Strube pressure scattering formula (table 1.1). It is proportional to the glottal area change divided by the neighbour tube area. At the glottis this neighbour area is always comparatively large so the dynamic correction will be very small. (In the corresponding flow analogy formula the correction is proportional to the impedance change, and then it is by no means small).

The other group can not manage the cutoff: the P-U adjustment, the Maeda model, and the static flow wave analogy. Fig 6.4a shows the degenerate behaviour of the latter without losses. Once glottal flow has got started you cannot stop it. The P-U dynamic models may even be disastrously unstable in some low loss cases.

Again, this is no wonder, these dynamic models were explicitly designed for flow continuity, and the static one is it implicitly. Such continuity corresponds to a large series inductance trying to maintain flow. A danger is, that using elaborate enough loss resistances, they can be patched up to look about right, fig 6.4b. This inductance brings a
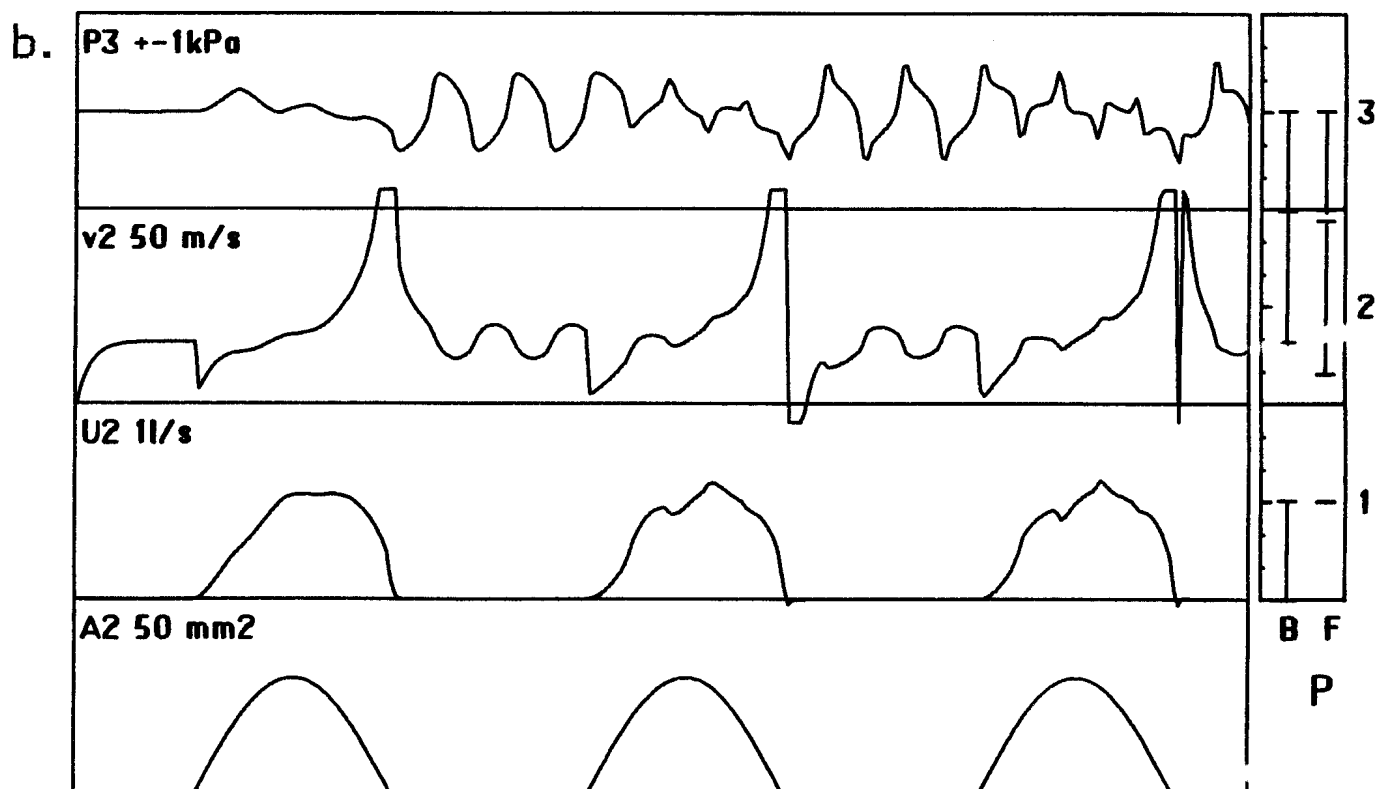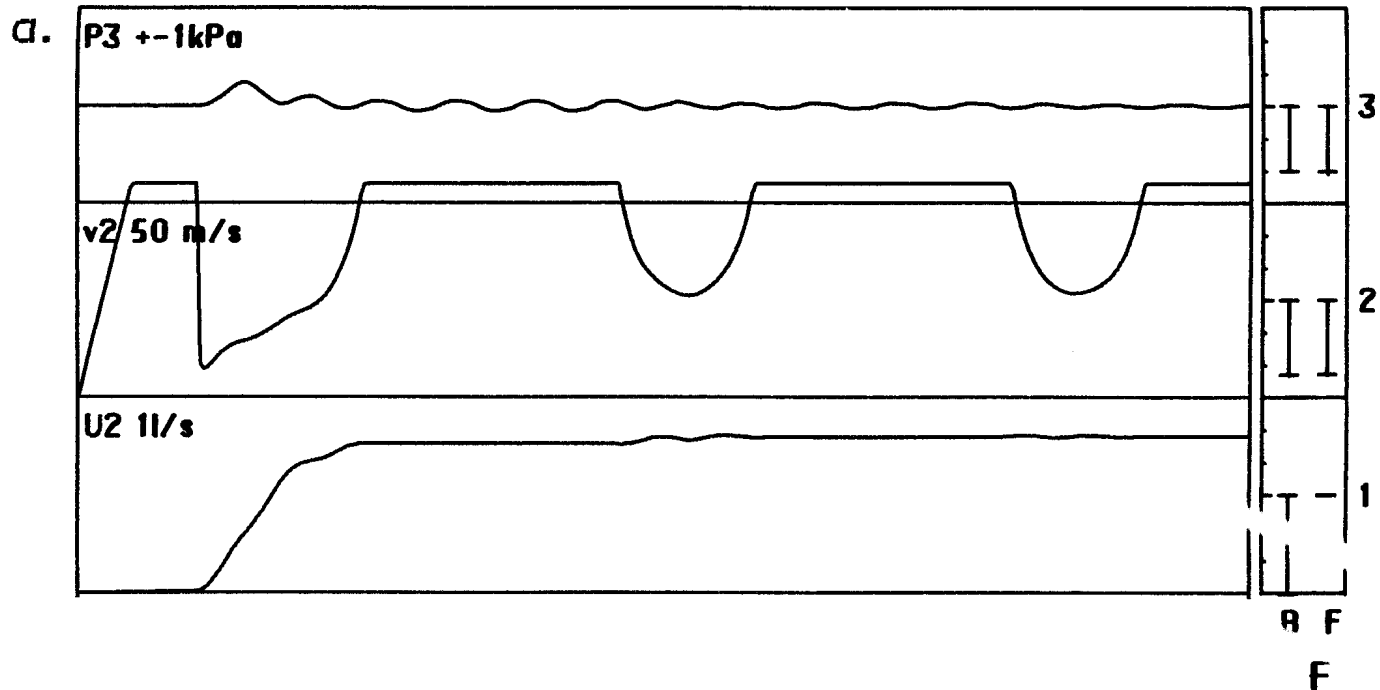
Fig 6.4. Simulation of glottis using a P-U continuity model.

a. Excluding all losses. The flow continuity has the absurd consequence of disabling the glottal cutoff.

b. With laminar and jet loss resistances. The system is marginally unstable at glottal closure. It has momentary negative flow and extreme partial wave magnitudes in the glottal tube 2.

The little frames to the right show partial wave magnitude ranges (except first 10 samples) in tubes 1, 2, and 3, separately for the backward and forward waves. All are normalized to the constant exciting forward wave in tube 1. In each frame the magnitude scale is logarithmic from 10^{-3} to 10^3. The frame sets marked P pertain to the pressure wave analogy and the ones marked F to flow.

Fig 6.4

pronounced time skew in glottal flow as compared to glottal area, something observed in real speech, but here it is artificial. As elaborated by Ananthapadmanabha and Fant (1982) this skew is due to the vocal tract load, and in reality the glottal inductance is probably insignificant. The skew is present to a realistic degree in fig 6.2. There it can easily be inferred from the slope of glottal velocity during open glottis, glottal area being symmetric.

The obvious method to account for the kinetic drop is to correct the scattering equations as was shown in section 1.13. This implies the evaluation of an uncorrected net flow, and from that the non-linear derivation of a correction flow to insert into the scattering equations.

A different method is to account for the same dynamically varying pressure drop by means of the jet loss of section 2.13. Then the jet loss factor is evaluated using the existing net flow. Then this loss factor is put into the series loss corrected scattering equations of section 1.8. Thus the kinetic drop is accounted for with just a resistive loss term, albeit flow dependent, on equal foot with any other resistance we would care to account for. I find this a significant simplification compared to the direct treatment of the kinetic drop.

These two methods account for the same thing, but the underlying derivations were set up from differing aspects. What should be warned against is of course to apply both corrections simultaneously. Putting the jet loss resistance at the input to the glottal tube gives a result indistinguishable from fig 6.2.

The jet loss resistor may also be moved to the next tube joint, the exit of 'glottis'. Then the particle velocity will get a slightly higher peak at glottal closure. One could say that too much air has been let into the glottal passage at closure, and at the exit it is a little too late to stop it.

```
10      DIM DELIN(50)    'DELAY LINE FOR UNIFORM VOCAL TRACT
20      DEFINT N         'INTEGER VARIABLE NAMES BEGIN WITH LETTER N
30      CEE=350          'm/s         SPEED OF SOUND
40      RHO=1.14         'kg/m3=Ns2/m4 DENSITY OF AIR
50      MU=.0000186      'Ns/m2       VISCOSITY OF AIR
60      PI=3.14159
70      AGMAX=.00003     'm2          PEAK GLOTTIS AREA, =0.3 cm2
80      AGMIN=.0000001   '            MINIMUM AREA, =.1 mm2
90      ATUB=.0004       'm2          UNIFORM VOC TR AREA, =4 cm2
100     PO=800           'N/m2        LUNG SUPPLY PRESSURE ~ 8cm H2O
110     TO =.01          's          PITCH PERIOD, FO=100 Hz
120     TAU=.0001        's          SAMPLING INTERVAL, 10 kHz SAMPL
130     DL=CEE*TAU/2     'm          TUBE SEGMENT LENGTH
140     NDEL=.175/DL     '           # OF DELAY LINE STEPS o VT LENGTH=.175m
150                      '           (ONE STEP FOR EACH TWO TUBE SEGMENTS)
160     ZO=RHO*CEE/ATUB 'Ns/m5       TUBE ACOUSTIC IMPEDANCE
170     '*******************************  REQUEST FUNCTION SWITCHES FROM OPERATOR *******
180     INPUT "Press analogy? (P)";SPRS$:XPRS=-1        'SWITCH PARTIAL WAVE ANALOGY
190     IF ((SPRS$="P") OR (SPRS$="p")) THEN XPRS=1      'XPRS=1:PRESS, =-1:FLOW
200     INPUT "Dyn: -2=Strube, -1=Maeda, 0=Static, 1=P-U adj, 2=f-v adj";XDMO
201                                      'XDMO SELECTS TYPE OF DYNAMIC MODEL
210     INPUT "Kinetic corr? ";XKIN      'XKIN SELECTS KINETIC CORRECTION
220     INPUT "1-2 Jet loss? ";XJ1       'XJn SELECT IF AND WHERE TO INSERT JET
230     INPUT "2-3 Jet loss? ";XJ2       '        LOSS RESISTANCE
240     INPUT "Lam resist? ";XLAM        'XLAM SELECTS LAMINAR RESISTANCE CORR
250     DLIM=9.9:IF XJ1+XJ2+XLAM<>0 THEN INPUT "D limit? ";DLIM    'MAX LOSS FACTOR
260     IF DLIM=0 THEN DLIM=9.9
270     PRINT "  PO  Dyn  Kin  J12  J23  Lam  Dmax";
280     IF XPRS=1  THEN PRINT "  PRESS"
290     IF XPRS<>1 THEN PRINT "  FLOW"
300     PRINT USING "####.#";PO;XDMO;XKIN;XJ1;XJ2;XLAM;DLIM
310     '---------------------------- INITIAL SETUP OF VARIABLES --------------------
320     FOR I=0 TO NDEL:DELIN(I)=0:NEXT I        'INITIALLY CLEAR V. T.
330     FOR3=0:FOR2=0:BAK3=0:BAK2=0     'TUBE 2 & 3 FORW & BACKW WAVES
340     AGLT=AGMIN                       'INITIALLY "CLOSED" GLOTTIS
350     FORO1=PO                         'REFERENCE INPUT FROM LUNG PRESSURE
360     IF XPRS<>1 THEN FORO1=PO/ZO      'CONVERT FOR FLOW ANALOGY
370     FOR1=FORO1/2                     'CONSTANT FORWARD FEED INTO TUBE 1
380     BAK1=FOR1                        '(THIS IS REDUNDANT FOR THE SCATTERING)
400     '**************************************** COMPUTATION LOOP ***************************************
410     FOR T=-.0010 TO .0070 STEP TAU   'DEFINE TIME SPAN OF LOOP
420     AOLD=AGLT:RFCO=RFAC              'SAVE PREVIOUS GLOTTIS AREA, REFL COEFF
430     AGLT=AGMAX*SIN(2*PI*T/TO)        'AND MAKE THE ACTUAL ONE AS A SINE ARC
440     IF AGLT<AGMIN THEN AGLT=AGMIN    'LIMIT AREA DOWNWARDS
450     RFAC=(ATUB-AGLT)/(ATUB+AGLT)     'REFLECTION COEFF AT JOINT 1-2,           (1.15)
460     '----------------------------- WAVE ADJUSTMENT IN TUBE 2 --------------------
470     IF XDMO<=0 GOTO 550              'MAEDA, STRUBE OR STATIC, NO ADJUSTMENT
480     IF XDMO>=2 GOTO 520
490     IF XPRS<>1 THEN PD=(FOR2+BAK2)*(1-AGLT/AOLD)/2       'P-U ADJ TUBE 2       (1.43)
500     IF XPRS=1  THEN PD=(FOR2-BAK2)*(AGLT-AOLD)*.5/AGLT   'WHEN XDMO=1
510     FOR2=FOR2-PD:BAK2=BAK2+XPRS*PD:GOTO 550
520     IF XPRS<>1 THEN PD=(FOR2-BAK2)*(1-AGLT/AOLD)/2       'f-v ADJ TUBE 2       (1.45)
530     IF XPRS=1  THEN PD=(FOR2+BAK2)*(AGLT-AOLD)*.5/AGLT   'WHEN XDMO=2
540     FOR2=FOR2-PD:BAK2=BAK2-XPRS*PD
550     '----- FIRST PASS: COMPUTE ODD NUMBERED FOR AND EVEN BAK PARTIAL WAVES --------
560     '----------------------------- COMPUTE AT JOINT 2-3, HAVING -RFAC -----------
570     PN=-RFAC*(FOR2-XPRS*BAK3)        'COMMON SCATTERING TERM                   (1.16), (1.23)
580     IF XDMO=-1 THEN PN=-RFCO*(FOR2-XPRS*BAK3)        'MAEDA                    (1.32-4)
590     DG=XLAM*2*PI*MU*DL/(RHO*CEE*AGLT)     'LAMINAR LOSS FACTOR/2, SWITCHED BY XLAM      (2.5)
600     DJ=0:IF XJ2=0 GOTO 630
610     U2=FOR2-BAK2:IF XPRS=1 THEN U2=U2*AGLT/(RHO*CEE)        '                  (1.12), (1.21)
620     DJ=ABS(.5*U2/(CEE*AGLT))         'JET LOSS FACTOR                          (2.12)
630     D=DG+DJ*XJ2:IF D>DLIM THEN D=DLIM        'TOTAL SERIES LOSS FACTOR
```

Tab 6.1 Experimental program in Basic language to simulate glottal valve and a uniform resonator tube, fig 6.1. Depending on switches set by the operator any of the simulation models can be used. Kinetic correction and jet and laminar losses can be applied.

Bracketed numbers to the right indicate corresponding equations in the text.

