DE GRUYTER

FUNDAMENTALS OF PHONETICS, II: ACOUSTICAL MODELS, GENERATING THE FORMANTS OF THE VOWEL PHONEMES

JANUA LINGUARUM. SERIES MINOR

FUNDAMENTALS OF PHONETICS, II

JANUA LINGUARUM

STUDIA MEMORIAE NICOLAI VAN WIJK DEDICATA

edenda curat

C. H. VAN SCHOONEVELD
INDIANA UNIVERSITY

SERIES MINOR
NR. XXVI/2

1970
MOUTON
THE HAGUE · PARIS

FUNDAMENTALS OF PHONETICS

II: Acoustical Models Generating the Formants of the Vowel Phonemes

by

H. MOL
UNIVERSITY OF AMSTERDAM

1970
MOUTON
THE HAGUE · PARIS

LIBRARY OF CONGRESS CATALOG CARD NUMBER: 70-110954

Printed in The Netherlands by Mouton & Co., Printers, The Hague.

For Nancy and Laurence Batchelder

ACKNOWLEDGEMENTS

The author has to thank Mr. J. G. Blom, staff-member of the Institute of Phonetic Sciences of the University of Amsterdam, for carefully and, above all, critically reading the typescript and for making many valuable suggestions. Moreover, Mr. Blom is entitled to a special word of thanks for making the computer-programs presented in the Appendix.

The author wishes to acknowledge his indebtedness to Miss N. Schokking, who prevented him from committing at least the most flagrant sins against the English language.

The author wishes to express his thanks to his students, who urged him to develop methods for satisfying their scientific curiosity.

CONTENTS

I

INTRODUCTION

In her review of Gerold Ungeheuer's *Elemente einer akustischen Theorie der Vokalartikulation* (Springer-Verlag 1962), Ilse Lehiste[1] states: "The question might be asked whether, after the appearance of Fant's comprehensive volume[2], it is at all possible to make a new and original contribution to the acoustic theory of vowel production". Frankly speaking, I do not like such discouraging statements and I am sure dr Fant does not like them either. Being an engineer myself, I greatly admire Fant's work and certainly appreciate its originality. Nevertheless, my approach to the problem of the prediction of the formant frequencies of the vowels is different from that of Fant.

This monograph has been written with a purely linguistic purpose. In the Netherlands there exists a new generation of linguistic students and also a group of mature linguists who are no longer content with the traditional verbal labels of the phonemes. It is the duty of the physically and mathematically trained phonetician to provide these malcontents with a manageable concept of the relation between articulation and acoustics. More specifically for the vowels, there is a need for a workable model of the generation of the formant positions.

Predictions of the formant positions based on a simple model need not necessarily bear the hall-mark of inaccuracy. By suitably choosing the values of its few parameters with an eye to the real shape of the vocal tract as seen in X-ray photographs, one can form

[1] *Language*, Volume 39, Number 3 (1963).
[2] G. Fant, *Acoustic theory of speech production* (Mouton, The Hague 1961).

fair estimations that compare favourably with the results of calculations of more pretentious models of the vocal tract.
The model around which this monograph centres is called the loss-free twin-tube resonator. A formula producing the formants of the twin-tube resonator can be found in literature but we shall show in the following chapters that until now the possibilities of its graphical solution have not been sufficiently explored. Maybe the reason for this neglect is that the twin-tube was seen as a bad imitation of the vocal tract rather than as a good model of it. It can even be shown, as in Chapter VII, that Hellwag's vowel triangle, published as early as 1781, comes to the fore as a vector diagram of the first two formants as predicted by the twin-tube model.

An acoustical model is, as a rule, representative of SOME properties only of a real vocal tract. Therefore it is always an approximation of the physical reality in the real vocal tract. Nevertheless, the calculation of the model must be taken seriously and its results must be checked as much as possible by measurements on 'hard ware' models, conveniently made of hard plastics.

For not too high frequencies, a hard-walled tube with uniform cross-area may be treated as a one-dimensional problem, that is, plane waves may be supposed to be running to and fro in the tube following the direction of its axis.

We found it necessary, however, to introduce in our models sections with non-uniform cross-area, such as diabolos, defined in this monograph as exponential twin-horns. One needs a manageable differential equation for describing the physical behaviour of such sections. Some authors, for instance Ungeheuer[3], advocate the application of Webster's horn equation even to the complete vocal tract. We only succeeded in solving Webster's horn equation in special cases like, for instance, the exponential horn. As this was the horn we needed we were able to introduce it in a variant of the twin-tube model.

To an outsider possessing a smattering of phonetics a formant

[3] G. Ungeheuer, *Elemente einer akustischen Theorie der Vokalartikulation* (Springer-Verlag 1962).

may seem to be a comparatively problemless tangible thing, lying in waiting to be measured by a suitable method. Even in the phonetic sciences one runs the risk of being accused of riding a hobby-horse when one underlines the arbitrary nature of a formant definition. In spite of the (fortunate) fact that fundamentally different formant definitions often lead to frequencies in the same order of magnitude, in spite of the fact that contrasts between formant positions prove to be more important than absolute formant positions, we should not consider formant theory as a free-for-all for loose or arbitrary formulations. Definitions labelled as traditional or practical do not improve the clear comprehension of the mechanism of speech and hearing. Our ultimate scientific aim should be that clear comprehension. Chapter III should be considered as a first, faltering step towards this goal.

This monograph is hopefully announced as a contribution to linguistics. However, its author is vividly aware of the fact that the average linguist will not be able to digest all physical and mathematical problems without the assistance of an acoustically trained physicist or engineer. Nevertheless, we do hope that the interested linguist will be able to appreciate the conclusions drawn from the calculations. This monograph is also meant for the acoustician who wishes to team up with linguists in phonetic research. He may be irritated somewhat by the elaborate treatment of what he considers as details or general knowledge. On the other hand, it is quite healthy for engineers and kindred spirits to come down to earth every now and then in order to realize how and if their mathematical concepts fit into the real mechanism of speech and hearing. To-day, phonetics is no longer a one-man science. Insight into the mechanism of speech and hearing can only be gained via a close co-operation between linguists, engineers, physicists, physiologists, psychologists, speech therapists, anthropologists and so on. Let us suppose, for a moment, that one wants to concentrate on articulatory problems only, waiving all other aspects of phonetics for whatever reason. As long as the articulatory terms and descriptions have the character of real and effective articulation recipes they are very practicable: in the ideal case they permit us to realize in the long run the speech sounds of a language so that a native talker of that language is able to identify them correctly. Articulation recipes, however, only show HOW to produce a series of contrasting speech sounds, but do not explain WHY the listener is able to discriminate between them, though we know THAT he is able to do so. The aim of the phonetic sciences as a whole is to get an over-all picture, including production, transmission and sensory reception of the speech sounds. One is, of course, completely free to restrict oneself to empirically derived articulation recipes but in that case one shuts out a broad and interesting field, namely, acoustics and sensory reception. A very striking example of the one-sidedness of articulation recipes is met in audiology: the audiologist needs acoustic labels of the speech sounds and, moreover, he wants them in such a form that he can understand WHY a defective organ of hearing fails to discriminate between them.

II

PHYSIOLOGICAL CONSIDERATIONS

In this short monograph a detailed treatment of the anatomy of the vocal tract would certainly be out of place. For greater description the reader is referred to the well-known text-books. This admittedly brief chapter is merely intended to persuade the reader to see the vocal tract as a slim tube extending from the vocal cords to the mouth opening. Though this view has a typically modern ring it dates back much further than is generally recognized. As early as in 1765 De Brosses describes the vocal tract as a tube that can be varied in diameter and that can be lengthened or shortened, see figure II.2.