Tab 6.1

```
640    IF XPRS=1 GOTO 660
650    URF=((1+RFAC)*FOR2-(1-RFAC)*BAK3)*D/(1+D):URB=URF:GOTO 670         '                 (1.86)
660    URF=(FOR2-BAK3)*D/(1+D):URB=URF*(1+RFAC):URF=URF*(1-RFAC) 'LOSS CORR              (1.87)
670    FOR3=FOR2+XPRS*PN-URF              'SCATTERING WITH LOSS CORRECTION
680    BAK2=BAK3+PN+URB
690    IF XDMO<>-1 GOTO 720               'SENSE IF MAEDA DYNAMIC CORRECTION
700    IF XPRS<>1 THEN BAK2=BAK2+(FOR2+BAK3)*(AGLT-AOLD)/(AOLD+ATUB) 'MAEDA FL           (1.32-4)
710    IF XPRS=1 THEN BAK2=BAK2-(FOR2-BAK3)*(AOLD/AGLT-1)/(1+AOLD/ATUB) ' M PR
720    IF XDMO<>-2 GOTO 760               'SENSE IF STRUBE DYNAMIC CORRECTION
730    IF XPRS<>1 THEN USTR=FOR2*(1-AGLT/AOLD)/(1+AGLT/ATUB)    'FLOW ANALOGY
740    IF XPRS=1  THEN USTR=FOR2*(AGLT-AOLD)/(AGLT+ATUB)        'PRESS ANALOGY           (1.35-6)
750    FOR3=FOR3-USTR:BAK2=BAK2-XPRS*USTR
760    '----- SECOND PASS: COMPUTE EVEN NUMBERED FOR AND ODD BAK PARTIAL WAVES -------
770    '------------------------------- COMPUTE AT JOINT 1-2, HAVING +RFAC -----------
780    PN=RFAC*(FOR1-XPRS*BAK2)          'COMMON SCATTERING TERM                (1.16), (1.23)
790    U2=FOR2-BAK2:IF XPRS=1 THEN U2=U2*AGLT/(RHO*CEE)         '              (1.12), (1.21)
800    PE=0:IF XKIN=0 GOTO 880           'SKIP IF NO KINETIC CORRECTION
810    IF XPRS <>1 GOTO 860
820    POB=2*(FOR1-BAK2):PE=-XKIN*POB*ABS(POB)/(1.5*ABS(POB)+2*RHO*CEE*CEE)     '        (1.127), (1.1210+1.125)
830    FOR2=F'....+(1+RFAC)*(POB+PE)/2    'KINETIC CORR, PRESSURE ANALOGY               (1.1211)
840    BAK1=FOR1-(1-RFAC)*(POB+PE)/2
850    GOTO 900
860    U0=FOR1-BAK2-PN:V0=U0/AGLT         'UNCORRECTED FLOW AND VELOCITY                (1.122)
870    PE=-XKIN*U0*ABS(U0)/(1.5*ABS(U0)+2*CEE*AGLT)     'KINET CORR FLOW (APPR)          (1.125)
880    FOR2=FOR1+XPRS*PN+PE:BAK1=BAK2+PN-PE         'SCATTER W KINET CORR               (1.112)
890    IF XJ1<>0 THEN DJ=ABS(.5*U2/(CEE*AGLT))       'JET LOSS FACTOR                   (2.12)
900    D=DG+DJ*XJ1:IF D>DLIM THEN D=DLIM             'SETTLE LOSS FACTOR
910    IF D=0 GOTO 970                               'SKIP IF NO RESISTIVE LOSSES
920    IF XPRS=1 GOTO 950
930    URF=((1-RFAC)*FOR1-(1+RFAC)*BAK2)*D/(1+D)     'LOSS CORR WITH FLOW               (1.86)
940    URB=URF:GOTO 960
950    URF=(FOR1-BAK2)*D/(1+D):URB=URF*(1-RFAC):URF=URF*(1+RFAC)'CORR W PRESS           (1.87)
960    FOR2=FOR2-URF:BAK1=BAK1+URB                   'CORRECT FOR SERIES LOSSES
970    '------------------------------- SIMULATE V. T. AS UNIFORM TUBE ---------------
980    BAK3=-1!*DELIN(NDEL)                          'RETURNING FROM TOTAL REFL AT LIPS
990    DELIN(1)=FOR3                                 'TUBE 3 GOES INTO DELAY LINE FOR V.T.
1000   FOR I=NDEL-1 TO 1 STEP -1:DELIN(I+1)=DELIN(I):NEXT I    'UPDATE DELAY
1010   '*********************************** DISPLAY RESULTS ON CRT SCREEN ***************
1020   U1=FOR1-BAK1:U2=FOR2-BAK2:IF XPRS<>1 GOTO 1040          'NET FLOWS 1 AND 2     (1.21)
1030   U1=U1/Z0:U2=U2*AGLT/(RHO*CEE)                 'PRESS ANALOGY
1040   V2=U2/AGLT                                    'VELOCITY 2
1050   P1=(FOR1+BAK1):P3=(FOR3+BAK3):IF XPRS=1 GOTO 1070       'TOT PRESS 1 AND 3     (1.12a)
1060   P1=P1*Z0:P3=P3*Z0                             'FLOW ANALOGY                    (1.21)
1070   '
1080   SUT$=".           1           .           .           .           .           1           .           "
1082                                    'SCREEN DISPLAY RASTER DIVISIONS
1090   NS=AGLT*1E+06:  SK$="a":GOSUB 1230            '10 mm2/div, GLOTTAL AREA
1100   NS=10+U2*50000!:SK$="U":GOSUB 1230            '200 cm3/sec.div, GLOTTAL FLOW
1110   NS=V2:          SK$="v":GOSUB 1230            '10 m/s.div, GLOTTAL VELOCITY
1120   NS=60+P3/100:   SK$="p":GOSUB 1230            '1 kPa/div, ORAL PRESSURE
1130   NS=60+P1/100:   SK$="s":GOSUB 1230            '1 kPa/div, SUBGL PRESSURE
1140   NS=50+10*FOR2/FOR1:IF XPRS=1 THEN NS=50+FOR2/(2*FOR1)
1150   'SK$="f":GOSUB 1230                           'NORMALIZED FORWARD WAVE 2
1160   NS=50+10*BAK2/FOR1:IF XPRS=1 THEN NS=50+BAK2/(2*FOR1)
1170   'SK$="b":GOSUB 1230                           'NORMALIZED BACKWARD WAVE 2
1180   PRINT SUT$
1190   NEXT T         '******** END MAIN LOOP ***********
1200   GOTO 180                                      'START ALL OVER
1220   '************** SUBROUT TO INSERT CHARACTER SK$ INTO CRT DISPLAY LINE **********
1230   IF NS>78 THEN RETURN              'SKIP IF OVERRANGE
1240   IF NS<0 THEN RETURN
1250   SUT$=LEFT$(SUT$,NS)+SK$+MID$(SUT$,NS+2,78-NS)
1260   RETURN
```
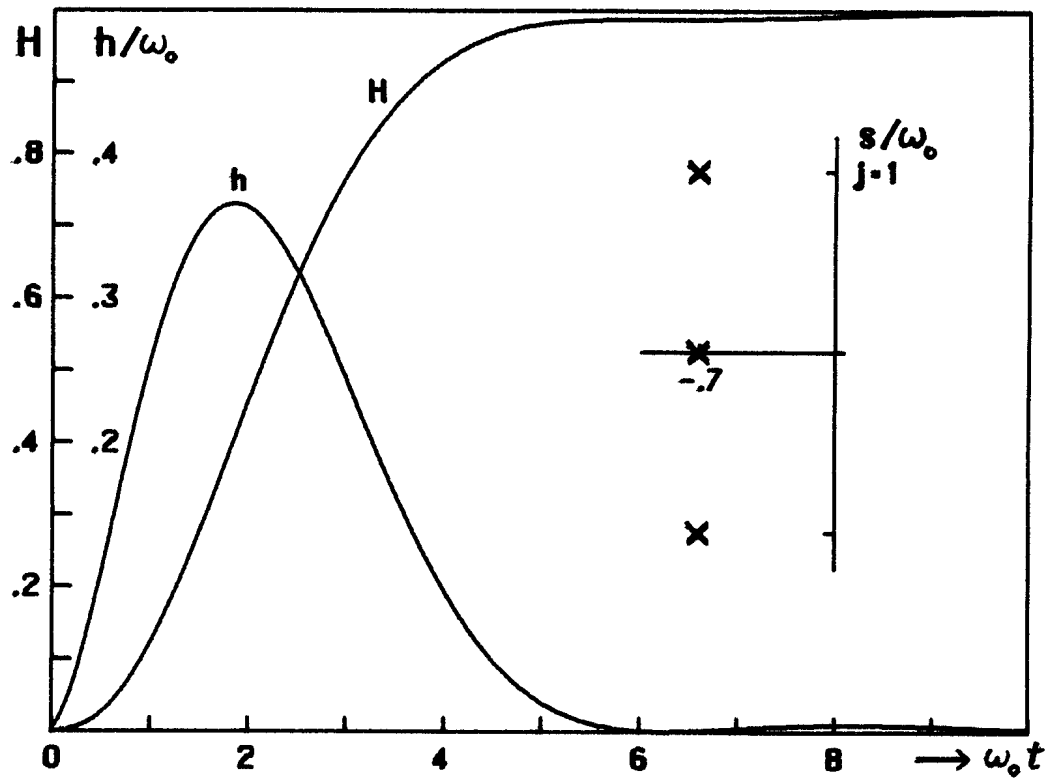
Tab 6.1

Fig 7.1. Normalized impulse and step responses of third order parameter smoothing filter.
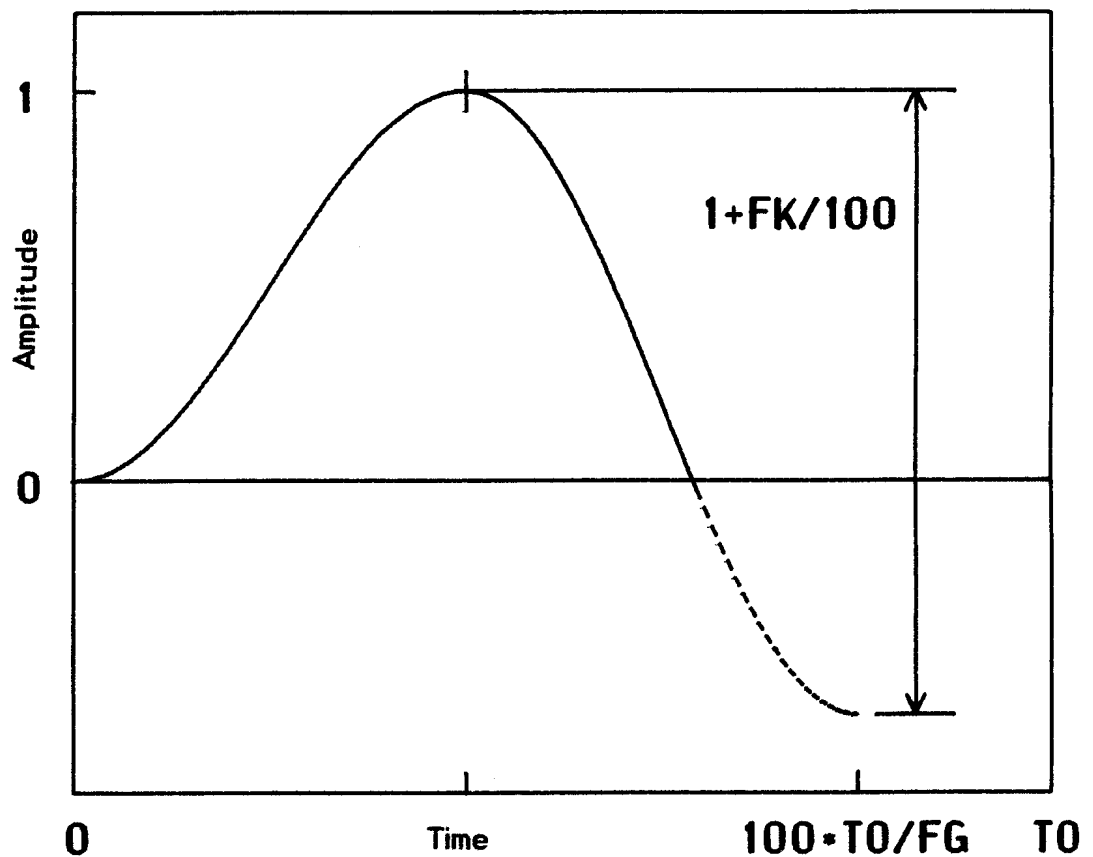

Fig 7.2. Meaning of FK and FG parameters controlling the Fant glottal waveform model. TO is the pitch period. The shape is built from two half period cosine arcs with the glottal frequency being FG percent of the pitch 1/TO. The magnitude of the second arc is scaled by the FK parameter. Only the positive part of the shape is used.

7. DESCRIPTION OF PROGRAMS

Structure and editor for synthesis parameter files. Control parameters. Synthesis main program and line simulation subroutine.

Some key programs for operation of the line analog will be outlined. The core is a subroutine for the actual line simulation which generates the output synthetic speech wave. This is called by a main program reading an input file of synthesis parameters and storing the generated waveform. An editor to generate a phoneme library file can also generate control parameter files for rule synthesis by simple phoneme concatenation. Other auxiliary programs are used for display of parameters and results, including formant frequencies computed from synthesis parameters in the articulatory domain.

7.1 FILE STRUCTURES

The different files involved in the operation of the line analog are compatible to the general purpose interactive ILS program system ((R) Signal Technology Inc, S:ta Barbara, CA) in that sense that they all contain an ILS file header. The purpose of this is primarily to make it possible to use existing ILS programs and subroutines for analysis and display of signals. The header occupies the first 64 16-bit words of the file (IHDR(0) to IHDR(63)) and holds miscellaneous information such as utterance identification etc.

There are two different kinds of files involved. One is files with synthesis parameter data used to control the line simulation, the other is signal waveform files produced as outputs from the simulations.

7.11 SYNTHESIS PARAMETER FILES

A key entry in the synthesis parameter file header is a number NCHAN=IHDR(57) specifying the frame size. All data after the file header is thus arranged in frames, each containing a corresponding number of 16-bit words. Two frames following the file header contain special information for the interpretation of the subsequent data. The parameter file is thus organized as follows:

a. 64 word ILS file header, including value of NCHAN
b. one NCHAN word frame with parameter names
c. one NCHAN word frame with smoothing time constants
d. an arbitrary number of NCHAN word frames with parameter data
e. one end frame containing the number −1 as its first word

Item b. is a list of parameter names where each name is represented by two ASCII characters. The first two parameter names are always MK (for MarK) and DR (for DuRation). Other parameter names are defined by the operator using the parameter file editor, and unused space is padded with zeroes.

Item c. defines individually for each parameter a smoothing time constant in milliseconds. This is used by the synthesizer program. After the data frames have been interpreted the result is smoothed with the time constant given here.

In item d. several consecutive frames may form a group that have an identical value of the marker parameter MK. The value of MK is one or two ASCII characters. It serves as an identifier in the parameter editor program where it may signify for instance a phoneme, a diphone, or any desired sequence of parameter values.

The essential purposes served by this structure are that the file within itself documents its contents, and that the number and order of the data items is arbitrary.

7.12 WAVEFORM FILES

In the waveform files produced by the simulation program the header includes the sampling frequency, codes for the presence of pre-emphasis etc. NCHAN=1 for ordinary speech files, but by application of switches in the simulation program multichannel files can be generated. These may for instance contain oral, nasal, and laryngeal waveforms in separate channels.

7.2 PARAMETER FILE EDITOR

The editor for synthesis parameter files is a revised version of OVCON, originally written for control of OVE cascade speech synthesizers, Liljencrants (1969).

Before the editor is entered the parameter file is assigned as a secondary file in the ILS system. When the editor is entered this file is created if it does not already exist, and opened. The detailed operations of the editor are performed in a work buffer in the computer memory. The parameter file can now be set up and altered in many ways using various single character commands, when applicable preceded by one or more numerical arguments. The more essential of these commands are:

N   Define or delete a parameter name. Parameter names are always two characters, the first of which is to be either A, B, F, G, K, or P. The parameter names MK and DR are not accessible for change.

T   Define smoothing time constant associated with parameter.

M   Define the data interpretation mode. In the time vs parameter domain only certain points are assigned a value. In the so called 'polygon mode' these points are joined by straight lines in the value vs time plane. Alternatively, in the 'square mode', the parameter value of one data point is held constant until the time the next data point is reached.

W   After this command the editor requests a string of characters. The parameter file is scanned for frames having MK parameter values matching those of the string, and the corresponding data frames are copied into a work buffer. The marks in the MK parameter may be single characters of pairs of characters. In the matching process a matching pair takes precedence over a single character match. This way one or several groups of data frames are brought into the work buffer. In this buffer individual parameter values can be accessed in terms of frame number and parameter name, and altered as desired. One application of this command is to generate synthesis by rule control patterns by simple phoneme concatenation.

L   List contents of the work buffer.

P   Plot contents of the work buffer as a set of waveforms in time.

U   Add contents of the work buffer to the parameter file. The command has to be augmented with a one or two character string. This string is inserted as the MK parameter value in all frames from the work buffer.

Using these commands a reference parameter file can be constructed. This file is generally called a library, it may for instance contain typical patterns on a phoneme or diphone basis. Using the W command several of these can be combined in the work buffer. This will then form the control pattern of an utterance of coherent speech. With a special command (R) the work buffer can now be written as a separate parameter file. Later this is used as input to the synthesis program.

Each data frame has a specified value of its duration in the DR parameter. Other parameter values given apply to the beginning of this time interval. An important feature is that parameter values may be unspecified in a frame. This is marked with a negative parameter value; all parameters are assumed to be positive or zero.

The exception to that is the duration parameter. DR is always specified and it may take negative values. This is useful especially to impose a pitch pattern on a phoneme sequence. One may as an example devise a pair of data frames where the first defines the pitch F0 to be 150 Hz and has a duration of 100 ms. The second frame defines F0=120 Hz and has the duration -100 ms which brings us back to the original time. All other parameters are left unspecified. In 'square mode' this is then interpreted as a plateau at 150 Hz and a step down to 120 Hz. When inserted between data frames of phonemes it will do nothing to the mutual timing of these.

7.3 SYNTHESIS MAIN PROGRAM

The main program for synthesis handles service functions like decoding arguments from the operator, performing operations with the parameter and waveform files, and calling a separate subroutine for the line simulation. Its key task is to interpret the synthesis parameter file and construct the corresponding area function. It also contains a routine to plot sequences of area functions.

We will now take a closer look at the interpretation of parameter data. Initially the first two data frames of the parameter file are read. These tell which parameters are present in the input and their corresponding smoothing time constants. The program has another table of parameter names, namely those required for the operations. This is compared to the input parameter names and a parameter pointer table is constructed. This table then tells which word in an input data frame that holds the value of any specific parameter. The reason behind this is that the mutual order of parameters in the input file should be irrelevant, something very helpful to the researcher. Should any required parameter be missing in the input the operator is informed and a corresponding constant default value is taken from a special table.

Another part of the preliminaries is to compute the coefficients of a set of third order z-domain IIR lowpass filters used to smooth the parameter values from the interpretation. These filters are characterized by a conjugate pole pair with an imaginary part equal a nominal cutoff frequency ω_o together with a real pole:

$$s_{1,2} = (-.7 +- j) * \omega_o$$
$$s_3 = -.7 * \omega_o$$

All poles have the same real part. This does to some extent degrade the symmetry of the filter time response, but a desirable effect of giving zero overshoot. For this particular filter the smoothing time constant will be $2.81/\omega_o$ and the delay time $2.13/\omega_o$. The impulse and step responses of this filter are shown in fig 7.1. When the filter coefficients are

set up then a table of 'durations valid' is also made, one entry for each parameter. It is initialized to minus half the filter delay time. This time bias is to make the half value point of the smoothed parameter steps to coincide in time with the unsmoothed. Fig 7.5 shows an example of control parameters interpreted in 'square mode', as a tabulation and as unsmoothed and smoothed functions of time.

' ..er, during simulation, we proceed forward in steps along the time scale. At each such step the entries in the table of 'duration valid' is decremented accordingly. When an entry becomes negative it shows that time has expired for that parameter. The next value and a new interval of validity is then established by searching the input frames. It should be stressed that it is by no means necessary that the different parameters change synchronously, this because of the feature of possibly unspecified data in the frames.

Several parameters define the excitation of the model. So far no more elaborate modelling of the glottis has been implemented, that is, no interaction from the air on the glottal area is considered. Instead the glottis area model suggested by Fant (1979) is used, which inherently approximates some of these effects. At present the following are used to define a forcing function for the glottis area:

FO   (100)   Pitch in Hz
FG   (125)   Glottis characteristic frequency in percent of FO
FK   (100)   Steepness of glottis closure in percent
AV   (0)   Glottal peak area, in dB relative to 0.3 cm^2
AC   (0)   Glottis opening for unvoiced operation, in dB rel 0.3 cm^2

Values in brackets show default values used in absence of a parameter or any value of it. Zero with the dB scale denotes cutoff, all parameters are to take positive values. FG and FK apply to the Fant model and are further illustrated in fig 7.2. AV and AC can mostly be considered as binary on/off signals.

Oral area function data can be defined in two ways. One is by supplying explicit areas for 16 tube sections from glottis to mouth in terms of parameters BO - BF. This is of interest for test purposes and for using area data obtained from X-ray tracings of the vocal tract. Alternatively an oral area function is set up from the assumption of a sinusoidal tongue profile as suggested by Liljencrants (1971). The magnitudes of a sine and a cosine coefficient for the profile are given by parameters FS and FC. The diameter of tube section number n is computed from
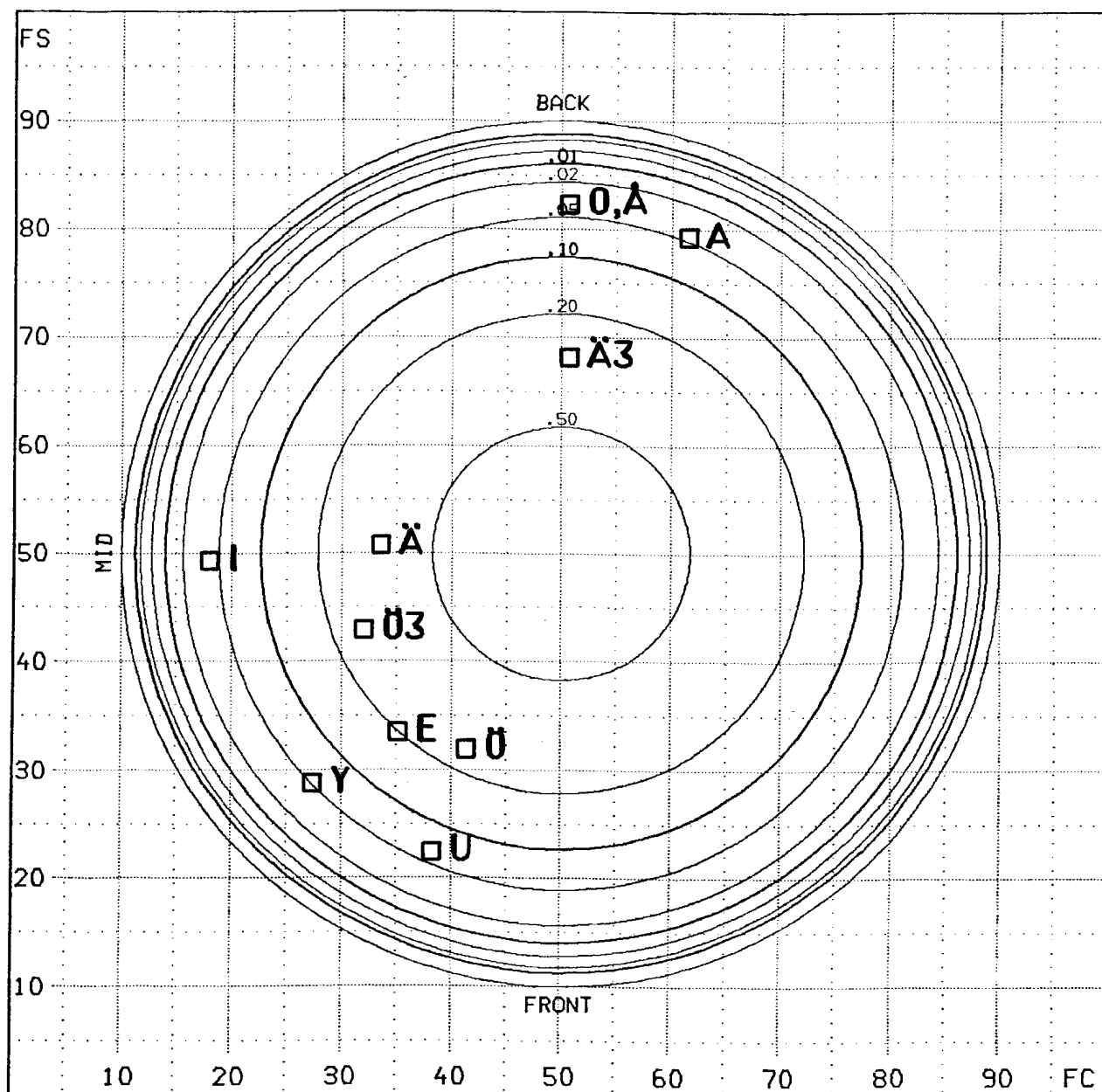
Fig 7.3. Plane of tongue contour Fourier coefficient parameters
FC and FS. Circles represent contours of equal minimum area
relative to a normalizing area. Loci for equal place of
constriction go radially. Symbols show typical positions of some
Swedish vowels.

Fig 7.3

$$\Phi(n) = 2*\pi*(n+.5)/16$$

$$D(n) = 1 - \frac{FS-50}{40} * \sin(\Phi(n)) - \frac{FC-50}{40} * \cos(\Phi(n))/40$$

and the area is then set as

$$AR(n) = ARNOM * D(n) * /D(n)/$$

where ARNOM is a normalizing area, for instance 4 cm^2.

The FS and FC parameters thus have a bias of 50 (percent) units. Deviation from this of more than 40 units in their combined magnitude will give a negative diameter and will represent a cutoff in the oral passage. Effects of variation in the FS and FC parameters in a Cartesian coordinate system are shown in the diagram if fig 7.3. The circles in this are contours of equal constriction area, marked in fractions of the normalizing area.

It is obvious that the same information could just as well be conveyed using a pair of parameters in a polar coordinate system. A phase angle could then represent the place of constriction and a magnitude could represent the degree of constriction.

The polar representation is simpler to handle conceptually but has not been used for the reason of its adverse effect when it comes to interpolation between shapes pertinent to different vowels. In fig 7.3, consider the transition between vowels /a/ and /i/, widely distant in the plot. A straight interpolation in the FS-FC domain will give inter- mediate shapes with less constriction, even approaching a uniform tube in the mid interval. Conversely, an interpolation in polar coordinates would give little difference between the extremes in the radial direc- tion, and an interpolation would follow an arc of almost constant area of the constriction. This corresponds to clearly articulated interme- diate vowels, contrary to what is the case in real speech. The Cartesian representation in the FC and FS coefficients also bears a resemblance in effect to the articulatory coordinates used by Lindblom and Sundberg (1971) or by Coker (1976).

The sinusoidal shape is not used in the extreme pharyngeal region. Instead in the first sections the areas are linearly interpolated between a fixed value at the glottis and that of the fifth section.

Fig 7.4 shows a comprehensive plot of the vocal tract model and how the different parameters control it. In this plot, as well as in the ones produced for monitoring the main program, fig 7.6a, diameters are shown rather than areas. Generally areas are taken as proportional to the diameters squared.
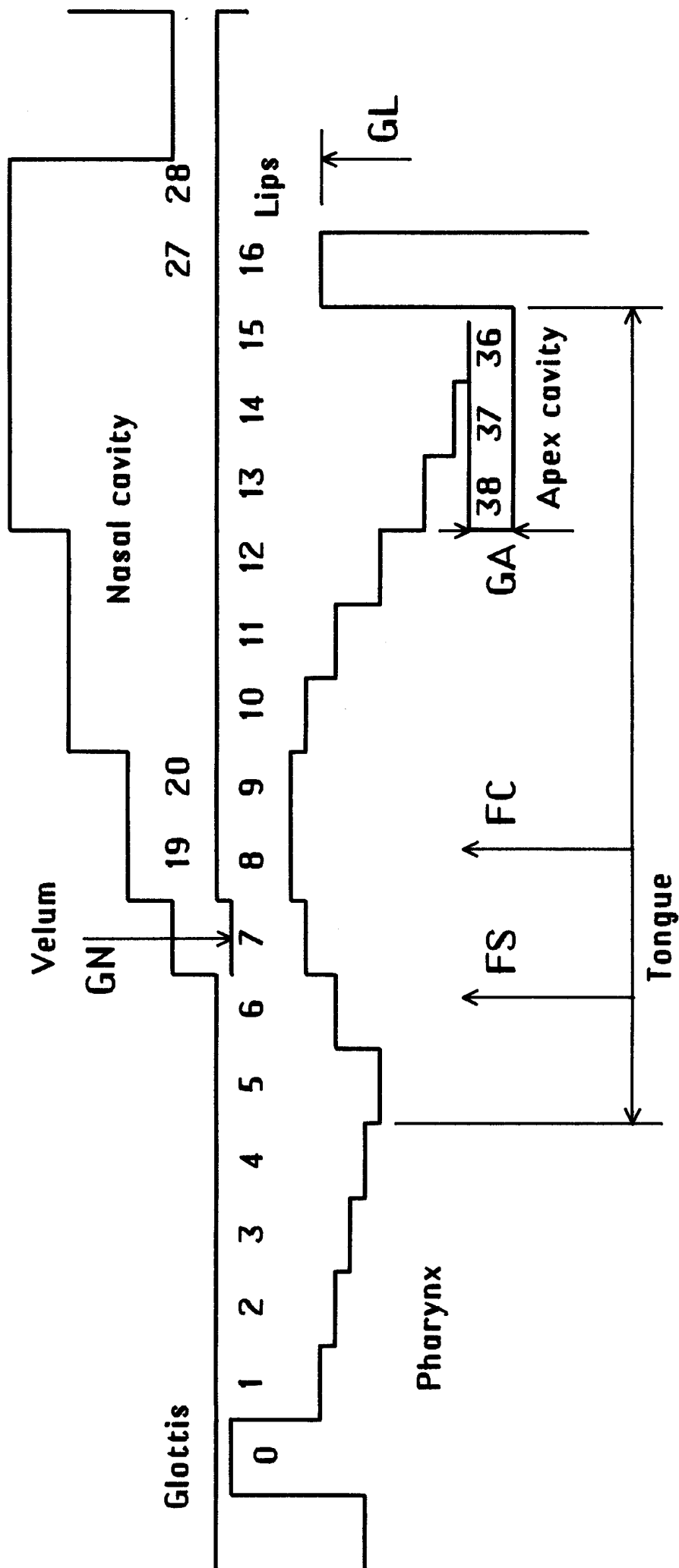
Fig 7.4

Fig 7.4. Layout of vocal tract model showing where the control parameters operate. FS and FC are Fourier coefficients defining the tongue profile. Shapes are in terms of diameters; areas are taken as diameters squared. The velar and apical trapdoor parameters GN and GA operate on a percentage basis on the areas.

Apart from the tongue shape coefficients FS and FC, or the section by section specification in terms of B0 – BF, the vocal tract shape description is supplemented with the following parameters:

GL   (50)   Lip opening
GN   (0)   Nasal gate
GA   (0)   Apex gate

GL controls the area of the last tube section, independently of the tongue shape. It has the same bias and scale factor as the FS and FC coefficients, that is, 50 gives a nominal area and 10 gives the limit of closure. GL has a marked influence on the first formant and also controls the magnitude of the lip radiation reactance, that is, the end correction.

GN controls a trap door corresponding to velum, it opens the passage into the nasal tract. The nasal tract has a fixed shape defined in a table of areas and the velar valve connects it to tube 6 of the vocal tract. GA controls another valve at tube 15 opening into a closed three-section line. This is to simulate a cavity under the apex of the tongue.

The area function is derived from the parametric input using a time step of 5 ms. Once the areas of the complete vocal tract have been calculated control is given to the line simulation subroutine that will compute a corresponding sequence of waveform samples. The glottis area will of course have to be evaluated on a sample by sample basis, so this is left to a special section in the simulation routine to which the excitation parameters are passed.

7.4 LINE SIMULATION SUBROUTINE

In the line simulation subroutine the signal waveforms are computed after a number of supplementary operations. In earlier sections several topics have been covered mostly with a theoretical view. Here we will now deal with some additional practical aspects on their application.

The input is here the excitation parameters and the complete set of tube section areas, except the area of the glottis section. We are going to compute a sequence of 80 time samples corresponding to 5 ms at 16 kHz sampling. A major part of the operations of course have to be performed each sample interval, including determination of the glottis area and loss coefficient.

The reflection and loss coefficients vary slowly in our pseudo-static

system, so these can be recomputed at longer intervals. It was soon found that a 5 ms update interval is far too long, especially when opening or closing a tube section where the relative step in area can be very large. Hence the area function as delivered from the main program is interpolated in smaller steps. A suitable area update interval appears to be 4 to 8 samples, this defines the area under-sampling factor. It is hardly worthwhile trying to expand the update interval much further since the extra saving of computer time will then be less significant.

A convenient way to interpolate is to use a simple first order lowpass filtering of the input area AR(n) to obtain the smoothed

ARS(n) := (AR(n)-ARS(n))/NAV + ARS(n)

as further commented with (5.22). Observe that the averaging time constant is now multiplied with the under-sampling factor. One could of course just as well use some other interpolation, for instance linear.

This is now the time to impose an important limiting on permissible areas. Clearly zero or negative areas will cause severe numerical trouble in the computations fortcoming. The conventional solution is to limit areas downward to a small minimum, let us say .1 mm^2. Provided losses are taken in account this will give a good enough cutoff in the line with laminar losses dominating.

Here we can explain why for instance oral cutoff with GL is defined at the value 10 rather than 0. The input data may well tend toward a negative area. After the smoothing the control parameter will then reach zero with possibly some speed. But if the unsmoothed input gives zero area, then the smoothing will reach this only asymptotically and you tend to lose control of the exact moment of closure, it can even happen that you may never reach it. The essential thing is thus that the limitation of area is done after the smoothing.

Once the final area function is established we can proceed with the dynamic adjustment of partial waves as in section 1.4. Also the set of reflection coefficients are computed with (1.15), and the special reflection coefficients required at the branches with (1.71).

With (3.5) the coefficients for the lip radiation impedance are settled. In my experience this must be done every sample interval, an under-sampling gives spurious sampling frequency subharmonics in the output when the lip area changes rapidly.

A special array of short time average particle velocity is also maintained. When areas are updated I also compute the net flow in each section, divide it by the area and make a running average by first order lowpass filtering as above. These data and the areas are used to

evaluate the different series loss factors using (2.26) and appropriate values from table 2.1. The larger of these form the result in a set of loss coefficients, one for each tube section. Care is taken to include the jet loss only when both the neighbouring areas are greater than the current.

At this stage all information is available to excecute the scattering computations and the corrections. At appropriate places the three-way branches are treated. To simplify the program structure the oral and nasal parts may be arranged and processed as one long linear array. Afterwards the specific output waves at the branches are overwritten with the results of the branch scattering equations for which the inputs are still untouched. This is done in two passes, one for the even and one for the odd junctions as outlined in section 1.6. Then the z-domain recursions for wall impedance and lip radiation are excecuted.

Finally the narrowest place along the vocal tract line is located and the particle velocity here is established. I keep a special (one-dimensional) short time average of this constriction velocity, disregarding in which tube it may occur. Using (5.23) I find a noise amplitude, and if applicable a corresponding random number is added to one of the partial waves as elaborated in section 5.

Closing operations in the simulation subroutine is to collect various values as selected by operator switches and put them into buffers for transfer to output files.

MK	DR	FO	GN	FK	GL	FG	FC	FS	GA	AV	AL	AC
		150	20	50	50	50	50	50	10	10	100	10
1	0	100		100		125					60	
1	0				90		66	66	0	32		0
1	40		0									
1	50											
1	0		0		20		75	24	0	0		36
1	70		0				75	24		0		
1	0				90		22	42	0	32		0
1	50		0									
1	40											
1	0		0				35	1	0			
1	25									0		32
1	50		0		50		35	1		0		
1	0				42		50	82	0	32		0
1	50		0									
1	90											
1	-50		50									0
1	50				50		66	12	100	32		0
1	70											0
1	0				50		50	50	0	32		0
1	50		0									
1	50											
1	30	100		0			50	50	0	0		0
1	30											
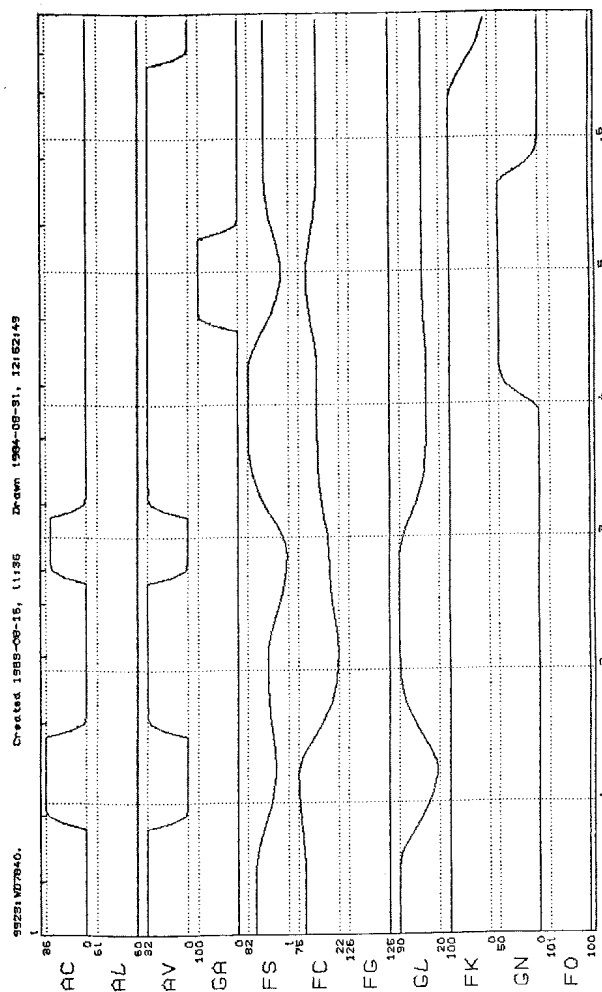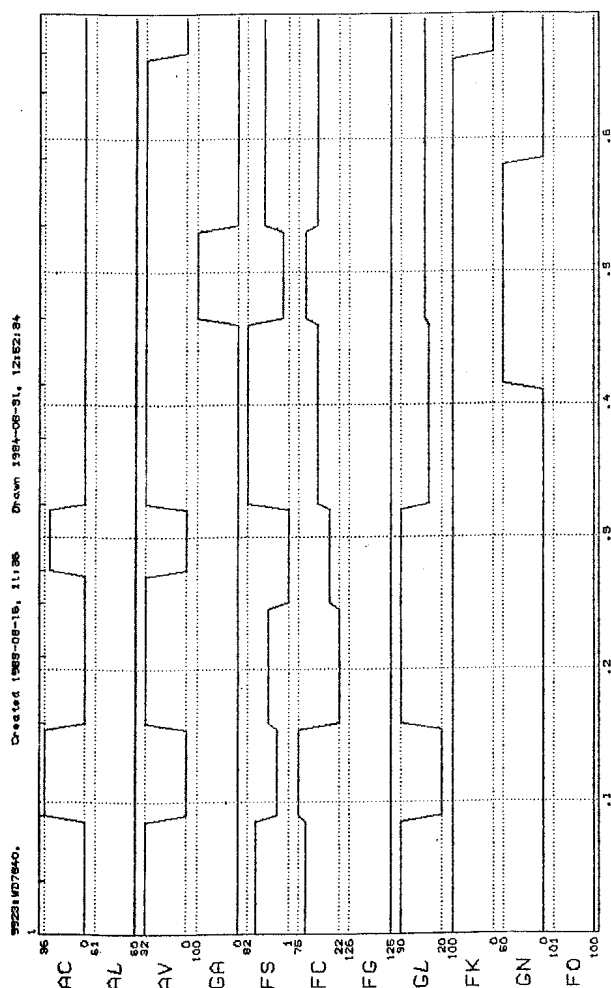
Fig 7.5. Tabulation of synthesis parameter file with one frame per line, and time waveforms of interpreted control parameters, unsmoothed and smoothed.

The parameters define a nonsense utterance /asiton/ with constant pitch, and these data are coherent with figs 7.6 – 7.8.

Fig 7.5

a.

b.