The modern motive for preferring to see the vocal tract as a slender tube stems from acoustic theory. If the total length of a tube is much larger than its diameter, its acoustic behaviour may be streamlined to such a degree that the calculation boils down to a one-dimensional problem (see Chapter III).

Figure II.3 demonstrates the oblong shape of the vocal tract. The wall of the tube is for the larger part formed by the wall of the pharynx, the epiglottis, the tongue, the soft- and the hard palate, the teethridges (including the teeth) and the lips. The axial length l of the vocal tract is indicated by the dot-stripe line between the arrow-heads.

As a rule the vocal tract is supposed to end at a point just outside the lips. In certain vowels, however, the vocal tract must be assumed to end at a point inside the mouth, even behind the teeth.
During the production of the vowels the vocal folds (cords) are generally regarded as a hard wall bounding the vocal tract at the

other end, in spite of the fact that they open and close rhythmically at an average rate of at least 150 times per second. As slow motion moving pictures of the larynx show, the just mentioned wall moves to and fro in the axial direction of the vocal tract, but we shall ignore the resultant length modulation.

During phonation the larynx may be regarded as a machine-gun projectilling air puffs into the vocal tract. In normal voices the vocal folds open slowly whereas they close with a snap. It is this closing snap, this abrupt ending of the air puff that is able to excite the vocal tract to action in the form of powerful damped oscillations. As a rule the reaction to the slow onset of the puff is of no practical importance.

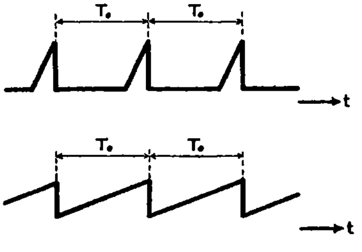


Figure II.1

Two examples of wave forms that are able to excite the vocal tract in a proper way.

The fact that the closing snap is the main source of excitation permits us to apply Heaviside's method to the vocal tract, see Chapter III.

The time interval between two subsequent snaps may be called the REPETITION period T_o. Formally speaking, there is no objection against calling

$$F_o = \frac{1}{T_o} \tag{II.1}$$

the repetition frequency of the snaps.

For periodic wave forms as shown in fig. II.1, T_o and F_o are constants. For other wave forms T_o and F_o may vary from snap to snap. Confining ourselves to periodic wave forms, we mention poitrine peut fournir l'air. Les conſonnes ſont les articulations de ce même ſon que l'on fait paſſer par un certain organe, comme à travers d'une filiere, ce qui lui donne une forme. Cette forme ſe donne en un ſeul inſtant & ne peut être permanente. Que ſi elle paroît l'être dans quelques articulations fortes qu'on appelle *eſprits rudes*, ce n'eſt plus un ſon clair & diſtinct ; ce n'eſt qu'un ſiflement ſourd qu'on eſt obligé d'appeller du nom contradictoire de *voyelle muette*. Ainſi la *voix* & la *conſonne* ſont comme la matiere & la forme, la ſubſtance & le mode. L'inſtrument général de la voix doit être conſidéré comme un tuyau long qui s'étend depuis le fond de la gorge juſqu'au bord extérieur des levres. Ce tuyau eſt ſucceptible d'être reſſerré ſelon un diametre plus grand ou moindre, d'être étendu ou racourci ſelon une longueur plus grande ou moindre. Ainſi le ſimple ſon qui en ſort repréſente à l'oreille l'état où on a tenu le tuyau en y

Figure II.2

Photostatic reproduction of page 109 of: De Brosses, *Traité de la formation méchanique des langues et des principes physiques de l'étymologie*, Tom. I, nr 101 (Paris 1765).

that more often than not T_o is called the fundamental period and F_o the fundamental frequency. The term: fundamental is a concession to Fourier analysis and refers to the possibility of developing the glottal wave into a Fourier series containing the fundamental frequency F_o and its harmonics $2F_o$, $3F_o$, $4F_o$ *etc.* As we do not make use of this mathematical possibility in this monograph we only mention it for the sake of completeness.

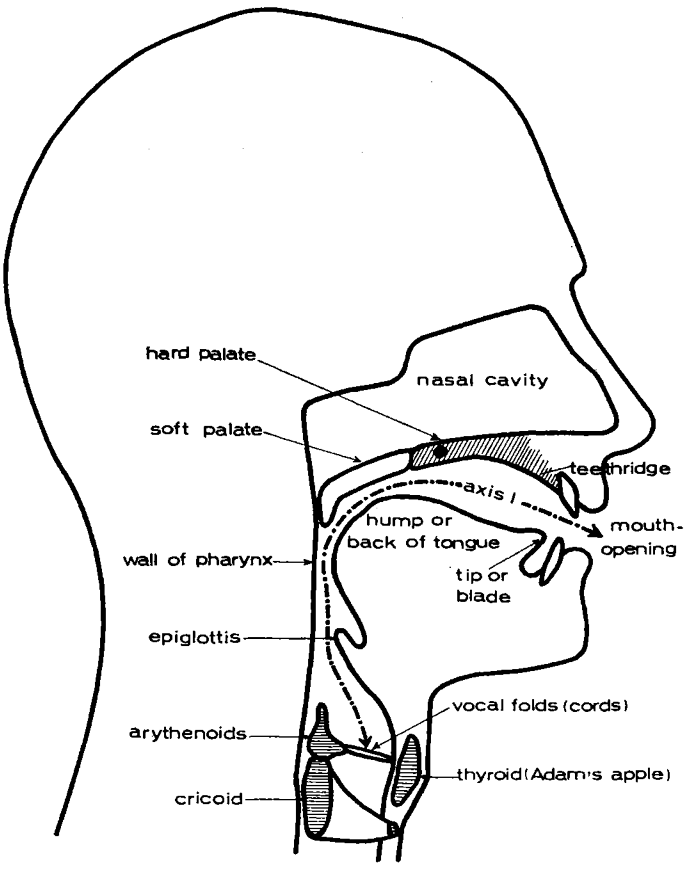


Figure II.3
Semi-schematic representation of the anatomy of the vocal tract.

Already one snap is sufficient to pursuade the vocal tract to produce its complete set of damped oscillations, in other words its formants, see Chapter III. Speaking in terms of information theory: all subsequent snaps are redundant as long as the vocal tract is kept in the same articulatory position. This statement is not weakened by the fact that, in order to arrive at a decision, the nervous system needs a time interval of several repetition periods. It is not necessary, and, in my opinion, it is even incorrect, to explain this time-delay as an inherent feature of a linear filter with a limited band width.

The acoustic effect of lowering the soft palate during the production of a vowel is called nasalisation. In some languages nasalisation is a speech defect or just a mannerism. In other languages it is a phonemic tool.

In this monograph we shall very briefly discuss nasalisation in Appendix § 1.

III

ON THE DEFINITION OF THE FORMANTS

A browse through the literature on vowels of the past 100 years leads to the conclusion that the investigators have always intuitively felt that the interpretation of vowel sounds by a listener might be related to phenomena often loosely referred to as resonance, tuning, natural frequencies, amplified overtones, formants, zones of amplification, *etc etc.*

In judging the scientific level of the early literature on vowels one must bear in mind the lack of physical tools and mathematical insight which was characteristic of many investigators of those days. In the course of time a great confusion of terms came into being.

In 1928 Hermann Gutzmann[1] still distinguishes between the FORMANT theory of L. Hermann and the OVERTONE theory of Helmholtz.