Ug

P1

Uw

1-1/X
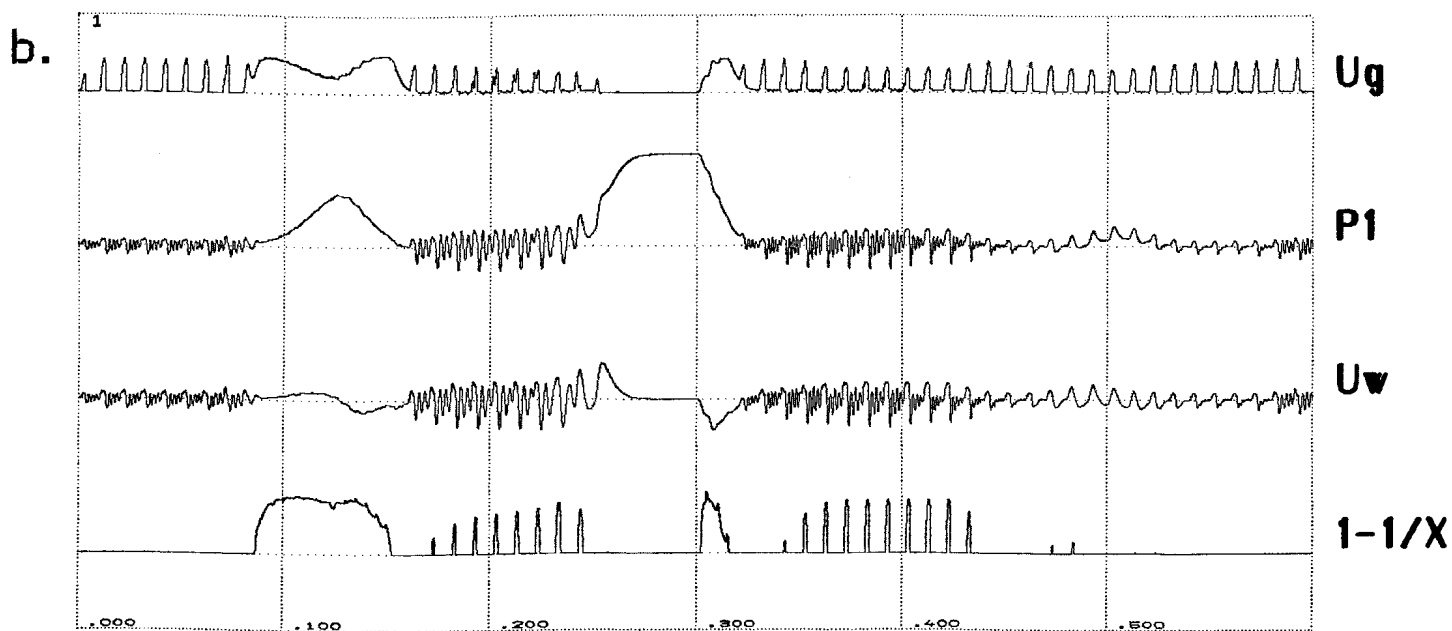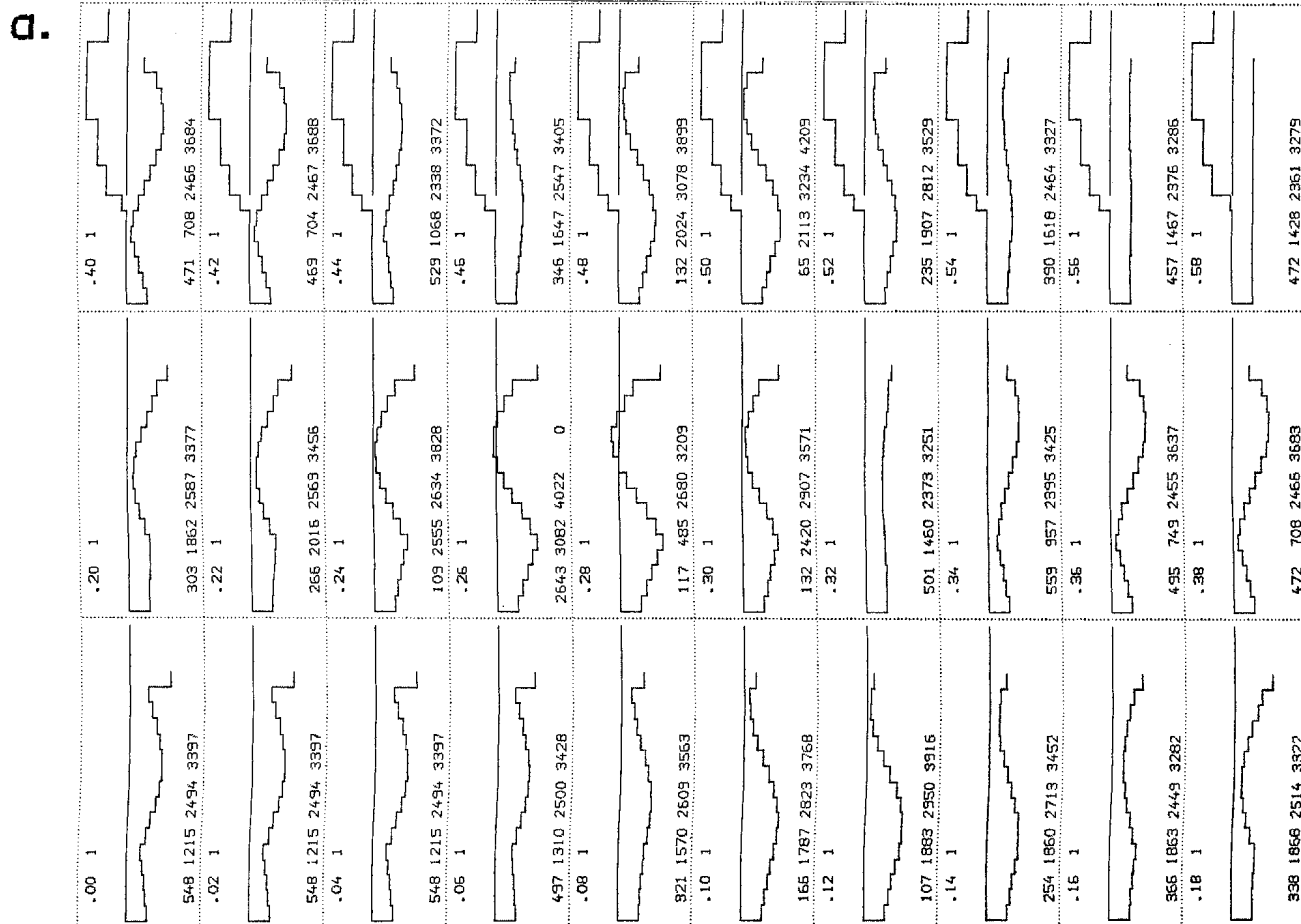
Fig 7.6. Diagnostic plots from synthesis of nonsense word /asitone/.

a. Vocal tract model shape at time increments of 20 ms.

b. Glottal flow, intraoral pressure, wall flow, and noise conversion factor. During the /s/ the supplying flow is partially pinched off, essentially without resuction in resulting noise level (see fig 7.7). At the /t/ release the wall flow contributes the air charged in the wall stiffness. Noise bursts come periodically in some vowels.

Fig 7.6
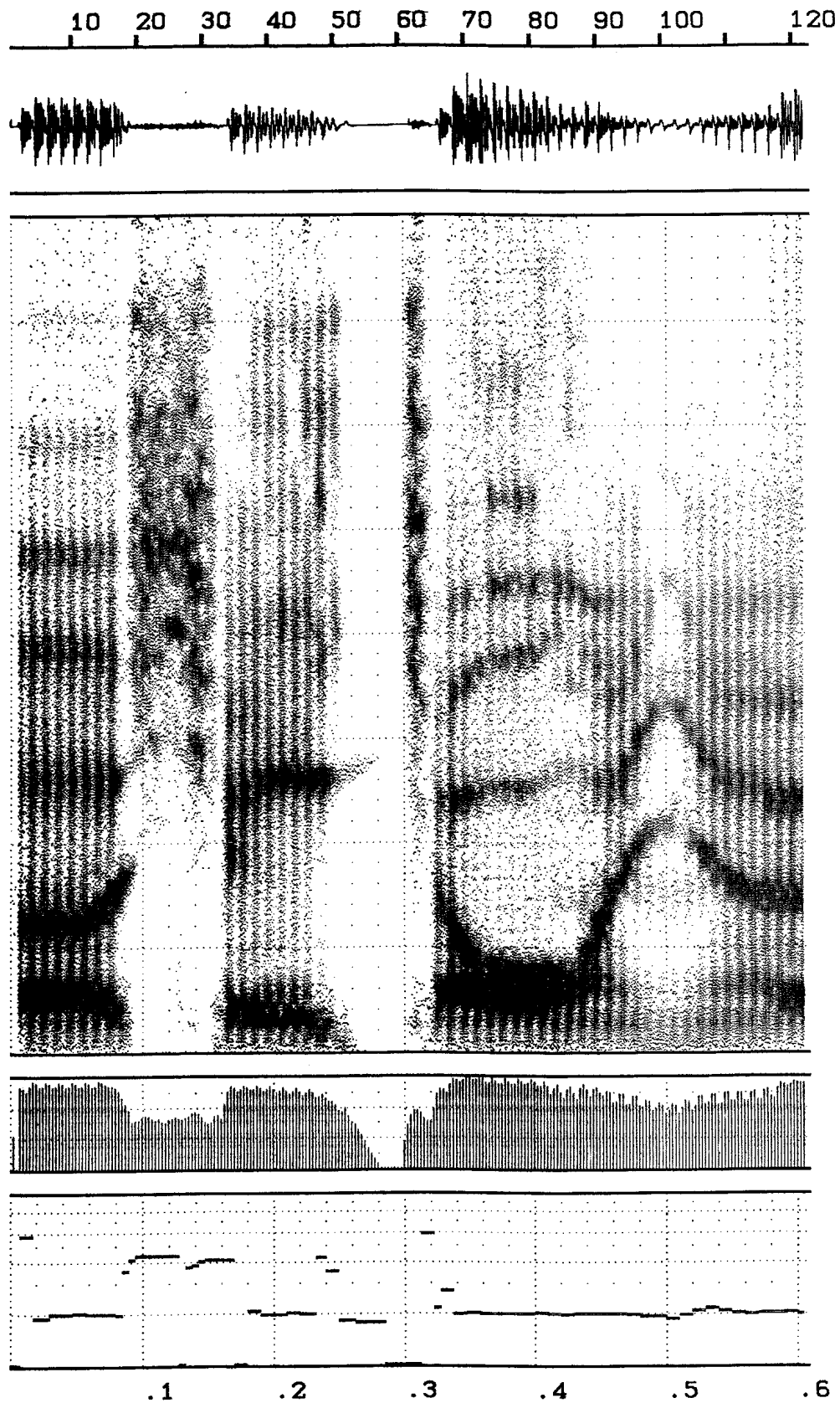
Fig 7.7. Computer spectrogram. Bottom traces are level and pitch on logarithmic scales.

Fig 7.7
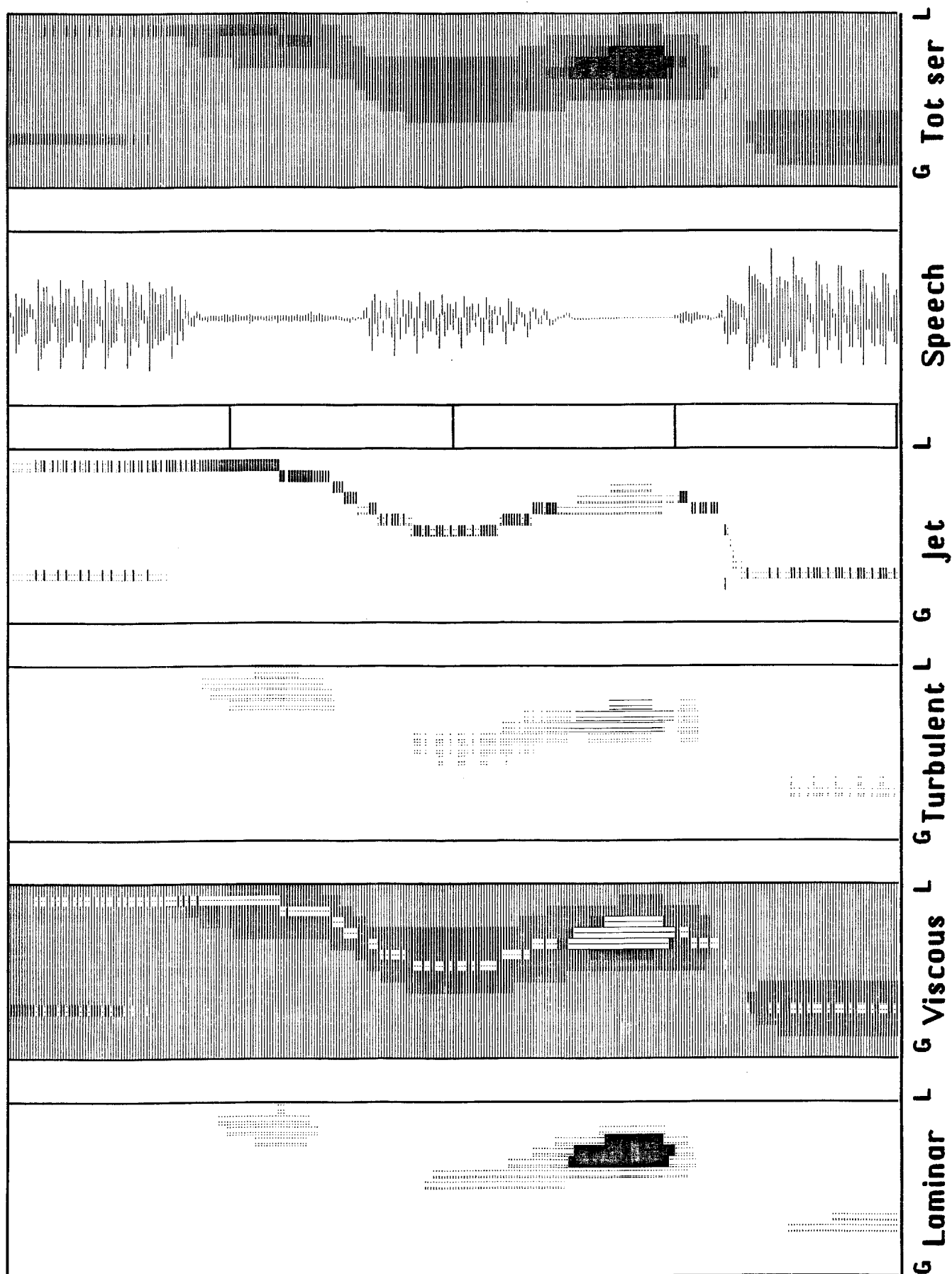
Fig 7.8. Spatial distribution of series loss components vs time.
L and G indicate lip and glottis ends. At top the total losses,
then the speech waveform followed by the individual loss
components from jet loss, turbulence, viscosity, and laminar
flow. In the lower four, horizontally dotted loss components are
overridden by one or more of the others.

Fig 7.8

8. CONCLUSIONS

Several physical phenomena relevant to speech synthesis have been reviewed. They were formalized in a homogeneous manner in order to make it easy to apply the results to the Kelly-Lochbaum reflection line analog. The simple model has been augmented with a number of tricks to render a comprehensive research tool for simulation studies of the vocal tract. As many as possible of those tricks have been deliberately put into the form of (mostly small) corrections to the classical static scattering equations. One reason for this approach is that it makes it particularly easy to connect and disonnect features for research studies. The specific signal flow and storage pattern allows scattering computations for one instant of time to be followed by one or more rounds of additive correction computations.

Modifications to the scattering equations to account for dynamic changes in the areas were discussed and a novel simple method was suggested. One type of dynamic scattering equations, based on continuity in pressure and flow, tends to diminish spurious signals due to undersampling of area data. Such undersampling can be done, but with certain care to avoid interference with the delay time though the line. The other type is based on continuity in force and velocity which implies a physical consistency and makes it capable of handling bandwidth changes due to area time variation. The relevance of this for speech synthesis however appears to be somewhat marginal. This model is more sensitive to area undersampling.

The much neglected analogy of flow waves rather than pressure waves has a conceptual advantage and generally gives a smaller range of numerical variations, but proper dynamic and loss corrections are imperative for handling area cutoff situations.

The reflection principle applied for glottal flow modulation is demanding because of the high speed area change. This would require a dynamic model with f-v continuity, but specifically because of the glottal area being relatively small the static pressure wave model happens to perform approximately as well.

Parallel branches to the line makes it possible to implement the nasal sounds and their complex patterns of pole-zero pairs in addition to the all-pole characteristics of the single line. Also shunting cavities and lines can be introduced.

Series losses are important in that they make it possible to cut off the DC air flow in a controlled manner. Without them other sounds than pure vowels are difficult to handle. The laminar flow resistance controls cutoff while the frequency dependent viscous resistance regularly influences formant bandwidths in the mid-frequency range. It is easily overshadowed by the non-linear "jet loss" when marked constrictions

occur. Losses due to turbulent flow appear to be mostly negligible, they are normally overridden by one of the other three.

The lip radiation impedance is the dominant controller of high frequency formant bandwidths. It also accounts for considerable formant frequency shifts, dependent on the lip area and the associated end correction. A technique of interfacing the reflection line analog to a radiation impedance model formulated in the z domain was illustrated.

Shunt losses are dominated by the vibrations of the vocal cavity wall. The major effect of this is on the first formant where it elevates the frequency and often dominates the bandwidth. The losses from direct radiation into the walls are negligible. Heat conduction loss is of little consequence to the sounds produced, but can have a function in damping out high frequency waves within closed cavities. The wall impedance model used is also defined in the z domain. It contains a capacitive element to handle air storage under internal pressure buildup in a closed vocal tract.

Automatic generation of noise at constrictions was theoretically reviewed and a practical procedure was defined.

9. BIBLIOGRAPHY

Ananthapadmanabha T V, Fant G (1982): Calculation of true glottal flow and its components. STL QPSR 1/1982, 1-30. Speech Comm 1, 167-184.

Beranek L L (1954): Acoustics, McGraw-Hill NY.

Coker C (1976): A model for articulatory dynamics and control.
IEEE Proc 64.4.

Degryse D (1981): Temporal simulation of wave propagation in the lossy vocal tract. FASE 4.

Fant G (1960): Acoustic theory of speech production.
Mouton, S'Gravenage, 2nd ed 1970.

Fant G (1972): Vocal tract wall effects, losses, and resonance bandwidths. STL QPSR, 2-3/1972, 28-52.

Fant G (1975): Vocal-tract area and length perturbations.
STL QPSR, 4/1975, 1-14.

Fant G (1976): Vocal tract energy functions and non-uniform scaling.
Acoust Soc Jap, Univ Tokyo.

Fant G (1980): The relations between area functions and the acoustic signals. Phonetica 37, 55-86.

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

Fettweis A (1971): Digital filter structures related to classical filter networks. Arch Elektr Uebertragung, 25-2, 79-89.

Fettweis A, Meerkötter K (1975): On adaptors for wave digital filters.
IEEE Tr ASSP-23, 516-525.

Flanagan J L (1965): Speech synthesis, analysis, and perception.
Springer.

Flanagan J L, Ishizaka K (1976): Automatic generation of voiceless excitation in a vocal-cord/vocal-tract speech synthesizer.
IEEE Tr ASSP-24, 163-170.

Flanagan J L, Ishizaka K, Shipley K (1975): Synthesis of speech from a dynamic model of the vocal cords and vocal tract. BSTJ 54, 485-506.

Goldstein S ed (1938): Modern developments in fluid dynamics.
Clarendon Press, Oxford. 2nd ed Dover Publications, NY 1965.

Heinz J M (1957): Model studies of the production of fricative consonants. MIT RLE QPR: oct-dec 1956, jan-mar 1957, jul-sep 1957

Ingard U (1953): On the theory and design of acoustic resonators.
JASA 25-6, 1037-1061.

Ingard U, Ising H (1967): Acoustic nonlinearity of an orifice.
JASA 42-6.

Ishizaka K, French J C, Flanagan J L (1975): Direct determination of vocal tract wall impedance. Trans IEEE ASSP-23 No 4, 370-373.

Jospa P (1977): Consequences acoustiques des deformations dynamiques du conduit vocal. In: Carre R, Descout R, Wajskop M, editors: Articulatory modeling and phonetics, GALF, 49-64.

Kabasawa Y, Ishizaka K (1980): Speech production system based on traveling wave representation considering vocal tract losses.
Univ Tokyo.

Kelly J L, Lochbaum C C (1962): Speech synthesis.
ICA 4, G42, 1962, and SCS, Stockholm.
Reprinted in: Flanagan J L, Rabiner L R, editors: Speech Synthesis, Dowden/Wiley 1973

Laine U K (1982): Modelling of lip radiation impedance in the Z-domain. IEEE Proc ICASSP.

Lherm B (1984): Realisation d'un outil de simulation harmonique et de synthese temporelle du type analogue du conduit vocal. Thesis, Grenoble.

Liljencrants J (1969): Speech synthesizer control by smoothed step functions, STL QPSR 4/1969, 43-50.

Liljencrants J (1971): Fourier series description of the tongue profile. STL QPSR, 4/1971, 9-18.

Liljencrants J, Fant G (1975): Computer program for VT-resonance frequency calculations. STL QPSR, 4/1975, 15-20.

Lindblom B, Sundberg J (1969): A quantitative model of vowel production and the distinctive features of Swedish vowels. STL QPSR, 1/1969, 14-32.

Maeda S (1977): On a simulation method of dynamically varying vocal tract: reconsideration of the Kelly-Lochbaum model. In: Carre R, Descout R, Wajskop M, editors: Articulatory modeling and phonetics, GALF, 281-288.

Maeda S (1982): A digital simulation method of the vocal-tract system. Speech Comm, 1, 199-229.

Mermelstein P (1973): An articulatory model for the study of speech production. JASA 53.4.

Meyer P, Strube H W (1984): Calculations on the time varying vocal tract. Manuscript TBP.

Meyer-Eppler W (1953): Zum Erzeugungsmechanismus der Geräuschlaute. Zeitschrift fuer Phonetik, Vol 7, 196-212.

Morse M P, Ingard K U (1968): Theoretical Acoustics. McGraw-Hill, New York

Möller J, Strube H W, Kretschmar J (1977): Measurement and acoustic estimation of articulatory parameters. In: Carre R, Descout R, Wajskop M, editors: Articulatory modeling and phonetics, GALF, 289-296.

Portnoff M R (1973): A quasi-one-dimensional digital simulation for the time-varying vocal tract. Thesis, MIT.

Rubin P, Baer T, Mermelstein P (1979): An articulatory synthesizer for perceptual research. Haskins Laboratories, SR-57, 1-15.

Ruiz P M (1971): A digital simulation of the time-varying vocal tract. JASA 49, 123, abstract.

Schlichting H (1958): Grenzschichttheorie. G Braun, Karlsruhe.

Shadle C (1983): Turbulence noise in the vocal tract. ICA.

Stevens K N (1971): Airflow and turbulence noise for fricative stop consonants: static considerations. JASA 50.4(2), 1180-1192.

Strube H W (1982): Time-varying wave digital filters and vocal tract models. IEEE Proc ICASSP, 923-926.

Strube H W (1982): Time-varying wave digital filters for modeling analog systems. IEEE Trans ASSP-30 no 6.

Titze I R (1983): Parametrization of the glottal area, glottal flow, and vocal fold contact area. University of Iowa, manuscript TBP.

Van den Berg JW, Zantema J T, Doornenbal P (1957): On the air resistance and the Bernoulli effect of the human larynx. JASA 29.5, 626-631.

Wakita H, Fant G (1978): Toward a better vocal tract model. STL QPSR, 1/1978, 9-29.

ACKNOWLEDGEMENT

APPENDIX

In this version legend of table II-A-III is corrected and a missing word on page 12, line 2 from bottom, is added.

FOURIER SERIES DESCRIPTION OF THE TONGUE PROFILE *

J Liljencrants

Introduction

In the study of the movements of the articulatory organs the sagittal projection is often used. Such images can for instance be obtained by x-ray photography.

The purpose of this report is to show how data on the tongue profile obtained that way may be described and modelled in terms of a Fourier series. Capitalizing on the generally smooth shape of the tongue, the description is compact, and the accuracy can be selected with the number of coefficients included. The method can thus be conveniently used both to describe profiles of live speakers as well as for operating simplified artificial models.

Attention is given to some practical aspects like placement of the coordinate system, and interpretation of the Fourier coefficients.

Data collection

The material investigated is a set of x-ray photographs of two speakers, each uttering ten different sustained vowels. Some processing has also been made on material with the same persons singing, both subjects being educated singers. The material was courteously supplied by Sundberg and Lindblom and is the same as used by them for other investigations, see ref. (1).

The mid-sagittal contour of the tongue body was traced from the photographs and put on a coordinate system, as shown in Fig. II-A-1. The system has a polar part covering the oral cavity and a Cartesian part for the pharyngeal region. The contour was sampled at $10°$ intervals in the polar system and at 5 mm intervals in the Cartesian. The data was given to the computer as a matrix with a 30-point element for each vowel. In general the profile occupied samples number 2 to 28. The remaining samples number 0,1, and 29 were filled in with values linearly interpolated between samples number 28 and 2.

All coordinate values in this study are in cm. The scale pertains to the x-ray photographs which are larger than the subjects by a factor of 1.2.
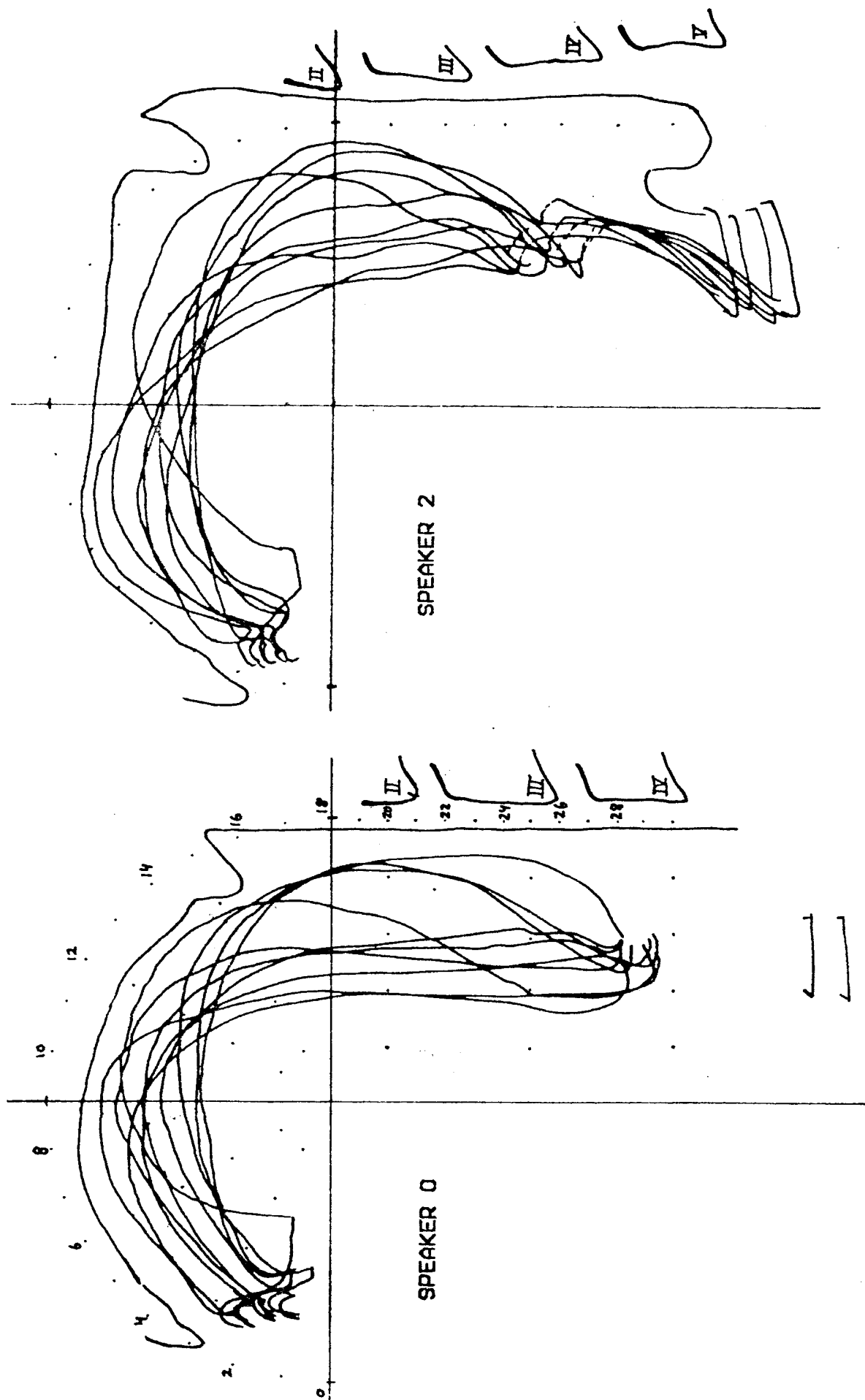
Fig. II-A-1.   Tracings of tongue profiles of two speakers uttering different vowels. In
the left figure the coordinate system and sample numbers are indicated.

To arrive at a suitable placement of the coordinate system the different tracings for one subject were aligned with respect to the upper incisors and the frontal protrusion of the second vertebra. Then the origin was placed at a visually determined center to the oral part of the tongue profile and the vertical axis was made approximately parallel to the rear pharyngeal wall.

Fig. II-A-2 graphically shows the input data for one subject arranged in a linear manner. The horizontal coordinate is the sample number. The vertical coordinate is the radius in the polar system of the mouth, and the distance from the vertical axis in the pharyngeal system.

Different subjects and placements of the coordinate systems are indicated with speaker codes in the illustrations as explained in Table II-A-I. Also the vowels used are indicated with codes as given in Table II-A-II.

Philosophy

By inspection of Figs. II-A-1 and II-A-2 we can make some elementary observations. First we see that the mean value of the "excursion coordinate", plotted vertically, for the different profiles does not change significantly between vowels. This can be regarded as a consequence of the "conservation of mass" in the tongue body. Of course this statement has only a limited validity since the sagittal third dimension has been ignored. Secondly, the strong coherence between successive samples is apparent, that is, the geometrical magnitude of the fine structure is much smaller than that of the overall gross shape variation. Many of the shapes have a strong resemblance to a sinusoid. These are the reasons for the proposition of this experiment, that the shape could efficiently be described in terms of a Fourier series.

To test this a number of computations have been performed.

Basic processing

The profiles were analyzed into a number of Fourier coefficients. For convenience a representation in terms of sine and cosine terms was selected:

$$CC_r = \frac{2}{N} \sum_{n=0}^{N-1} Y_n * \cos(2\pi rn/N)$$

$$CS_r = \frac{2}{N} \sum_{n=0}^{N-1} Y_n * \sin(2\pi rn/N)$$

$$\left. \begin{array}{c} \\ \\ \end{array} \right\} \quad r \neq 0$$

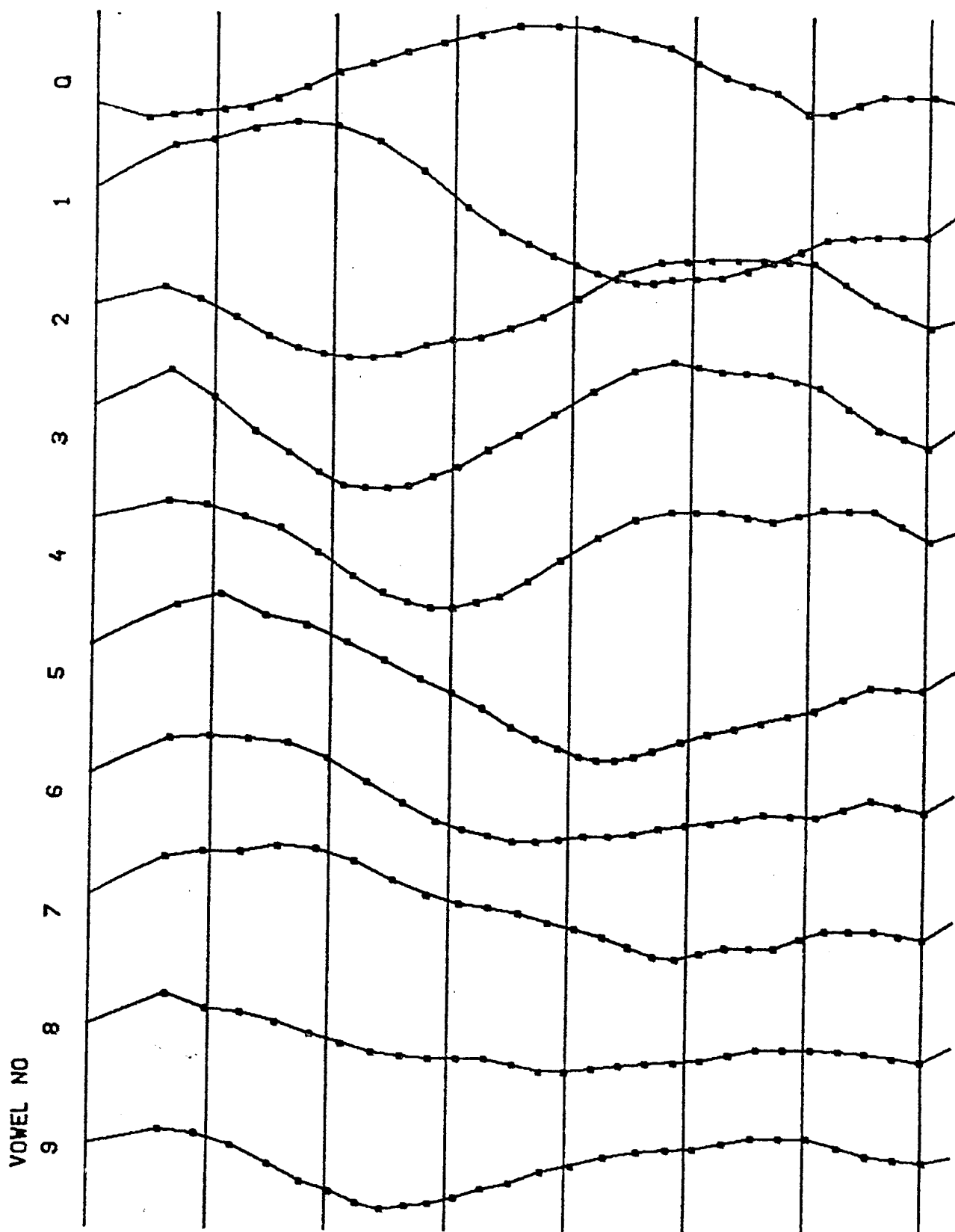$$CC_0 = \frac{1}{N} \sum_{n=0}^{N-1} Y_n$$

$$CS_0 = 0$$

Fig. II-A-2. Tongue profiles on a Cartesian coordinate system. The data
has been normalized using the distance along the contour as the
abscissa. Dots indicate points in the original coordinate system.

Table II-A-I.

Speaker codes in the illustrations

Speaker	Comments	
0	Subject RL,	fixed coordinate system
2	Subject JS,	fixed coordinate system
5	Subject RL,	moving coordinate system, following the mandible
6	Subject RL,	fixed coordinate system, translated .5 cm upwards and .5 cm forwards as compared to speaker 0

Table II-A-II.

Vowel codes in the illustrations

Vowel	IPA
0	u
1	i
2	α
3	o
4	oe
5	ʉ
6	ø
7	e
8	ε
9	œ

Table II-A-III.

Inter-speaker correlation coefficients

	CC_0	CC_1	CS_1	CC_2	CS_2
Speakers 0, 2	0.794	0.895	0.957	0.365	0.637
Speakers 0, 5	0.977	0.987	0.956	0.882	0.968
Speakers 0, 6	0.961	0.948	0.994	0.650	0.970

Y_n are the ordinate values of the N samples from the original profile. CS_r and CC_r are the sine and cosine Fourier coefficients in the single-sided (towards positive spatial frequencies) line spectrum of the periodic continuation of the shape.

Alternatively the Fourier coefficients may be represented with magnitude and phase as

$$R_r = \sqrt{CC_r{}^2 + CS_r{}^2}$$

$$\varphi_r = atan(CS_r/CC_r)$$

Having computed the Fourier coefficients the shape can be reconstructed using the inverse transform. Let us then include only the lower terms, up to number U:

$$Y_n{}' = CC_0 + \sum_{r=1}^{U} \left(CC_r * cos(2\pi rn/N) + CS_r * sin(2\pi rn/N) \right)$$

The omission of the higher terms is equivalent to a spatial lowpass filtering, and only the gross features of the original shape will be preserved. Thus the computation will render an approximate model of the contour. The accuracy will depend on the number of coefficients included. For $U = N/2 - 1$ there will be no approximation, the re-synthesized contour will match the original exactly.

In Figs. II-A-4 and II-A-5 the input data are shown as small squares and the resynthesized data, using $U = 2$, are drawn as solid lines. At the lower left corner of each plot the Fourier cosine and sine coefficients are given. Also, preceded by an E, is the RMS deviation between the data and the model. This error represents the total "power" of all the higher harmonics, not used in the re-synthesis.

As can be seen in the plots the error is in general quite small, even using the DC term and the fundamental only (Fig. II-A-3): This is a primary indicator that the Fourier expansion is an efficient descriptor of the profile shape.

It may be meaningful to measure the fidelity of the Fourier model to the input data in terms like harmonic distortion. If the fundamental is the highest harmonic used this distortion can be defined in a conventional manner:

$$d = \left(\sum_{r=U+1}^{N/2-1} R_r{}^2 \right)^{1/2} * \left(\sum_{r=0}^{N/2-1} R_r{}^2 \right)^{-1/2}$$

Using $U = 2$. If also some harmonics higher than the fundamental /are/ incorporated in the model we can use U greater than 2.

Fig. II-A-3 - 7.   Input data (dots) and Fourier model (solid) for the various speakers and a set of 10 vowels. (Cf. Tables II-A-1 and II-A-2). Legend at lower left in each plot shows vowel number, RMS error, and cosine and sine coefficients.
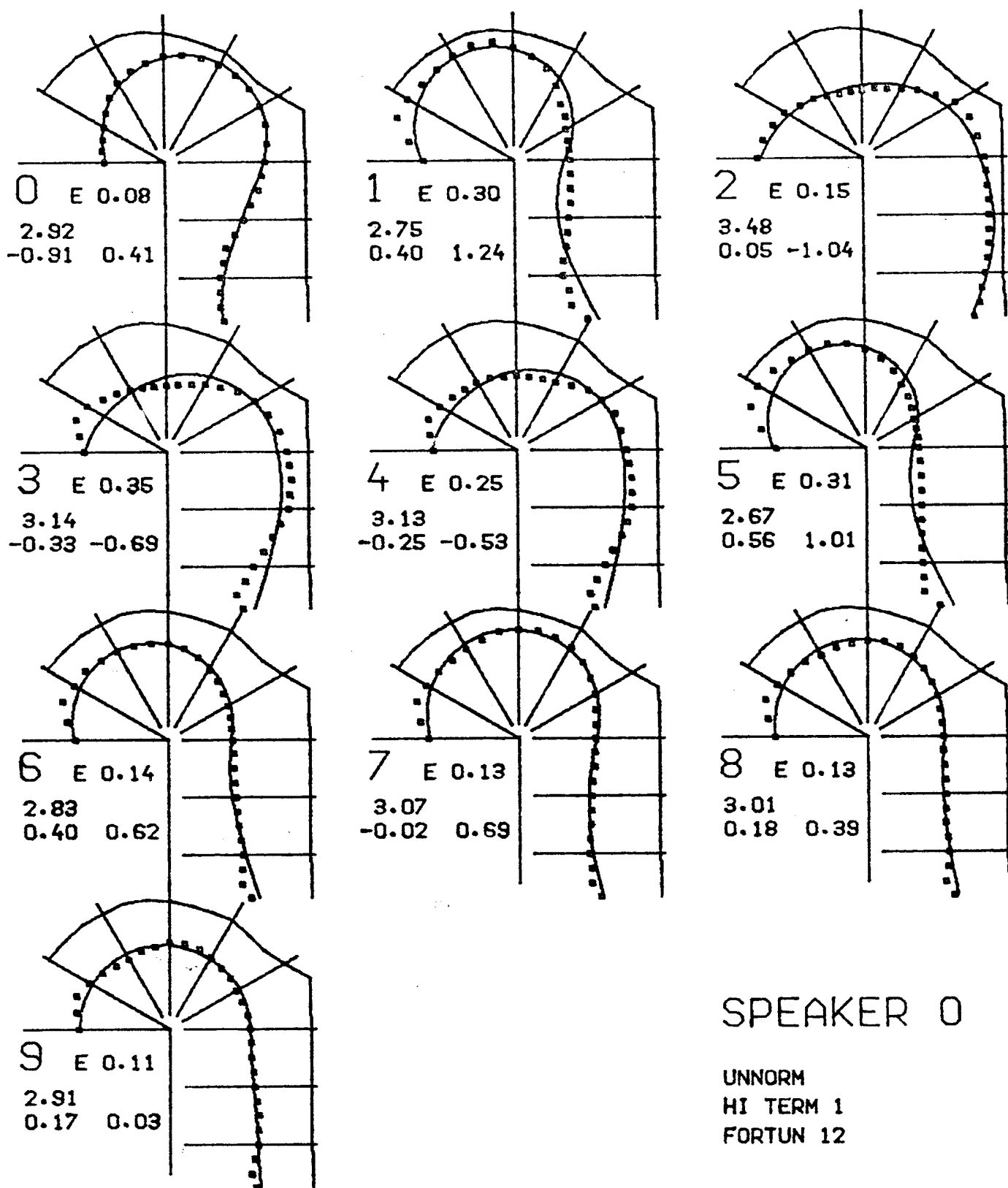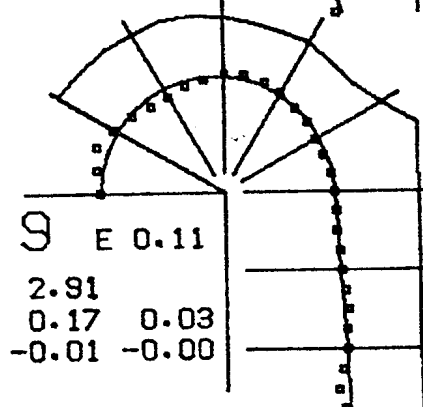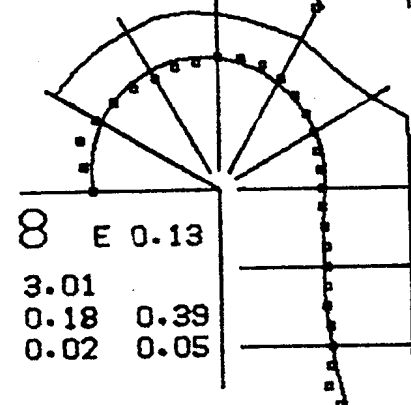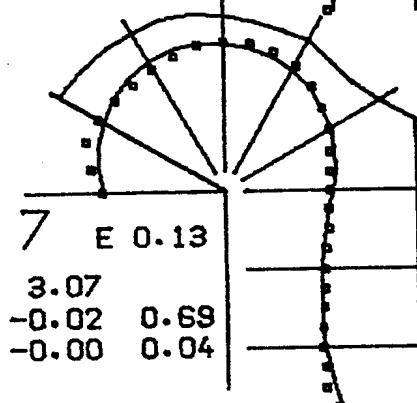
0   E 0.08
2.92
-0.91   0.41

1   E 0.30
2.75
0.40   1.24

2   E 0.15
3.48
0.05 -1.04

3   E 0.35
3.14
-0.33 -0.69

4   E 0.25
3.13
-0.25 -0.53

5   E 0.31
2.67
0.56   1.01

6   E 0.14
2.83
0.40   0.62

7   E 0.13
3.07
-0.02   0.69

8   E 0.13
3.01
0.18   0.39

9   E 0.11
2.91
0.17   0.03

SPEAKER 0

UNNORM
HI TERM 1
FORTUN 12

Fig. II-A-3.