In 1926 Carl Stumpf[2], however, attaches a different meaning to the term "formant" originally coined by L. Hermann as the frequency of a damped oscillation set up in the vocal tract by an air puff emitted by the vocal cords. He applies it to a certain region of overtones. Much confusion has been caused by this deplorable generalisation. As it is not advisable to put the clock back the term formant in its present vagueness should be maintained in the hope that the authors of to-day will not fail to indicate what they mean by this term in their publications.

Strictly speaking, an investigator who calculates the vibrations

[1] H. Gutzmann, *Physiologie der Stimme und Sprache* (Braunschweig 1928, 2 Auflage), p. 128.

[2] C. Stumpf, *Die Sprachlaute* (Springer Verlag Berlin 1926), p. 63.

of the vocal tract does not need to be primarily interested in perception. His first aim is to find the time course of the vibrations that leave the mouth during the production of vowel sounds. No doubt he does know that these time functions carry 'something' to the ear of the listener and that certain changes in that 'something', brought about by changes in the configuration of the vocal tract, convey a meaning to the listener, but for the time being this is not his problem yet. Nevertheless, he may hope that the mathematical equations describing the behaviour of the vocal tract allow him to define mathematical quantities (with the dimension of a frequency or a time) that are characteristic of a certain configuration or geometric changes of the tract. It may be expected that there will be more than one way to do so.

When there is sound in the air the air pressure P at a certain location in space shows small variations denoted by p around its average value P_o:

$$P = P_o + p \tag{III.1}$$

The variation p is called the sound pressure. It is a so-called scalar which means that it is completely given by only one number though that number depends on time and on the coordinates in space of the point where we measure the sound pressure. This dependence is mathematically indicated as follows:

$$p = p(x,y,z,t) \tag{III.2}$$

where x, y and z are the three dimensions in space and t represents time. We say that p is a function of x, y, z and t.

Also the density ϱ, defined as the mass of the air per unit of volume, displays a variation s around its average value ϱ_o:

$$\varrho = \varrho_o + s \tag{III.3}$$

The variation s is called the condensation. It can be proved that, for the usual small variations met in practice, p is proportional to s in the following way:

$$p = c^2 s \tag{III.4}$$

where c is the well-known velocity with which sound travels in free space. For air of vocal tract temperature we take that velocity as high as 350 m/sec. In practice, because of formula (III.4), p only is mentioned, not in the least because it is easy to measure p by appropriate microphones.

The vibrating air particles move to and fro around their positions of equilibrium. In doing so they cover very small distances and develop very low velocities. The so-called particle velocity v is defined as the distance an air particle travels during a very short time-interval, devided by that interval. Even in loud sounds v reaches peak values of less than 1 m/hour, a snail's pace considered to be paradoxical by many but nevertheless being correct. People without a physical background are inclined to mix up the particle velocity with the velocity of propagation c. One should bear in mind, however, that the particle velocity pertains to the transport of mass whereas the velocity of propagation describes the transport of energy which can take place at a much higher velocity.

Generally speaking it is not sufficient to merely state the absolute value v of the particle velocity. The particle velocity is a vector which is a mathematical way of saying that it has a direction in space. The air particles move to and fro along the so-called streamlines. We need not, however, in this monograph, make an excursion into vector theory because the following simplifications eliminate the need for such an unwelcome digression.

In order to facilitate the calculations the vocal tract which is virtually boomerang-shaped, is bent in such a way that its axis becomes a straight line which at the sime time may serve as our x-axis. Moreover, the streamlines are supposed to be essentially parrallel to the x-axis so that only the velocity-component u in the direction of the x-axis has to be taken into account. The next simplification is to suppose that the particle velocity u is the same throughout a cross-area perpendicular to the axis. The same holds good for the sound pressure p. This means that u and p are functions only of x and t:

sound pressure

$$p = p(x,t) \tag{III.5}$$

particle velocity

$$u = u(x,t) \tag{III.6}$$

By straightening the vocal tract and idealizing its streamlines we have strait-jacketed our problem into a one-dimensional problem. This method is followed by the majority of investigators. Nevertheless, we should not neglect the possibility that the damage inflicted on physical reality by the one-dimensional strait-jacket may render certain refinements in our calculations far-fetched if we accept the accuracy of the calculated formant frequencies as a criterion. On top of that comes a problem which is important to the linguist: with what accuracy does the mathematically defined formant frequency describe the physiological data the nervous system derives from the time-functions presented to the ear?

An advantage of the one-dimensional approach is that the so-called volume velocity U can be defined in the following simple way:

$$U = S\,u \tag{III.7}$$

where S is the area of the cross-section of the tube over which the particle velocity u is supposed to be constant with respect to x. As a matter of fact Webster's equation is based on the volume

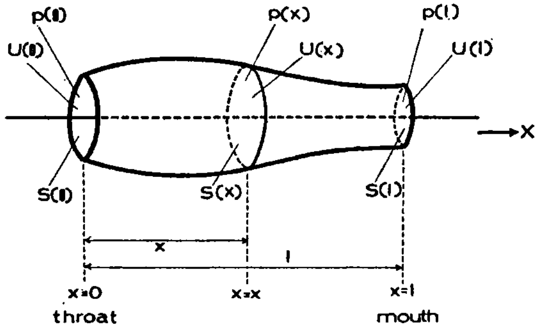


Figure III.1

The vocal tract seen as an acoustic device that transforms a sound pressure $p(o)$ and a volume velocity $U(o)$ at the throat into a sound pressure $p(l)$ and a volume velocity $U(l)$ at the mouth opening.

velocity in order to get the cross-area of the tube, which is a function of x:

$$S = S(x) \tag{III.8}$$

into the picture, that is the differential equation, at all.

From now on our calculations will aim at the determination of sound pressure p and volume velocity U in the vocal tract, especially at the beginning and the end.

Though the average layman will readily accept the loose formulation that the larynx produces a sound which is modified by the vocal tract in a way that is characteristic of the vowel under discussion, he might frown at the idea that in order to calculate that modification it is necessary to take into account two acoustic quantities, the familiar sound pressure and the less popular volume velocity. This necessity, forced on us by nature, has an advantage, however, because it places the vocal tract in the same class with the electric four-terminal networks the theory of which is already well-developed. In the calculation of electric networks the electric voltage as well as the electric current appear in the differential equations. It is very convenient to regard sound pressure and electric voltage as analogous quantities. The same can be said of volume velocity and electric current. We may even go so far as to denote analogous quantities by the same symbol, as is done in figure III.2,

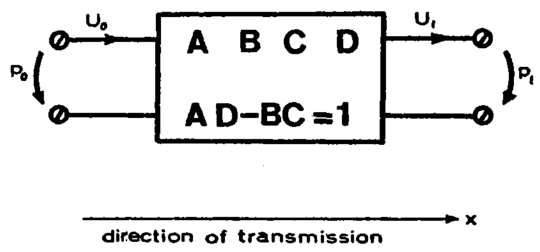


Figure III.2

Equivalent electric four-terminal network of the vocal tract
p_0 : input voltage, analogous to sound pressure at the throat
U_0: input current, analogous to volume velocity at the throat
p_l : output voltage, analogous to sound pressure at the mouth-opening
U_l: output current, analogous to volume velocity at the mouth-opening.

where a passive electric four-terminal network is depicted with the application of acoustic symbols for the electric quantities. In that way the vocal tract comes to the fore in the disguise of an electric network. The analogy between figure III.1 and figure III.2 is obvious.

Electric circuit theory shows, that for sinusoidal oscillations, the following linear relations exist between the quantities at the input and the output:

$$p_o = A p_i + B U_i \tag{III.9}$$

$$U_o = C p_i + D U_i \tag{III.10}$$

The coefficients A, B, C and D are called the general circuit parameters of the network. They depend on the configuration and on the values of the electric components of the network such as resistors, capacitors and solenoids. As a rule they are functions of frequency (remember we suppose sinusoidal vibrations!). In the vast majority of networks these components are so-called lumped, but they may also be distributed continuously as, for instance, in the telephone cable. The vocal tract is analogous to a cable: its acoustic 'components' are continuously distributed though we shall see that, for low frequencies, it may sometimes be regarded as being composed of lumped components. This is the case in the well-known Helmholtz-resonances with which we shall deal later on.

For the usual electric networks there is the following extra-relation between the general circuit parameters:

$$AD - BC = 1 \tag{III.11}$$

In the language of network theory we say that the network obeys the reciprocity theorem but we shall not press this point here. The critical reader will notice that the extra-relation (III.11) allows us to reduce the number of general circuit parameters to a mere three but for reasons of mathematical simplicity one always operates with four parameters, bearing in mind there is a useful relation between them.

If we succeed in determining A, B, C and D for the vocal tract they can be expected to depend on frequency, on the way in which the cross-section varies with respect to x and on the total length of the tract. In order to definitely find the parameters we must first find a differential equation for the vocal tract, next solve it for sinusoidal vibrations, then calculate separately sound pressure and volume velocity and finally introduce the boundary conditions at both ends of the tube. Unfortunately the average linguistic reader does not possess the necessary mathematical background to perform or to understand the sketched procedure. It is a consoling thought, however, that only in special cases this procedure is mathematically possible even for the expert mathematician.

In connection with the twin-tube resonator it will become necessary to calculate the general circuit parameters of a network consisting of two known four-terminal networks in tandem as is shown in figure III.3.

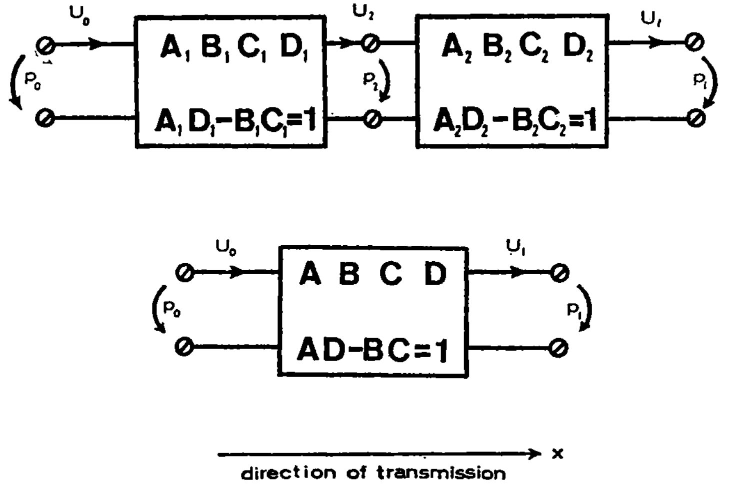


Figure III.3

Two four-terminal networks in tandem seen as one single, resultant network.

Each network has its own set of relations:

$$p_o = A_1 p_2 + B_1 U_2 \quad \text{(III.12)} \qquad p_2 = A_2 p_l + B_2 U_l \quad \text{(III.14)}$$
$$U_o = C_1 p_2 + D_1 U_2 \quad \text{(III.13)} \qquad U_2 = C_2 p_l + D_2 U_l \quad \text{(III.15)}$$

The calculation is based on the principle that the output volume velocity U_2 of the first network is at the same time the input volume velocity of the second network. Furthermore the output sound pressure p_2 of the first network is also the input sound pressure of the second network.

It is possible to eliminate p_2 and U_2 from the equations (III.12), (III.13), (III.14) and (III.15). Omitting the mathematics we get:

$$p_o = (A_1A_2 + B_1C_2)p_i + (A_1B_2 + B_1D_2)U_i \qquad \text{(III.16)}$$
$$U_o = (A_2C_1 + C_2D_1)p_i + (D_1D_2 + C_1B_2)U_i \qquad \text{(III.17)}$$

These expressions describe a new four-terminal network with the general circuit parameters:

$$A = A_1A_2 + B_1C_2 \qquad \text{(III.18)}$$
$$B = A_1B_2 + B_1D_2 \qquad \text{(III.19)}$$
$$C = A_2C_1 + C_2D_1 \qquad \text{(III.20)}$$
$$D = D_1D_2 + C_1B_2 \qquad \text{(III.21)}$$

Especially equation (III.21) will proove to be important for the calculation of the twin-tube model.

Before going on, we must realize that the general circuit parameters as we have defined them, refer to the hypothetical case where a sinusoidal vibration is being transmitted through the vocal tract from the throat to the mouth. This situation certainly does not represent the actual mode of action of the throat which does not produce sinusoidal vibrations.

Sinusoidal vibrations are introduced as mathematical tools for solving or simplifying the differential equations that govern the physical phenomena in question. Once the reaction of a network to a sinusoidal excitation has been determined, it is mathematically possible to predict the reaction of that network to an arbitrary time function. This possibility is based on the fact that, as a rule, an arbitrary time function may be considered as the sum of a, usually large, number of sinusoidal time functions with different frequencies. This step, back to reality, from the frequency concept to the actual excitation of the vocal tract, is more often than not, omitted. The consequences of this omission for the definition of the formants will be discussed in some detail in this chapter.

We shall now return to and start from fig. III.2 which, as has been said already, refers to the hypothetical case where a sinusoidal vibration is transmitted through the vocal tract. As visualized in fig. III.4, the throat is treated as a sinusoidal electromotoric force e in series with an internal impedance[3] Z_o. The mouth-opening is considered as a load-impedance Z_l.

Evidently

$$p_o = e - U_o Z_o \quad \text{(III.22)}$$

and

$$p_l = U_l Z_l \quad \text{(III.23)}$$

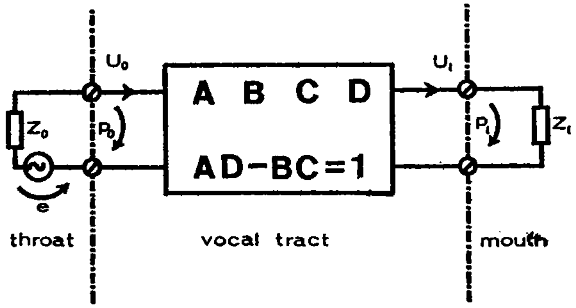


Figure III.4

The throat considered as a source of sinusoidal vibrations.

It is possible to calculate U_l as well as U_o by combining (III.9), (III.10), (III.22) and (III.23), which finally yields:

$$U_l = \frac{e}{B + DZ_o + AZ_l + CZ_oZ_l} \quad \text{(III.24)}$$

$$U_o = \frac{e}{Z_o + \dfrac{AZ_l + B}{CZ_l + D}} \quad \text{(III.25)}$$

In practically all calculations of the vocal tract one permits oneself the following, simplifying assumptions:

[3] The (acoustic) impedance is defined here as the sound pressure divided by the volume velocity.

$$Z_l \to 0 \qquad \text{(III.26)}$$

and

$$Z_o \to \infty \qquad \text{(III.27)}$$

Equation (III.26) persuades us to consider the mouth-opening as a dead short circuit. Consequently the sound pressure $p_l = 0$ (or at least very low), in other words there is always a pressure node in the mouth-opening.