0 E 0.05

2.92
-0.91 0.41
-0.02 0.07

1 E 0.17

2.75
0.40 1.24
-0.19 0.32

2 E 0.05

3.48
0.05 -1.04
0.17 0.10

3 E 0.15

3.14
-0.33 -0.69
-0.00 0.46

4 E 0.12

3.13
-0.25 -0.53
-0.06 0.32

5 E 0.15

2.67
0.56 1.01
-0.13 0.37

6 E 0.11

2.83
0.40 0.62
-0.05 0.13

7 E 0.13

3.07
-0.02 0.69
-0.00 0.04

8 E 0.13

3.01
0.18 0.39
0.02 0.05

9 E 0.11

2.91
0.17 0.03
-0.01 -0.00

SPEAKER 0

UNNORM
HI TERM 2
FORTUN 12

Fig. II-A-4.

Fig. II-A-8 - 11. Fourier coefficients for the profiles of Figs. II-A-4 - II-A-7.
Digits denote vowel number and are placed according to the
coefficients of the spatial fundamental. Vectors extending
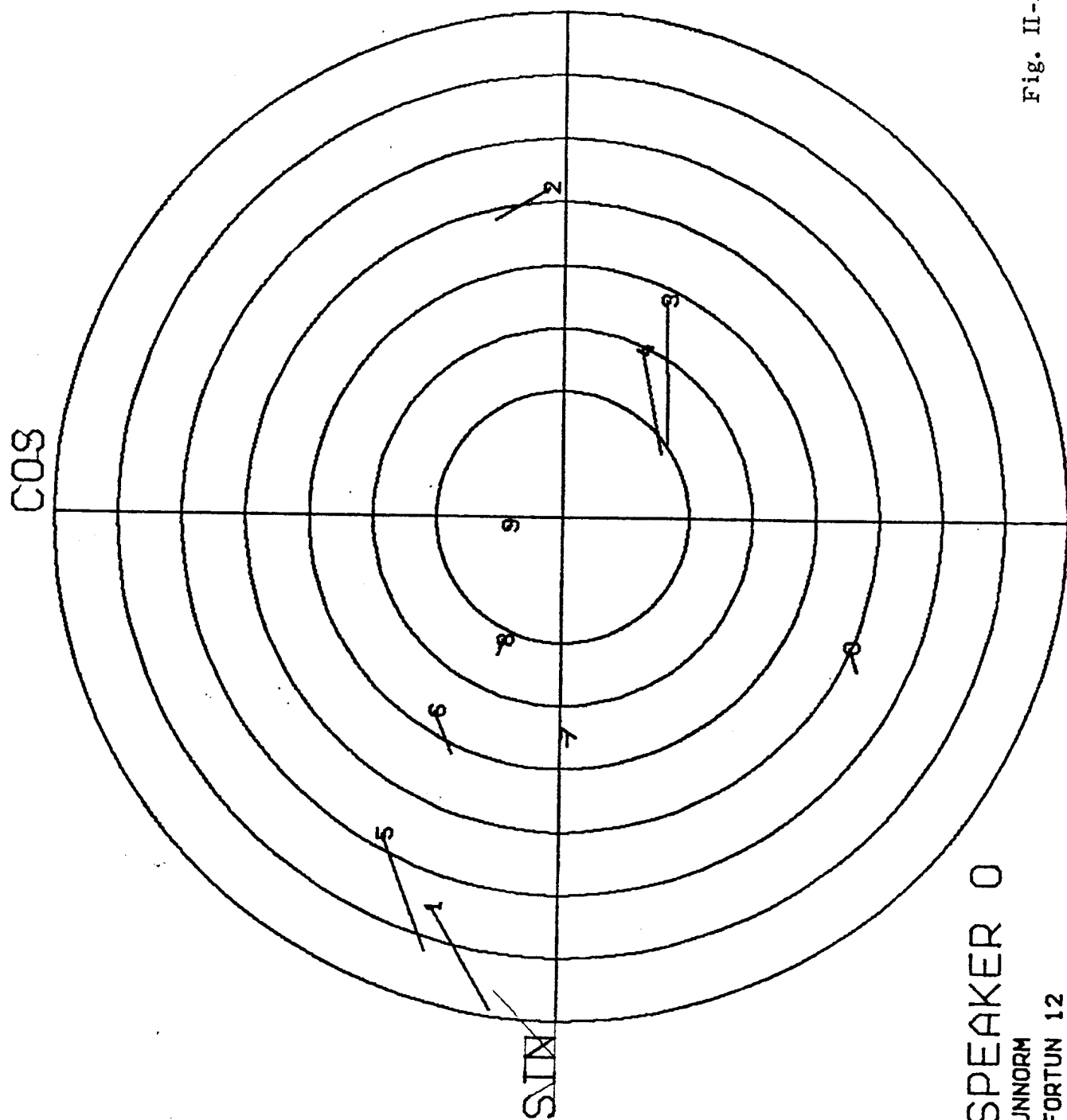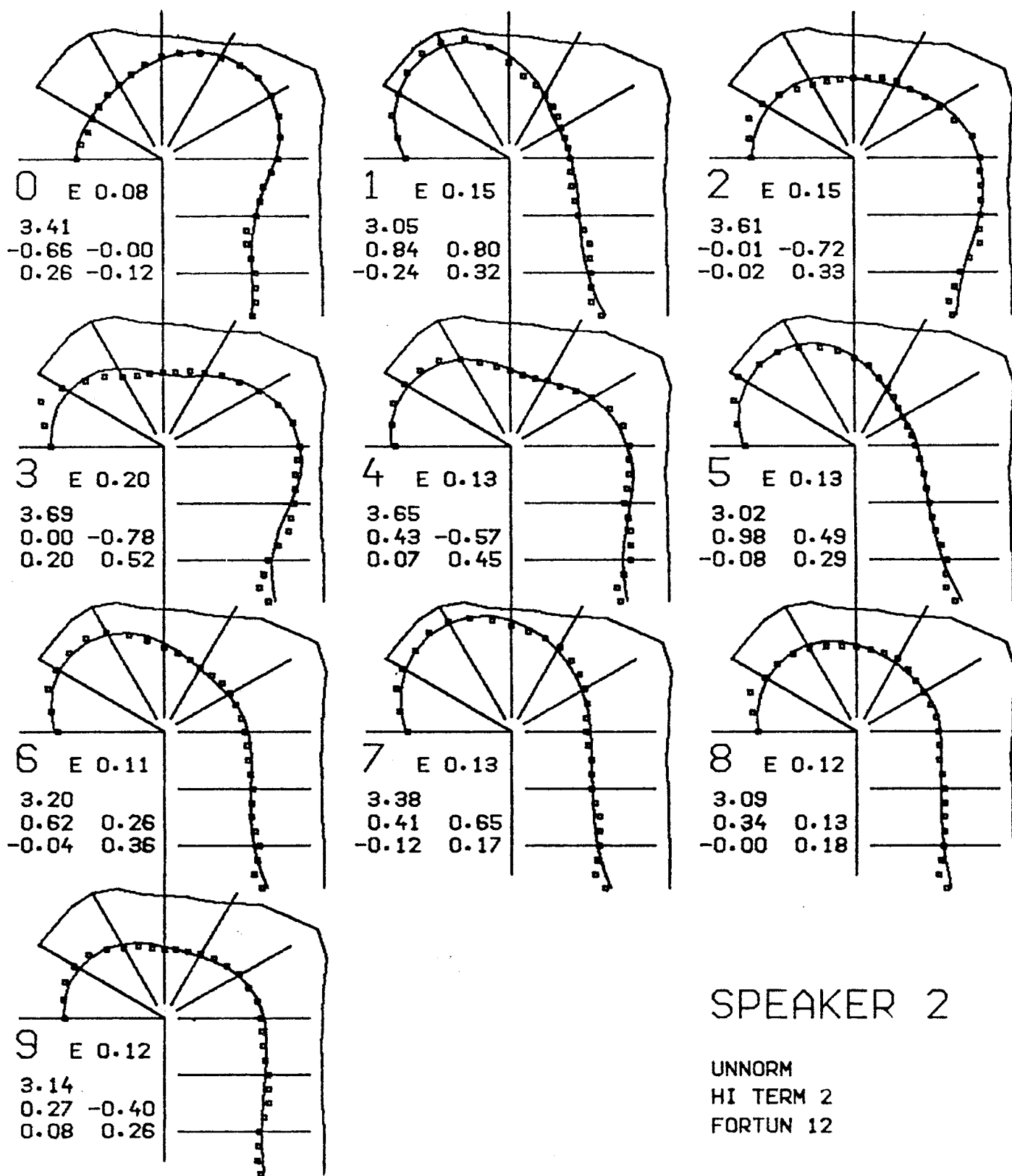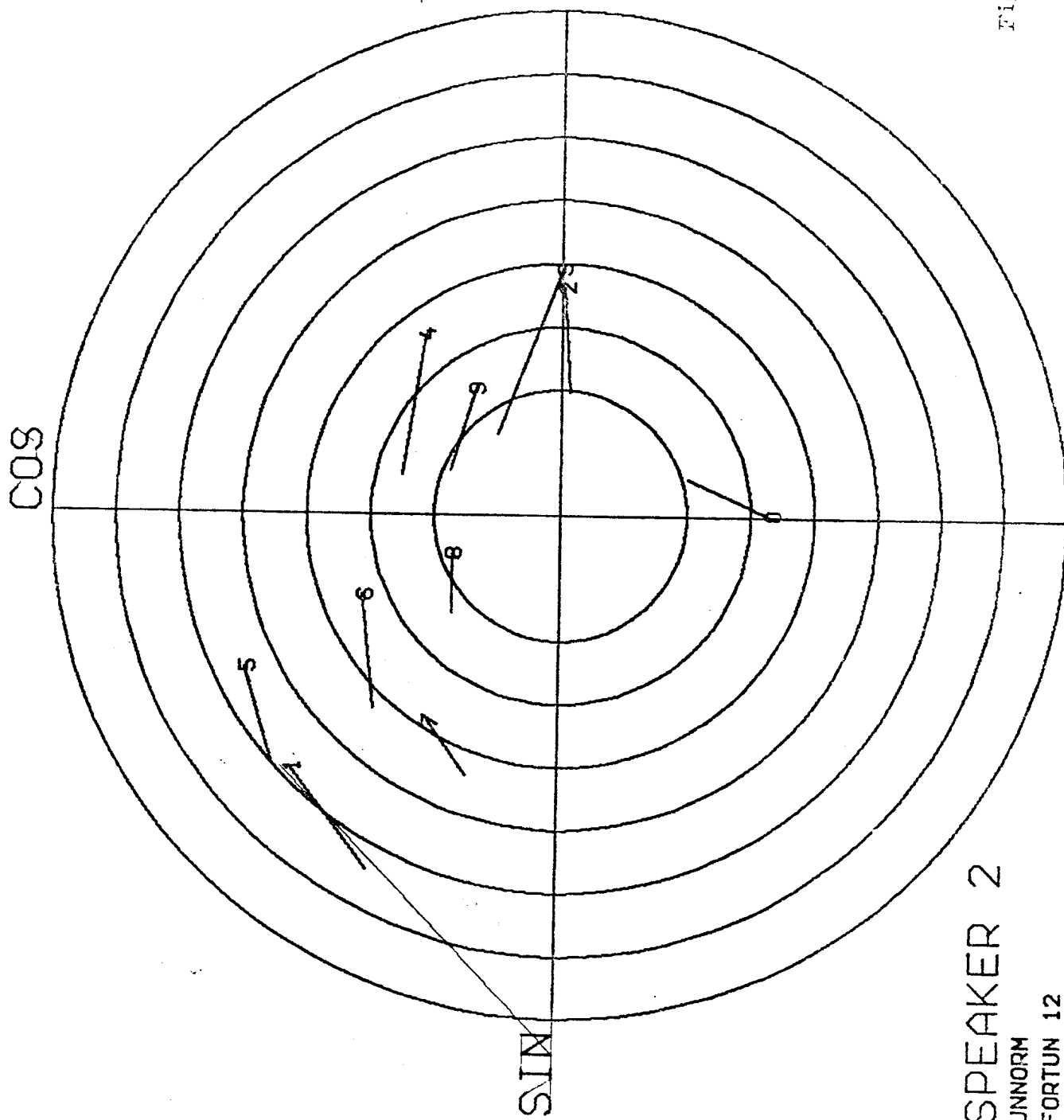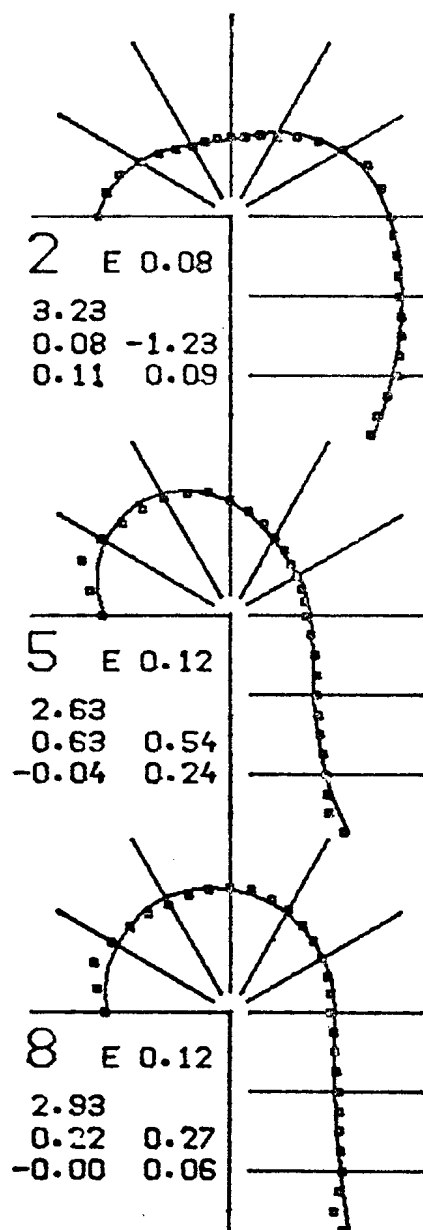from the digits denote the coefficients of the second harmonic.

Fig. II-A-8.

SPEAKER 0
UNNORM
FORTUN 12

0 E 0.08
3.41
-0.66 -0.00
0.26 -0.12

1 E 0.15
3.05
0.84 0.80
-0.24 0.32

2 E 0.15
3.61
-0.01 -0.72
-0.02 0.33

3 E 0.20
3.69
0.00 -0.78
0.20 0.52

4 E 0.13
3.65
0.43 -0.57
0.07 0.45

5 E 0.13
3.02
0.98 0.49
-0.08 0.29

6 E 0.11
3.20
0.62 0.26
-0.04 0.36

7 E 0.13
3.38
0.41 0.65
-0.12 0.17

8 E 0.12
3.09
0.34 0.13
-0.00 0.18

9 E 0.12
3.14
0.27 -0.40
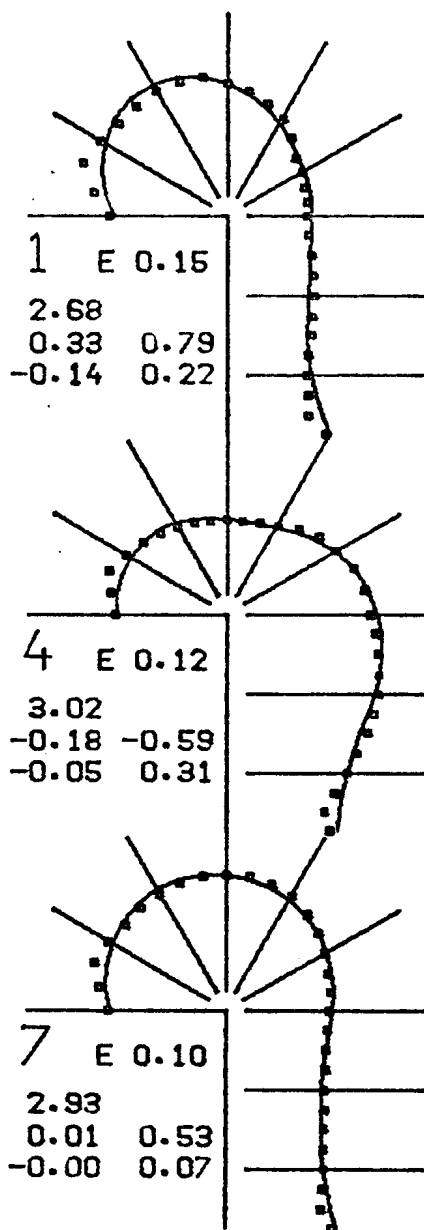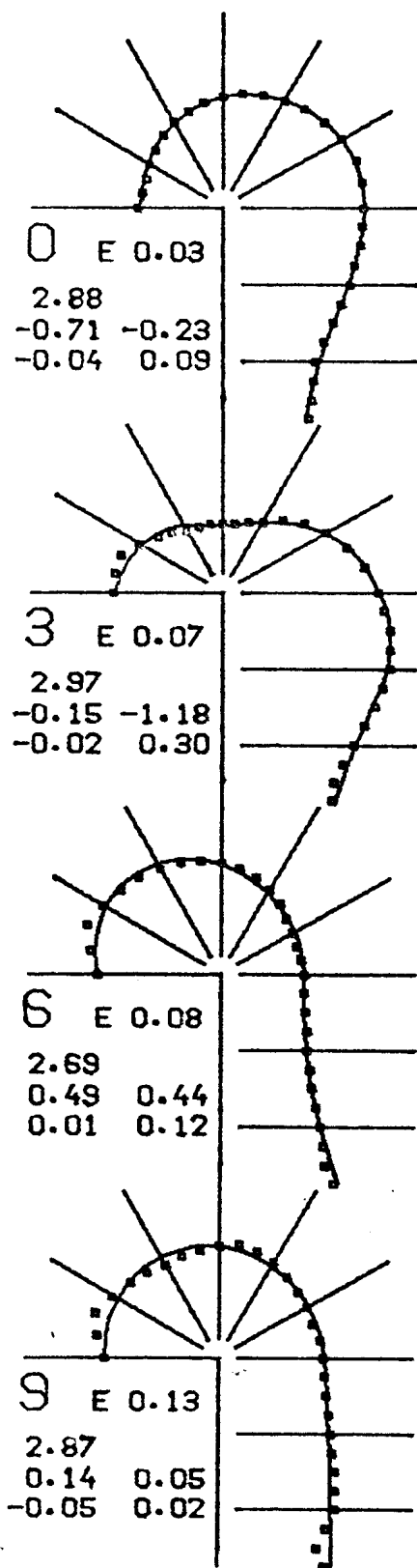0.08 0.26

SPEAKER 2

UNNORM
HI TERM 2
FORTUN 12

Fig. II-A-5.