Equation (III.27) expresses the belief that for sinusoidal vibrations the internal impedance of the throat is very high and independent of frequency.

By introducing (III.26) and (III.27) into (III.24) and (III.25) we simply get:

$$U_l = \frac{e}{Z_o D} \qquad \text{(III.28)}$$

and

$$U_o = \frac{e}{Z_o} \qquad \text{(III.29)}$$

Combination of (III.28) and (III.29) yields:

$$U_l = \frac{U_o}{D} \qquad \text{(III.30)}$$

Let us first precise the way in which the quantities in (III.30) depend on frequency.

The driving volume velocity U_o, determined by (III.29), has the same amplitude for all frequencies.

The general circuit parameter D of the vocal tract does depend on frequency. It is, mathematically speaking, a complex quantity with, in general, a real and an imaginary part. In order to express the variation of D with the frequency $\omega = 2\pi f$ it would be possible to write $D = D(\omega)$, but in view of what follows later on in this chapter it is advisable to use the following notation:

$$D = D(j\omega) \qquad \text{(III.31)}$$

with

$$j = \sqrt{-1} \qquad \text{(III.32)}$$

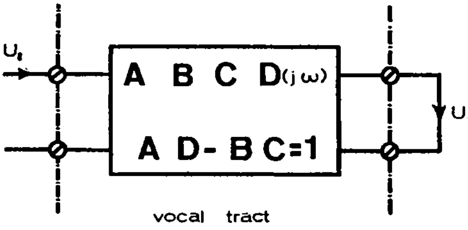


Figure III.5

The vocal tract driven by a sinusoidal constant volume velocity U_0 at the throat side and loaded by a short circuit (pressure node) at the mouth side.

From (III.30) it is very clear that U_l varies with frequency (both in amplitude and phase) because D does so. The graph depicting $|U_l|$, that is the amplitude of U_l, as a function of ω or f is known as the frequency response curve, an example of which is shown in fig. III.6.

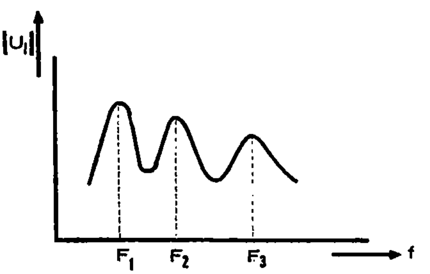


Figure III.6

Example of a frequency response curve. The frequency locations of the peaks are called the resonance frequencies and may be defined as the resonance FORMANTS.

As a rule the curve will show peaks (relative maxima).

The frequency locations of these maxima (as well as those of the minima!) are given by the roots of the following equation:

$$\frac{d}{d\omega}|D(j\omega)| = 0 \qquad \text{(III.33)}$$

The roots corresponding to the maxima are called the resonance frequencies ω_1, ω_2, ω_3 *etc* (F_1, F_2, F_3 *etc*). It is very tempting indeed to define the resonance frequencies as the formant frequencies. In order to avoid confusion, the formants defined in this way, we shall call the resonance formants.

The method of defining the formants as resonance frequencies is open to criticism. We might as well label this method as the sweep-frequency method because the throat is replaced by a constant volume velocity source U_o with a variable frequency. The way in which U_l at the (short circuited) mouth-side reacts to that source is used as a vehicle for defining the formant frequencies. In order to measure the resonance frequencies of a real talker it is necessary to remove his throat and to replace it by a small loudspeaker that produces a constant volume velocity. Most talkers would object against such a procedure.
As a matter of fact Jw. van den Berg has actually (and very ably) applied the sweep-frequency method in a patient whose larynx had been removed by surgery, as is illustrated in fig. III.7.

Now we have to accept the simple fact that a normal talker does not utilize a sweep-frequency method for conveying to a listener data on the shape of his vocal tract during the production of a vowel. His throat generates air-puffs instead. Even a spectrograph does not indicate resonance frequencies in its picture: it can at best depict spectral lines the frequencies of which are multiples of F_o, the fundamental frequency of the vocal folds, corresponding to the number of air-puffs produced per second by the larynx. Relatively strong spectral lines may be defined as FILTER FORMANTS, however, but they do not coincide with the resonance formants.

One might even doubt whether transmission theory with its etceteras is applicable to the vocal tract at all. Transmission theory was primarily created for the purpose of designing transmission links of which both the sending and receiving ends were available for measurement. At the sending end of the channel any type of signal could be expected. Transmission theory with its frequency concept provides a tool for predicting how an arbitrary signal will be distorted by the channel in question. It describes the channel, not the signal.

The ear of the listener, backed by the nervous system, is not a

Figure III.7

Replacement of the surgically removed pathological throat of a patient by a small loudspeaker producing a sinusoidal vibration. Taken from: J. van den Berg, Physica van de stemvorming, met toepassingen, 1953, Diss. Groningen.

transmission meter, it is a signal-detector, as it were caring very little for a process to which it has no access.

It is not necessary, however, to throw overboard the routine of transmission theory, though I still believe that habit is often stronger than progress. By keeping intact the frequency concept, we can predict how a vocal tract will react to the application of a so-called step-function at the throat-side.

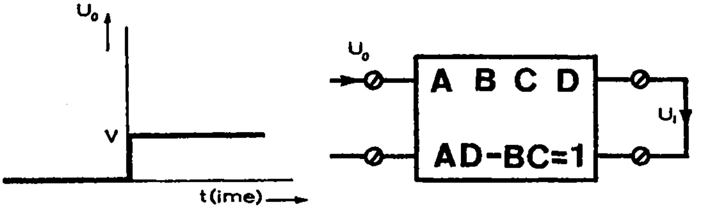


Figure III.8

Applying a volume velocity step to the vocal tract at the throat-side. The volume velocities U_o and U_l are no longer sinusoidal!

As shown in figure III.8, we suddenly make the (no longer sinusoidal!) volume velocity U_o jump from the value o to the value $U_o = V$ and see how the vocal tract reacts to that disturbance. This is called the method of Heaviside.

Mathematically speaking, this method of driving the vocal tract by a step-function is as arbitrary as driving it by a sine function but we shall see that it leads to another formant definition that is more in harmony with the mode of action of the larynx.

The Heaviside method, as it were parasitically, 'borrows' the concept of the general circuit parameter $D(j\omega)$ from its competitor, the sweep-frequency method. By way of improvement, however, it replaces the purely imaginary variable $j\omega$ by the complex variable p:

$$j\omega \to p \tag{III.34}$$

so that we get:

$$D = D(p) \tag{III.35}$$

with

$$p = a + jb \tag{III.36}$$

Taking (III.35) as a starting point we can make the following elaborations.

For $p = 0$ we have

$$D = D(0) \tag{III.37}$$

By differentiation we get:

$$\frac{d}{dp} D(p) = D'(p), \tag{III.38}$$

a matter of notation.
Furthermore we define

$$p_i = a_i + jb_i \tag{III.39}$$

as one of the roots of

$$D(p) = 0 \tag{III.40}$$

Now the scene is set for the presentation of Heaviside's formula

$$U_t = \frac{V}{D(0)} + V \sum \frac{\varepsilon^{p_i t}}{p_i D'(p_i)} \tag{III.41}$$

which is claimed[4] to describe the reaction of the vocal tract to the unit-impulse V.
The right hand member of (III.41) shows a constant component

$$\frac{V}{D(0)} \tag{III.42}$$

corresponding with a 'lift' of the sound curve and the, acoustically more interesting term

$$V \sum \frac{\varepsilon^{p_i t}}{p_i D'(p_i)} \tag{III.43}$$

which represents the superposition of a number of so-called damped oscillations. Without losing ourselves in too many details we can state that the roots p_i usually come in pairs and that, for instance a root

$$p_1 = a_1 + jb_1 \tag{III.44}$$

[4] The proof of Heaviside's formula can be found in any good textbook on operational calculus.

is accompanied by its conjugate

$$p_2 = a_2 + jb_2 = a_1 - jb_1 \tag{III.45}$$

Now its appears, that (III.43) produces sums like

$$\frac{\varepsilon^{p_1 t} - \varepsilon^{p_2 t}}{2j} \tag{III.46}$$

for which can be written when (III.44) and (III.45) are applied:

$$\frac{\varepsilon^{a_1 t}(\varepsilon^{jb_1 t} - \varepsilon^{-jb_1 t})}{2j} \tag{III.47}$$

or

$$\varepsilon^{a_1 t} \sin b_1 t \tag{III.48}$$

As calculations show that always $a_1 < 0$ this formula represents a damped oscillation as depicted in fig. III.9.

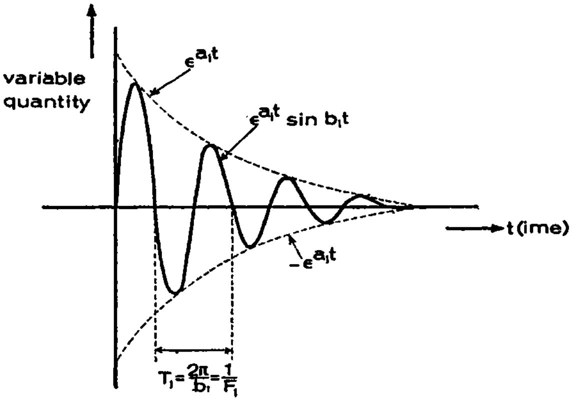


Figure III.9

Graphic representation of a damped oscillation characterized by the frequency $b_1 = 2F_1$ and the time constant a_1. The frequency F_1 betrays itself in the curve as the reciprocal of the time-interval T_1 (the period) between two subsequent zero-crossings with the same polarity. There are indications, that the nervous system of the listener is able to interpret in some way the interval T_1.

The frequency of this oscillation is given by

$$b_1 = 2\pi F_1 \tag{III.49}$$

whereas the rate of decay is governed by the time-constant $\frac{1}{a_1}$.

We shall not, in this chapter, occupy ourselves with the time-constant but concentrate ourselves on the frequency F_1 (or b_1) instead.

F_1 is called one of the NATURAL frequencies of the system because it represents the frequency of one of the collection of damped oscillations the system produces when left to itself, after it has been subjected to an initial shock-like disturbance. It is the frequency of a typical transient.

As apparent from figure III.8 the natural frequency is directly visible as the reciprocal of the time-interval T_1 between two subsequent zero-crossings with the same polarity.

As early as in the nineteenth century L. Hermann[5], one of the first to produce visible sound curves, drew attention to the damped oscillations so clearly discernible in the sound wave that leaves the mouth in real speech. These oscillations prove that the velocity puffs leaving the larynx must have a step-like ending (thereby betraying that the vocal cords close with a snap) which produces powerful damped oscillations in the vocal tract, and a much more gradual beginning which only generates weak oscillations. Therefore it is attractive to define the natural frequencies as formants, the NATURAL formants.

So at the moment we are faced with the choice between two possible definitions of the formants, the resonance formants and the natural formants. It is possible to keep this choice in the domain of mathematics. Both types have been derived from the same general circuit parameter $D(j\omega)$ by subjecting it to different mathematical procedures as shown by (III.33) and (III.40). Strictly speaking both procedures are arbitrary though the definition of the natural formant is better adapted to the actual production of the vowels.

We need not stress the point here because later on we shall see that both definitions will coincide when all dissipation in the vocal tract is neglected, as is, for instance, done in the method of Webster to be described in the next chapter.

[5] L. Hermann, *Phonophotographische Untersuchungen* (*Pflügers Arch.*, 1889-1895).

IV

ON WEBSTER'S HORN EQUATION

Though Webster's equation was originally created for the design of loudspeaker horns, it is very tempting indeed to try to apply it to the vocal tract. We shall briefly outline its derivation in order to expose the underlying simplifications.

In Webster's method the rigorous three-dimensional mathematical treatment with the walls as a bounding surface is playfully avoided by incorporating the wall already in the condition of continuity which, as shown in figure IV.1, is formulated via a one-dimensional approximation.

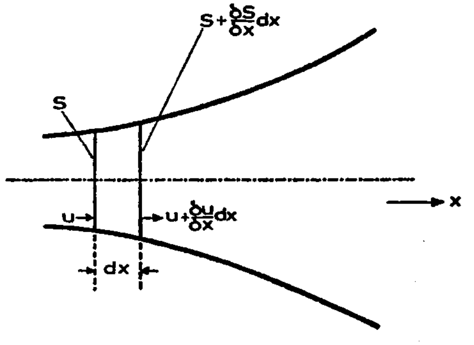


Figure IV.1
A hard-walled tube with variable cross-area.

The tube is cut into slices with a thickness of *dx*. Then one calculates what happens within a slice during the short time-interval *dt*.

During the interval dt the following mass has entered the slice through the surface S:

$$\varrho \,.\, u \,.\, S \,.\, dt \tag{IV.1}$$

where u is the particle velocity and ϱ the density (mass per unit of volume).

During the same interval dt the following mass escapes from the slice through the surface $S + \frac{\delta S}{\delta x} dx$:

$$\varrho \,.\, (u + \frac{\delta u}{\delta x} dx)(S + \frac{\delta S}{\delta x} dx)\, dt \tag{IV.2}$$

or, neglecting terms with $(dx)^2$:

$$\varrho \left[u \,.\, S \,.\, dt + \frac{\delta (Su)}{\delta x} dx\, dt \right] \tag{IV.3}$$

By subtracting (IV.1) from (IV.3) we find that the slice has seemingly produced the mass:

$$\varrho \, \frac{\delta (Su)}{\delta x} dx\, dt \tag{IV.4}$$

According to the principle of continuity no creation of mass is possible so that this mass has been delivered at the cost of the density ϱ in the slice. By decreasing its density the slice has contributed the following mass:

$$- S\, dx \frac{\delta \varrho}{\delta t} dt \tag{IV.5}$$

By demanding that (IV.4) equals (IV.5) we get, after dividing both members by the product $dx\, dt$:

$$\boxed{\varrho \frac{\delta (Su)}{\delta x} = - S \frac{\delta \varrho}{\delta t}} \tag{IV.6}$$

This equation is the crux of Webster's method. The cross area S has been elegantly included in a differential quotient, in that way even giving rise to the volume velocity Su already defined in the foregoing chapter. We must not be blind, however, to the cost of elegance. It was necessary to suppose that the air particles entered and left the slice at right angles, leaving in mid-air how such a miracle might be accomplished by goings-on in the slice. In other words, we have tampered with the stream-lines.

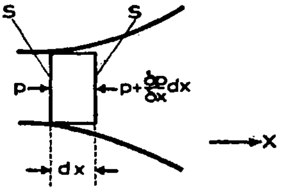


Figure IV.2
Deriving the dynamic equation.

Figure IV.2 shows how a thin cylindrical slice with the volume Sdx and the mass ϱSdx can be thought as moving to and fro in the direction of the x-axis under influence of a force which is being furnished by the pressure difference between both surfaces S.