COS

SIN

SPEAKER 2
UNNORM
FORTUN 12

Fig. II-A-9.

0   E 0.03
2.88
-0.71 -0.23
-0.04   0.09

1   E 0.15
2.68
0.33   0.79
-0.14   0.22

2   E 0.08
3.23
0.08 -1.23
0.11   0.09

3   E 0.07
2.97
-0.15 -1.18
-0.02   0.30

4   E 0.12
3.02
-0.18 -0.59
-0.05   0.31

5   E 0.12
2.63
0.63   0.54
-0.04   0.24

6   E 0.08
2.69
0.49   0.44
0.01   0.12

7   E 0.10
2.93
0.01   0.53
-0.00   0.07

8   E 0.12
2.93
0.22   0.27
-0.00   0.06

9   E 0.13
2.87
0.14   0.05
-0.05   0.02

SPEAKER 5

UNNORM
HI TERM 2
FORTUN 12

Fig. II-A-6.

COS

SIN
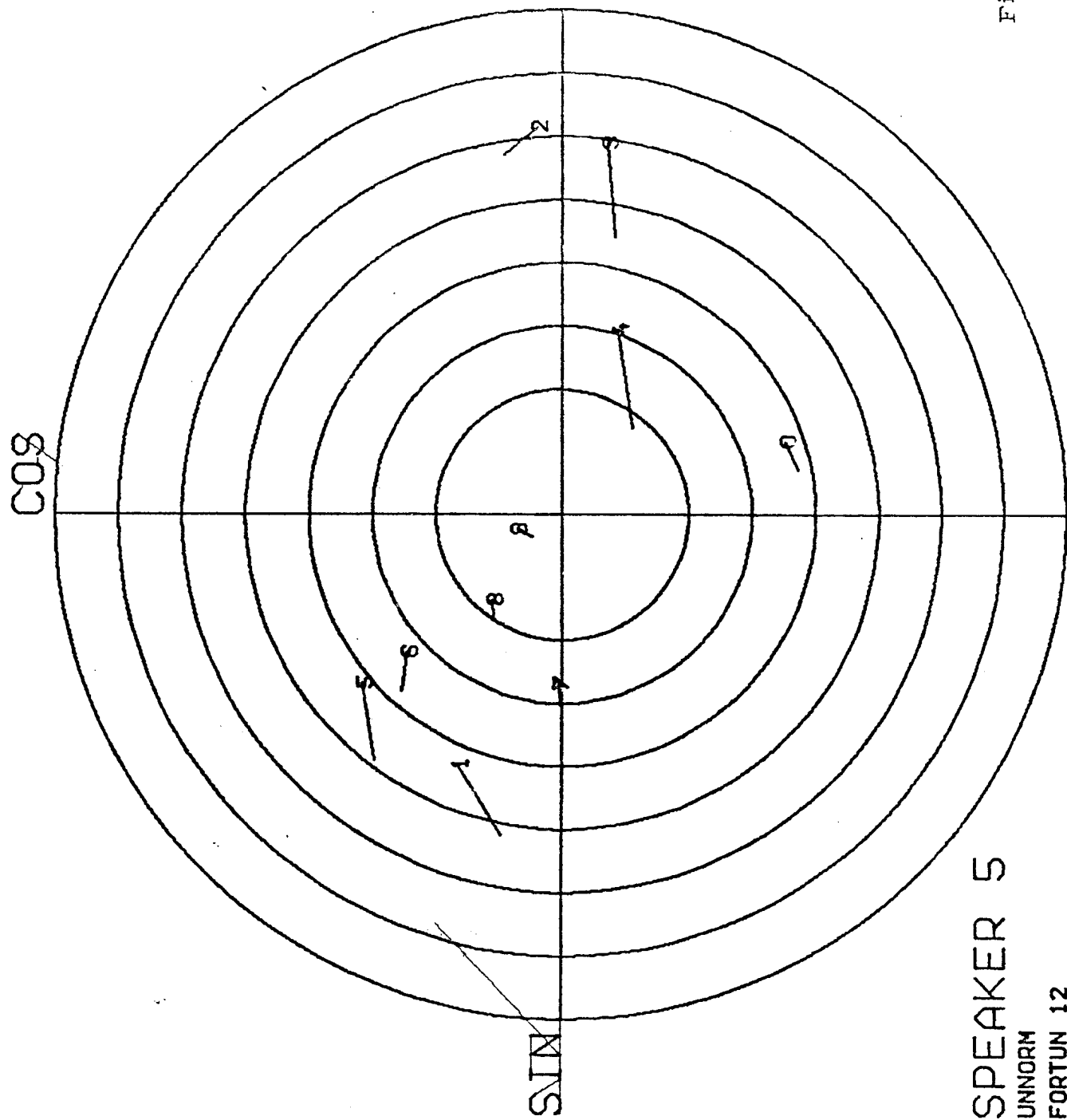
SPEAKER 5
UNNORM
FORTUN 12

Fig. II-A-10.

0   E 0.19
3.07
-0.84 -0.45
0.12 -0.10

1   E 0.13
2.81
0.04 0.65
-0.12 0.12

2   E 0.16
3.34
0.06 -1.70
0.09 0.02

3   E 0.05
3.08
-0.21 -1.43
-0.08 0.29

4   E 0.06
3.08
-0.19 -1.30
-0.10 0.14

5   E 0.10
2.70
0.30 0.46
-0.06 0.15

6   E 0.03
2.85
0.21 0.03
0.05 -0.04

7   E 0.06
3.07
-0.26 0.02
0.09 -0.07

8   E 0.04
2.99
0.02 -0.22
0.11 -0.06

9   E 0.05
2.90
0.08 -0.58
0.07 -0.16
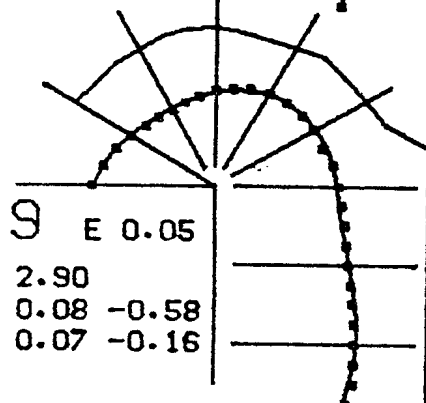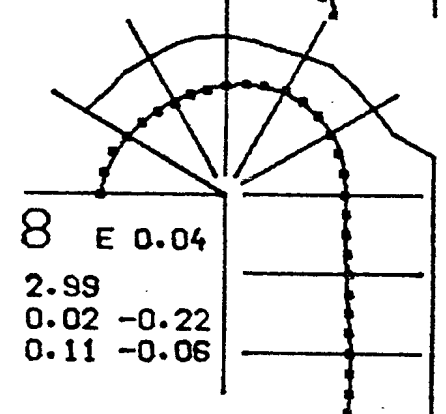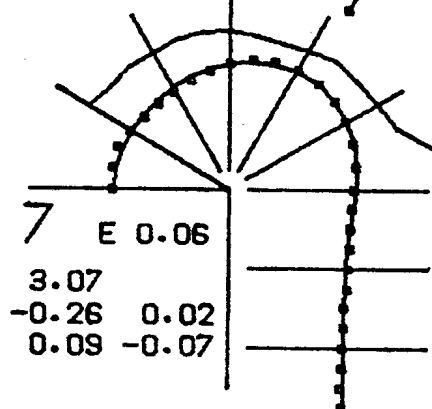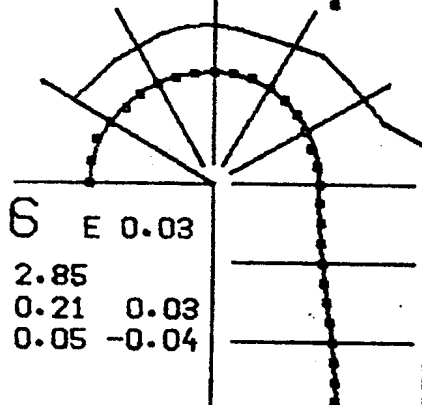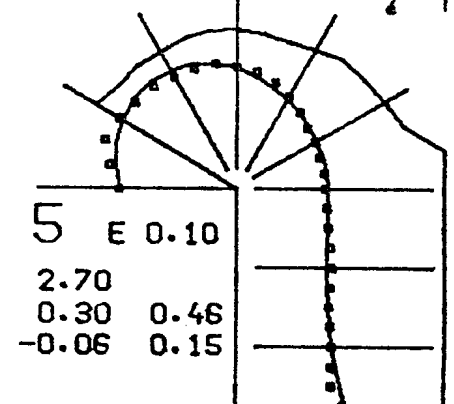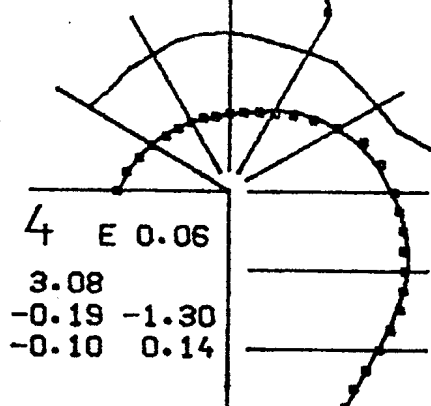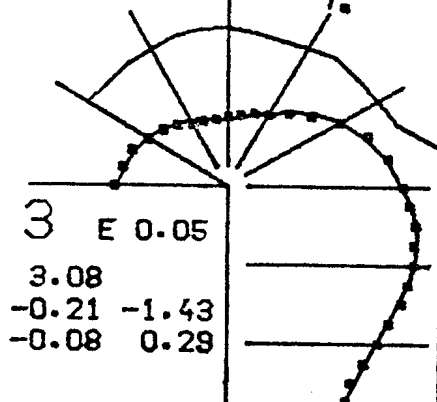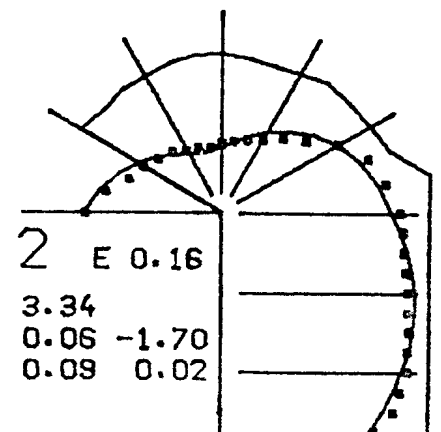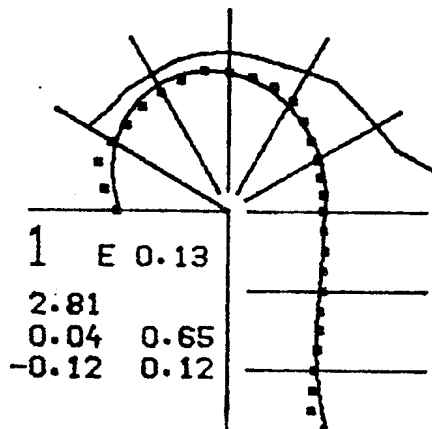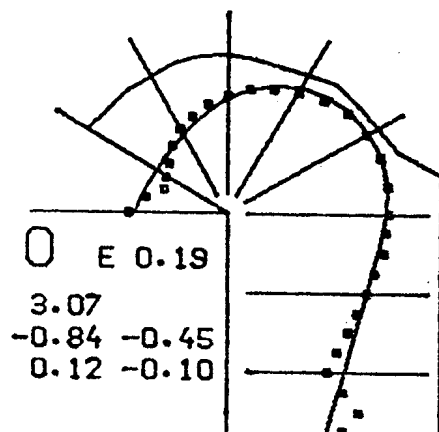
SPEAKER 6

UNNORM
HI TERM 2
FORTUN 12

Fig. II-A-7.

SIN

COS

SPEAKER 6
UNNORM
FORTUN 12

Fig. II-A-11.

0   E 0.00
3.00
0.00 -0.00
0.00 -0.00

1   E 0.00
3.00
0.70   1.22
0.00 -0.00

2   E 0.00
3.00
-0.70   1.22
0.00 -0.00

3   E 0.00
3.00
-1.40 -0.00
0.00 -0.00

4   E 0.00
3.00
-0.70 -1.22
0.00 -0.00

5   E 0.00
3.00
0.70 -1.22
0.00 -0.00

6   E 0.00
3.00
0.00 -0.00
0.00   0.60

7   E 0.00
3.00
0.70   0.70
0.00   0.40

8   E 0.00
3.00
-0.70   0.70
0.00   0.40

9   E 0.00
3.00
0.00 -1.00
0.40   0.40

HI TERM 2
FORTUN 12

Fig. II-A-12.   Plots of various artificial contours, using the two lowest
spatial harmonics, to illustrate the function of the model.
Cf. Fig. II-A-13.

Fig. II-A-13.   Coefficients for the artificial profiles of Fig. II-A-12.

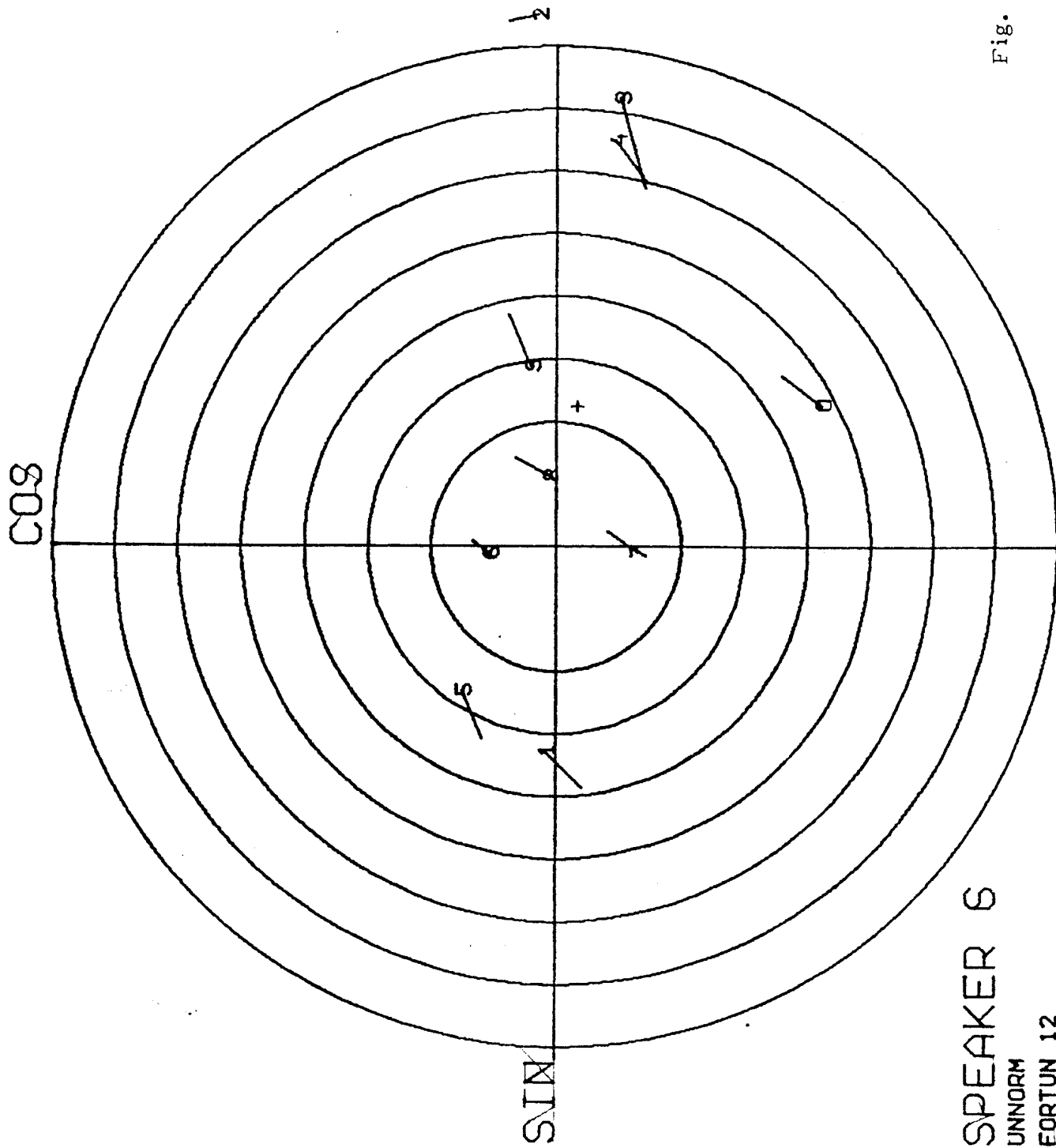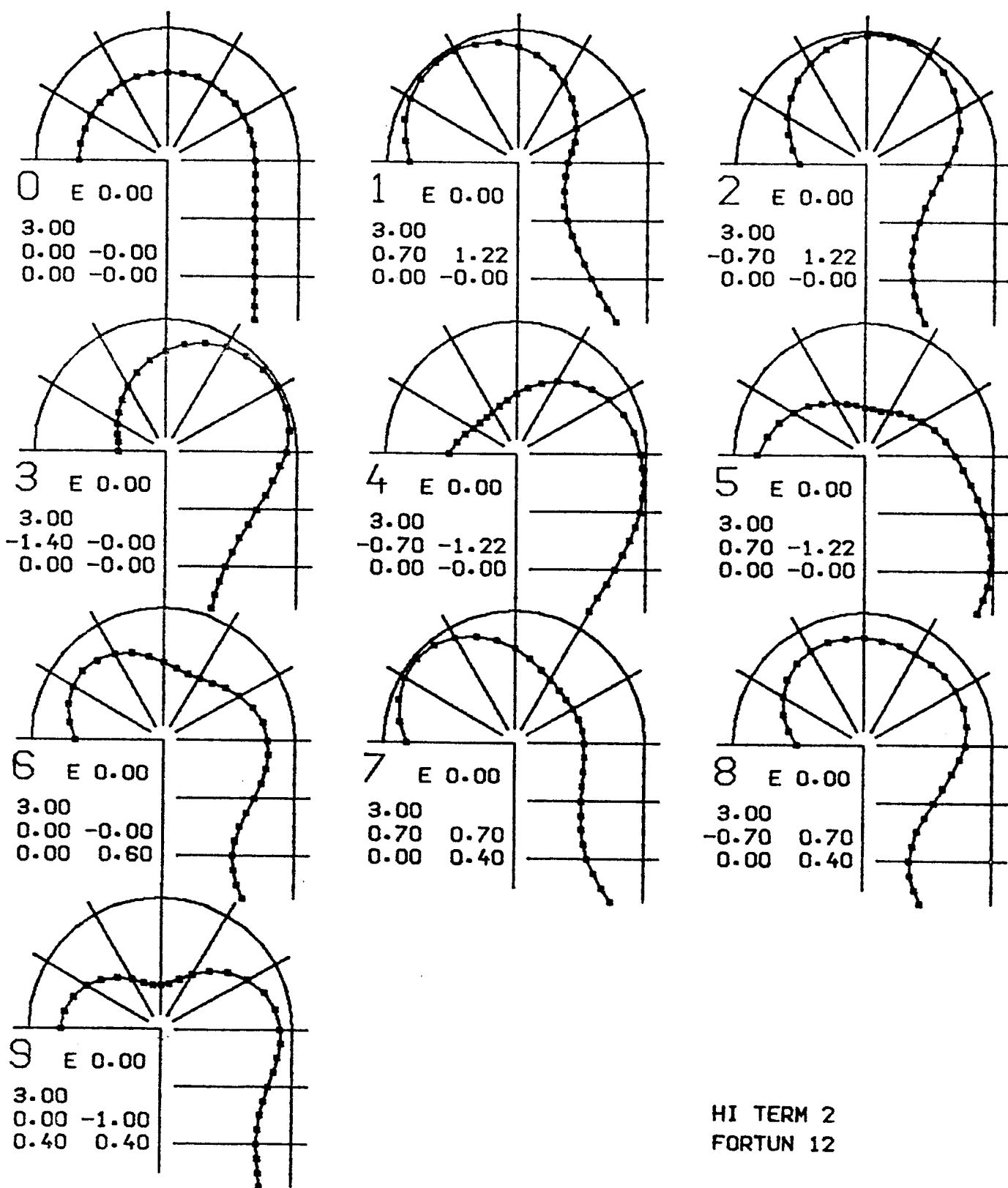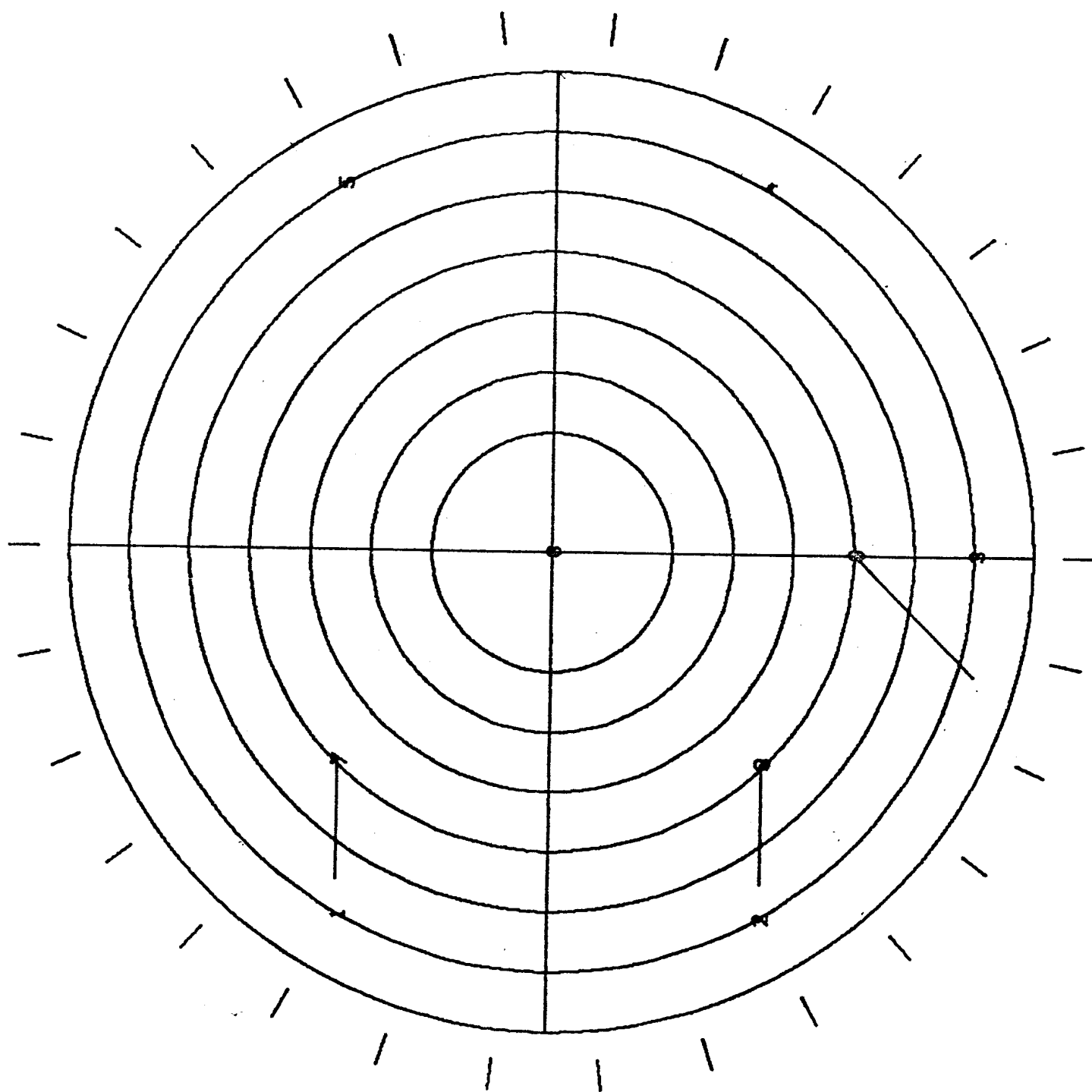This expresses the ratio of the "unwanted signal", (which is equivalent to the RMS error) to the "wanted signal", (namely the data). The formula also includes the DC term which is a deviation from usual practice in defining distortion on signals.

Visualization of analysis data

The representation of the tongue profile by means of a few Fourier coefficients may at first glance seem to be very far fetched and difficult to interpret. We shall now illustrate that this need not be the case.

In Figs II-A-8 and II-A-9 the sine and cosine coefficients of the fundamental is plotted on a Cartesian coordinate system. The values for each of the contours in Figs. II-A-4 and II-A-5 is indicated with numerals corresponding to the vowel number. Also from each of these points there is a vector showing the coefficients for the spatial second harmonic. The representation of the coefficients in terms of magnitude and phase is aided in the coefficient plot by the circles. These then indicate the magnitude coordinate of the spatial fundamental.

Figs. II-A-12 and II-A-13 show the profiles and coefficients for a simplified, artificial case. Item number 0 here has a DC coefficient only and corresponds to a case with "neutral" articulation. In items 1 to 5 a spatial fundamental having a constant magnitude is added, and its phase is systematically varied. In the correspondent plots of contours we see how the point of maximal excursion, or equivalently: point of narrowest constriction, moves from a "dental" position to a "pharyngeal". Thus the phase of the spatial fundamental is a measure of the place of articulation. This is also indicated in Fig. II-A-13 with markers around the circumference of the plot.

In the simple model with DC term and fundamental and having a fixed coordinate value for the contour of palate, velum, and rear pharyngeal wall it is also easy to find the locus of complete lingual closure in the coefficient plot. It is the circle where the magnitude of the fundamental equals the difference between the palatal-pharyngeal coordinate value and the DC coefficient.

When also a second harmonic is introduced as in items 6 to 9 in Fig. II-A-12, we see that the increased spatial resolution may be used in modeling for instance the tongue tip. It is then however more difficult to find a locus of complete closure, but it can still be done in an elementary way by vector addition in the coefficient plot, and keeping in mind to scale the phase of the higher harmonic vectors properly.

The function of the Fourier model of the tongue may also be further illustrated with a mechanical device as outlined in Fig. II-A-14. A circular disk has a central pivot. When the disk is tilted the height of its circumference will be approximately a sinusoidal function of the angle around the disk. The vertical movements of equidistant points on the disk are transferred with thin wires to pointers arranged in a coordinate system. The set of pointers will indicate the model contour. The cosine and sine coefficients of the spatial fundamental are given by the back-forth and the sideways tilt of the disk respectively. The DC term may be controlled by elevating the pivot of the disk.

Positioning of the coordinate system

At the outset of this experiment it was thought natural to use a coordinate system having the mandible as a reference, since a large part of the lingual muscles are joined with the mandible. Some data for this placement of the coordinate system is shown in Figs. II-A-6 and II-A-10.

A comparison with the data of Figs. II-A-4 and II-A-8, which pertain to the same subject, shows that the deviation of the model from the ideal is not improved, the mean error is essentially the same. So the accuracy aspect does not seem to be an argument for a coordinate system following the mandible. One could however still hope that the additional information necessary to specify the mandible position would give a dividend in form of a smaller variation in the coefficients describing the tongue shape. Especially this ought to be the case with the spatial DC term if the bulk of the tongue was to follow the jaw. A comparison of the data shows that the standard deviation of the DC term is indeed smaller in the mandibular coordinate system, but not very much. In the fixed coordinate system we have $\sigma_{Ro} = 0.219$ and in the moving $\sigma_{Ro} = 0.172$.

The conclusion will be that for the purpose of describing the profile shape there is no gain in taking the mandibular position into account, because the increased number of parameters will not give a better accuracy or a more easy interpretation. Also it will be more circumstantial with a moving tongue coordinate system to determine the distance between the tongue profile and the stationary structures of the speech organs.

A different problem is to determine how critical the positioning of the coordinate system is with relation to the stationary structure. Figs. II-A-7 and II-A-11 show data on speaker 6. These pertain to the same tracings as
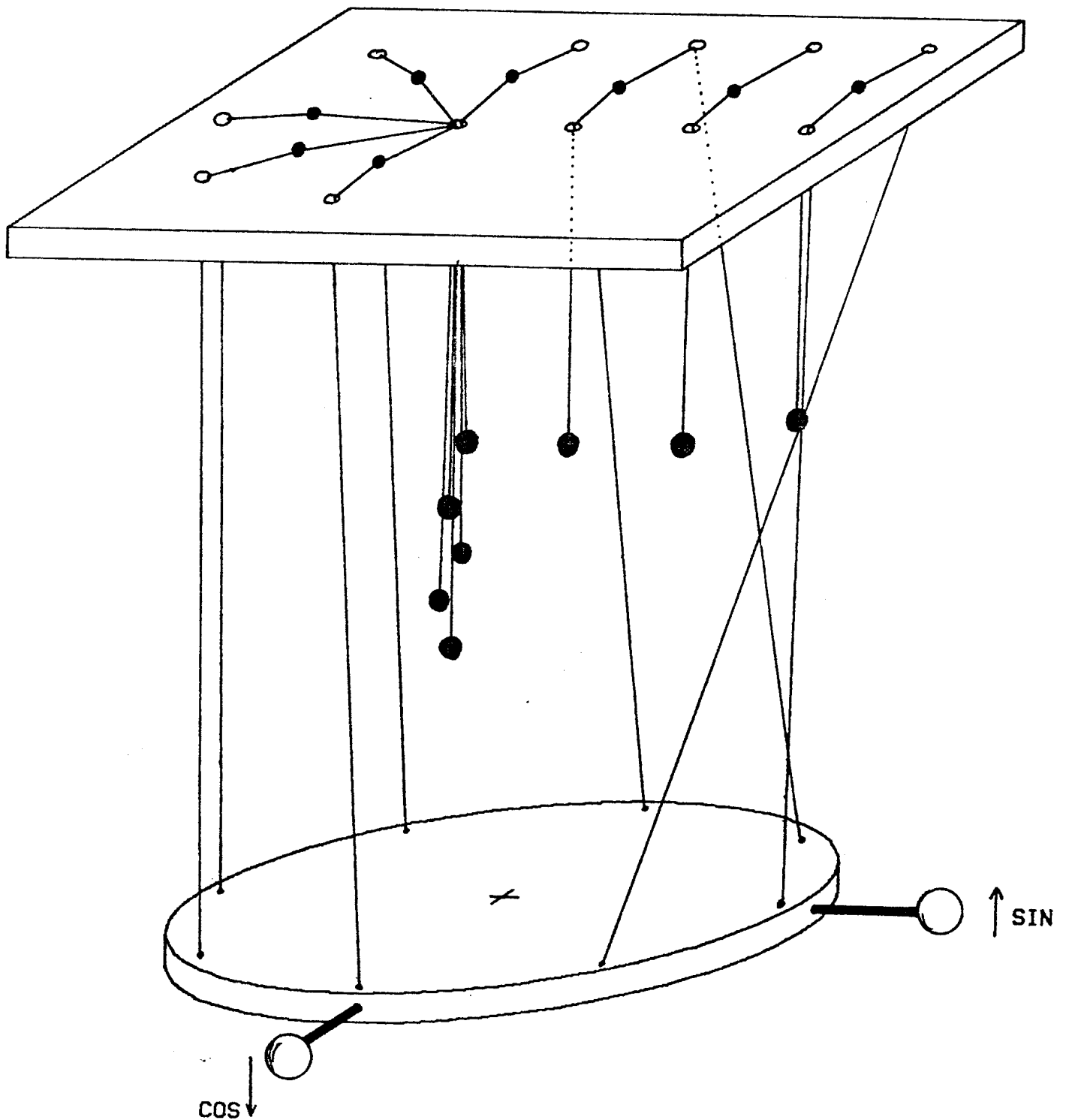
Fig. II-A-14.   Mechanical device to illustrate the Fourier model.
Sine and cosine coefficients of the spatial fundamental
control the tilt of circular disc with a central pivot.
Vertical movements of its periphery are transferred
with thin wires to coordinate system at top. Wires
are kept taut by small weights.

for speaker O but the coordinate system has been translated .5 cm forwards and .5 cm upwards. Inspecting the polar coefficient plots in Figs. II-A-8 and II-A-11 shows that the distribution of the coefficients is very closely similar. We see that the set of data points has been translated and somewhat rotated, but the interrelations are essentially unchanged.

Normalizing

When the tongue contour comes close to the origin of the polar system, then the sampling density, defined as samples per unit of contour length, becomes rather high at this portion. To compensate for that over-representation of a small part of the contour a length normalization was performed in part of the computations. To do this an auxiliary table was computed which contained the distances between successive samples. This table was then converted by summation into another table with cumulative distance along the contour. Finally, normalized input data was derived by interpolation from the original input in 30 points equidistant along the contour. Some data normalized this way are shown in Fig. II-A-2, where the original data points are marked with dots.

When the errors are compared in the two cases with and without normalizing with respect to contour length, then we find that the difference in general is so small that it is of no consequence. An important proviso is however that the coordinate system has been put reasonably central with respect to the population of contours so that these will not come too close to the origin of the polar system.

As a consequence the extra labour of normalizing the data was abandoned for a larger part of the investigation.

Correlation analysis of the Fourier coefficients

For a further illustration a number of correlation coefficients have been computed. In Table II-A-III (p. 11), different speakers are pairwise compared. For each of the Fourier coefficients the correlation coefficient has been computed, using the sine/cosine coefficients. For the Fourier coefficient X with ordinal number r, speakers k and l, and using index n for the N vowels, the correlation coefficient was computed as

$$\rho_{r,k,l} = \frac{\mu_{r,k,l}}{\sigma_{r,k} * \sigma_{r,l}}$$

with the covariance

$$\mu_{r,k,l} = \frac{1}{N} \sum_n (X_{r,k,n} - m_{r,k}) * (X_{r,l,n} - m_{r,l})$$

variance

$$\sigma^2_{r,k} = \frac{1}{N} \sum_n (X_{r,k,n} - m_{r,k})^2$$

and mean

$$m_{r,k} = \frac{1}{N} \sum_n X_{r,k,n}$$

A high correlation for different speakers using the sine/cosine coefficients will indicate that their corresponding coefficient plots (as in Figs. II-A-8 to II-A-11) are similar. The differences eliminated by the correlation analysis are those of a parallel translation and of magnitude scaling. A shortcoming is however that differences caused by rotation of the plots are not compensated for. Rotational differences do exist as can be seen from the plots, and if these had been compensated for the correlations of table II-A-III would probably have been still higher.

The purpose of these computations is to show that:

a. comparing different subjects (speakers 0 and 2), the DC term and the spatial fundamental are strongly correlated. This implies that the gross shapes are similar for the two sublects uttering different vowels. The second harmonic is not as strongly correlated which should be due to a more subject dependent fine structure.

b. comparing different coordinate systems, cranium based and jaw based, (speakers 0 and 5), all coefficients are strongly correlated. From this it is concluded that it is irrelevant which coordinate system is selected, so far as the relative distribution of the coefficients for the vowels is concerned. In other words, the points in Figs. II-A-8 and II-A-10 are arranged in essentially the same pattern.

c. considering a displacement of a cranium based coordinate system (speakers 0 and 6) the same conclusions as in b. seem to hold.

These results are not very revolutionary in themselves, but they confirm the usefulness of the Fourier coefficients as descriptors.

Another correlation study is shown in Table II-A-IV. For each speaker, the correlation between the various Fourier coefficients has been computed. The same formulae as above were used, but using r as subscript for the speaker and k and l as subscripts for two Fourier coefficients. Furthermore these computations were made on both the sine/cosine and the magnitude/phase representations. Then a normalized phase measure was used, varying from 0 to 1:

$$f_r = (\varphi_r + \pi)/2\pi$$

In the table, for simplicity, the nonexistent DC sine and phase components are represented by zeros.

Table II-A-IV.

	sin/cos CC_0		CC_1	CS_1	CC_2	CS_2
Speaker 0	1.000					
	0.000	0.000				
	-0.393	0.000	1.000			
	-0.873	0.000	0.423	1.000		
	0.876	0.000	-0.316	-0.754	1.000	
	-0.166	0.000	0.084	-0.051	-0.487	1.000
Speaker 2	1.000					
	0.000	0.000				
	-0.580	0.000	1.000			
	-0.740	0.000	0.541	1.000		
	0.548	0.000	-0.813	-0.689	1.000	
	0.286	0.000	0.483	-0.363	-0.206	1.000
Speaker 5	1.000					
	0.000	0.000				
	-0.518	0.000	1.000			
	-0.809	0.000	0.518	1.000		
	0.650	0.000	-0.016	-0.517	1.000	
	-0.157	0.000	0.049	-0.226	-0.377	1.000
Speaker 6	1.000					
	0.000	0.000				
	-0.445	0.000	1.000			
	-0.824	0.000	0.245	1.000		
	0.383	0.000	-0.273	-0.086	1.000	
	-0.057	0.000	0.151	-0.198	-0.836	1.000

	Magn/phase R_0		R_1	f_1	R_2	f_2
Speaker 0	1.000					
	0.000	0.000				
	-0.165	0.000	1.000			
	0.779	0.000	0.219	1.000		
	-0.093	0.000	0.555	0.421	1.000	
	-0.622	0.000	-0.390	-0.635	-0.174	1.000
Speaker 2	1.000					
	0.000	0.000				
	-0.153	0.000	1.000			
	0.709	0.000	-0.280	1.000		
	0.546	0.000	0.386	0.445	1.000	
	0.009	0.000	0.031	-0.096	-0.236	1.000
Speaker 5	1.000					
	0.000	0.000				
	0.261	0.000	1.000			
	0.744	0.000	0.668	1.000		
	-0.108	0.000	0.564	0.437	1.000	
	-0.514	0.000	-0.621	-0.490	-0.068	1.000
Speaker 6	1.000					
	0.000	0.000				
	0.668	0.000	1.000			
	0.669	0.000	0.512	1.000		
	-0.056	0.000	0.409	0.334	1.000	
	-0.253	0.000	-0.769	-0.103	-0.314	1.000

The more interesting part of the table is the one with the correlations between the magnitude/phase coefficients, because of the spatial fundamental can be interpreted as measures of effort and place of articulation.

Examination of the table may lead to the following conclusions:

a. In all cases there is a strong correlation between the DC term and the phase of the spatial fundamental. This indicates that these two terms are not orthogonal when regarded as articulatory parameters.

b. The fact that this holds also for speaker 5 (jaw based coordinate system) tends to counter indicate the usefulness of the mandibular position as an articulatory parameter.

Conclusions

The mid-sagittal profile of the tongue can be efficiently described in terms of a set of Fourier coefficients. For modelling purposes it should be sufficient with only two numbers to describe the profile, the magnitude, and the phase of the spatial fundamental. Possibly the additional inclusion of the DC term may be useful, but this is found to be correlated to the fundamental phase.

For increased accuracy in modelling of live subjects an arbitrary number of higher terms may be included, without changes in the values of the basic, low spatial frequency harmonics. There seem to be no tangible gains in using the mandibular position as a reference in the description. On the contrary, a moving coordinate system for the tongue profile will make estimations of vocal tract diameter and area function a good deal more complicated. It is thus recommended that the mandibular position is not used as a primary articulatory parameter, but rather as a secondary parameter, dependent on the others.

Reference

(1)   Lindblom, B. and Sundberg, J.: "Acoustical consequences of lip, tongue, jaw and larynx movement", J.Acoust.Soc.Am. 50, No. 2 (1971), pp. 1166-1179.