The resultant force in the positive direction of x equals:

$$-S\frac{\delta p}{\delta x}dx \tag{IV.7}$$

The acceleration $\frac{du}{dt}$ given by this force to the mass of the cylinder is determined by

$$-S\frac{\delta p}{\delta x}dx = \varrho S dx\frac{du}{dt} \tag{IV.8}$$

$$\frac{\delta p}{\delta x} = -\varrho\frac{du}{dt} \tag{IV.9}$$

The appearance of a total derivative in the right-hand member of (IV.9) is not the result of a printing error. The dynamic law dictates a total derivative and is also restricted to movements along the stream-lines. Strictly speaking

$$\frac{du}{dt} = \frac{\delta u}{\delta t} + \frac{\delta u}{\delta x} u \tag{IV.10}$$

but, for the small velocities met in phonetic problems, the second term in the right-hand member is always neglected so that we get

$$\frac{du}{dt} \approx \frac{\delta u}{\delta t} \tag{IV.11}$$

Combination of (IV.9) and (IV.11) yields

$$\boxed{\frac{\delta p}{\delta x} = -\varrho \frac{\delta u}{\delta t}} \tag{IV.12}$$

the equation of motion containing only partial derivatives, a property which is very convenient for the derivation of a manageable partial differential equation in p or u. For that purpose we also need

$$p = c^2 s \tag{III.4}$$

and

$$\varrho = \varrho_o + s \tag{III.3}$$

Because we wish to consider the vocal tract as an acoustical four-terminal network it will appear to be advantageous to derive a partial differential equation in θ, the so-called velocity potential which is defined in the following way

$$\boxed{u = \frac{\delta \theta}{\delta x}} \tag{IV.13}$$

Also p can be derived from θ as follows

$$\boxed{p = -\varrho_o \frac{\delta \theta}{\delta t}} \tag{VI.14}$$

This formula can be proved by partially differentiating both members to x, which yields

$$\frac{\delta p}{\delta x} = -\varrho_o \frac{\delta^2 \theta}{\delta t \delta x} = -\varrho_o \frac{\delta u}{\delta t} \tag{IV.15}$$

Equation (IV.15) is the same as (IV.12) provided we replace ϱ by ϱ_o. The replacement of ϱ by ϱ_o when ϱ appears as a coefficient is a sacrifice we have to accept anyhow if we want to arrive at a manageable differential equation.

By combining (IV.6), (III.4), (III.3), (IV.13) and (IV.14) we finally find:

$$\frac{\delta^2\theta}{\delta x^2} + \frac{1}{S}\frac{\delta S}{\delta x}\frac{\delta\theta}{\delta x} - \frac{1}{c^2}\frac{\delta^2\theta}{\delta t^2} = 0 \qquad \text{(IV.16)}$$

This expression is known as Webster's horn equation. In the literature on this subject-matter it is generally held that the following two drawbacks of Webster's equation prevent its application to phonetic problems:

1. Webster's equation does not contain terms that can take care of dissipation so that only loss-free tubes can be calculated.
2. It is impossible to present a general solution of Webster's equation; it can only be solved in special cases.

In this monograph, however, we confine ourselves to the calculation of loss-free models. Moreover, we only need to calculate the exponential horn, a special case in which Webster's equation can be solved indeed.

Let us first, as usual, suppose sinusoidal vibrations and define $\theta(x,t)$ as:

$$\theta(x,t) = \o(x)\varepsilon^{j\omega t} \qquad \text{(IV.17)}$$

Introduction of (IV.17) into (IV.16) yields

$$\frac{d^2\o}{dx^2} + \frac{1}{S}\frac{dS}{dx}\frac{d\o}{dx} + \frac{\omega^2}{c^2}\o = 0 \qquad \text{(IV.18)}$$

We may use total derivatives now because we have freed ourselves from the time dimension.

As is apparent from the formulas in fig. IV.3 the differential equation folds down to

$$\frac{d^2\phi}{dx^2} + m\frac{d\phi}{dx} + \frac{\omega^2}{c^2}\phi = 0 \tag{IV.19}$$

Suppose there is a solution

$$\phi(x) = A\varepsilon^{bx} \tag{IV.20}$$

where A and b represent constants, then introduction of this solution into (IV.19) yields:

$$b^2 + mb + \frac{\omega^2}{c^2} = 0 \tag{IV.21}$$

or

$$b = -\frac{m}{2} \pm \sqrt{\frac{m^2}{4} - \frac{\omega^2}{c^2}} \tag{IV.22}$$

This can be written as follows

$$b = -\frac{m}{2} \pm j\frac{\omega}{c}\sqrt{1 - \frac{m^2c^2}{4\omega^2}} = -\frac{m}{2} \pm j\frac{\omega}{v} \tag{IV.23}$$

where

$$v = \frac{c}{\sqrt{1 - \frac{m^2c^2}{4\omega^2}}} \tag{IV.24}$$

The role of v will become apparent when we study the complete solution for θ as a function of x and t:

$$\theta(x,t) = A_1\varepsilon^{-\frac{m}{2}x + j\omega\left(t + \frac{x}{v}\right)} + A_2\varepsilon^{-\frac{m}{2}x + j\omega\left(t - \frac{x}{v}\right)} \tag{IV.25}$$

There are two constants, A_1 and A_2, because there are also two roots of b that fulfil equation (IV.22).

The second term of the right-hand member of (IV.25) represents a wave travelling in the positive direction of x at the speed v, in other words travelling from the throat to the mouth.

The first term of the right-hand member represents a wave travelling in the negative direction of x at the speed v, so from the mouth to the throat.

As equation (IV.24) shows, however, for frequencies above $\frac{mc}{2}$, the velocity $v > c$, which means that the waves travel at supersonic speeds! Needless to say that this state of affairs does not correspond to physical reality. Let us call this consequence of Webster's horn equation the SUPERSONIC PARADOX.

For the frequency $\omega = \frac{mc}{2}$ the waves even travel at an infinite speed.

Below that frequency the velocity v becomes imaginary and there is no wave propagation at all in the tube. There the exponential horn is supposed to have a cut-off frequency. We must, however, refer this cut-off frequency to the domain of science-fiction because in reality the vocal tract easily conducts the breathstream which represents the frequency 0.

So we see, that Webster's equation leads to the supersonic paradox and is unduly breath-taking as well.

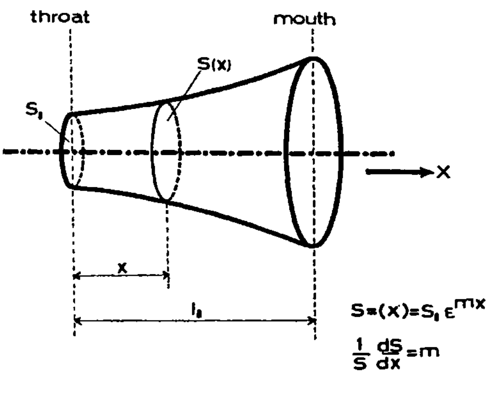


Figure IV.3

The exponential horn

The quantity m is called the flare of the horn.

There is no reason, however, for discrediting Webster's equation on basis of the supersonic paradox and the paradoxical cut-off frequency. These two paradoxes are nothing but seeming problems: the decomposition of the stationary solution $\theta(x,t)$ into two travelling waves as revealed by formula (IV.25), is nothing but a mathematical feature. Physically speaking, there are no two independent waves that run in opposite directions at supersonic speeds; there is nothing but the stationary wave pattern!

Interesting enough though, in a tube with constant cross-area, in other words with $m = o$, the mathematical waves 'travel' at the normal speed c, as shown by formula (IV.24). This seemingly normal behaviour lures our attention away from the fact, that even in that case there are no two independent travelling waves: physically speaking, there 'is' a stationary wave pattern.

Though we shall not, in this monograph, calculate the conical horn, we point to the interesting and, at the same time, misleading fact, that in a conical horn the mathematical waves also travel at the speed c, in that way providing themselves with a physical 'alibi'.

In the next chapter we shall, among other things, calculate the general circuit parameters of the exponential horn, at the same time presenting its formant formula.

V

THE GENERAL CIRCUIT PARAMETERS AND THE FORMANTS OF THE EXPONENTIAL HORN

In the foregoing chapters we have seen that a manageable calculation of the vibrations in the vocal tract boils down to the calculation of a much simpler one-dimensional model. Such a model does not pretend to predict the real values of sound pressure, particle velocity and density in every corner of the vocal tract!

Taking the formant frequencies of the model as a criterion for its worthwhileness, we should feel most satisfied when the formant positions predicted with the aid of the model are in fair agreement with those measured at the real vocal tract.

One of the 'selling points' of a workable model is that variation of its parameters corresponds with demonstrable articulatory actions. A model becomes a caricature, however, when it exhibits properties that are physically untenable. In the case of the exponential horn Webster's equation only seemingly leads to physical nonsense: the supersonic paradox and the paradoxical cut-off frequency are nothing but mathematical pleasantries.

Starting from the general horn equation:

$$\frac{d^2\phi}{dx^2} + \frac{1}{S}\frac{dS}{dx}\frac{d\phi}{dx} + \frac{\omega^2}{c^2}\phi = 0 \qquad \text{(IV.18)}$$

we may, without actually solving it, derive some interesting properties as to the influence of growth of the vocal tract on the formant positions. Let us first consider the factor

$$\frac{1}{S}\frac{dS}{dx} = \frac{d}{dx} \ln S \qquad \text{(V.1)}$$

in formula (IV.18). Suppose we multiply *all* cross-areas by the same factor a, then

$$\frac{d}{dx} ln(aS) = \frac{d}{dx}(ln\, a + ln\, S) = \frac{d}{dx} ln\, S \qquad \text{(V.2)}$$

This property gives rise to the law of proportional transversal growth: the formant positions of the tube do not change when all cross-areas are multiplied by the same factor, provided the boundary conditions do not depend on S.

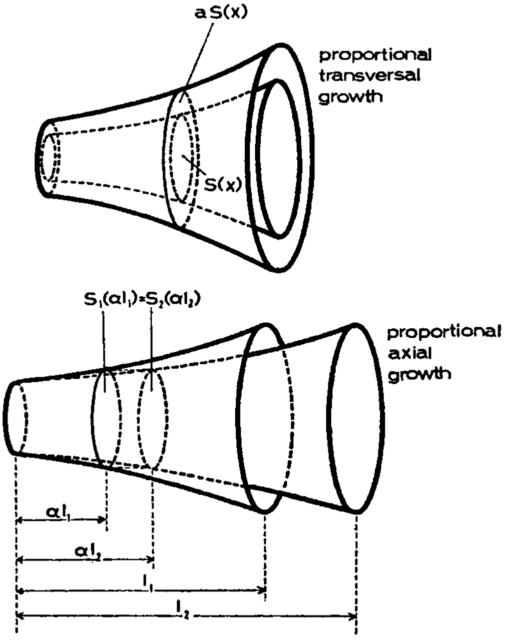


Figure V.1
The two types of proportional growth.

In order to illustrate another type of proportional growth we introduce the new variable α given by

$$x = \alpha l \qquad \text{(V.3)}$$

In that case the horn equation may be written as follows:

$$\frac{d^2\o}{d\alpha^2}+\frac{1}{S}\frac{dS}{d\alpha}\frac{d\o}{d\alpha}+\frac{\omega^2 l^2}{c^2}\o=0 \tag{V.4}$$

Let us first of all define proportional axial growth. Suppose we have a tube, characterized by the index 1, the cross-area of which depends on x in the following way:

$$S = S_1(x) \tag{V.5}$$

When we stretch this tube 1 so that all its axial dimensions are multiplied by the same factor, as is shown in figure V.1 a new tube, tube 2, comes into being.

The relation between the tubes 1 and 2 is given by

$$S_1(\alpha l_1) = S_2(\alpha l_2) \tag{V.6}$$

where a is identical for both tubes.

For tube 1 we have the following horn equation:

$$\frac{d^2\o}{d\alpha^2}+\frac{d}{d\alpha}[lnS_1(\alpha l_1)]\frac{d\o}{d\alpha}+\left(\frac{\omega l_1}{c}\right)^2\o=0 \tag{V.7}$$

For tube 2 we arrive at:

$$\frac{d^2\o}{d\alpha^2}+\frac{d}{d\alpha}[lnS_2(\alpha l_2)]\frac{d\o}{d\alpha}+\left(\frac{\omega l_2}{c}\right)^2\o=0 \tag{V.8}$$

In view of (V.6) it becomes clear that both tubes may be described by the same horn equation provided we drive them with different frequencies ω_1 and ω_2 between which the following relation exists:

$$\omega_1 l_1 = \omega_2 l_2 \tag{V.9}$$

This means that a formant frequency F_1, derived from (V.7), will correspond to a formant frequency F_2, derived from (V.8), in the following way:

$$F_1 l_1 = F_2 l_2 \tag{V.10}$$

which leads to the law of proportional axial growth: when all axial dimensions of the tube are multiplied by the factor a, its formants change inversely proportional to that same factor a.

These two laws of growth were also stated and derived by Ungeheuer[1].

With respect to the velocity of sound c we may formulate the following law: when the velocity of sound of the medium in the tube is changed by a factor a, the formant frequencies change proportional to the same factor.

We shall now calculate the general circuit parameters of the exponential horn, the shape of which is determined by:

$$S(x) = S_0 \varepsilon^{m.x} \tag{V.11}$$

It is possible to express the flare m in terms of the cross-areas S_0 and S_l at the beginning and the end of the exponential horn, and the length l of the horn as follows:

$$m = \frac{1}{l} ln \frac{S_l}{S_o} \tag{V.12}$$

This formula has the following consequences: during proportional axial growth of an exponential horn its flare changes inversely proportional to its length. On the other hand the flare is not affected by proportional transversal growth.

We shall now return to the exponential horn and calculate its general circuit parameters. For that purpose we shall rewrite (IV.25) as follows:

$$\theta(x,t) = A_1 \varepsilon^{b_1 x} \varepsilon^{j\omega t} + A_2 \varepsilon^{b_2 x} \varepsilon^{j\omega t} \tag{V.13}$$

with

$$b_1 = -\frac{m}{2} + \sqrt{\frac{m^2}{4} - \frac{\omega^2}{c^2}} = -\frac{m}{2} + j\sqrt{\frac{\omega^2}{c^2} - \frac{m^2}{4}} \tag{V.14}$$

and

$$b_2 = -\frac{m}{2} - \sqrt{\frac{m^2}{4} - \frac{\omega^2}{c^2}} = -\frac{m}{2} - j\sqrt{\frac{\omega^2}{c^2} - \frac{m^2}{4}} \tag{V.15}$$

Equation (V.13) enables us to determine both p and U.

We know already that

$$p = -\varrho_o \frac{\delta\theta}{\delta t} \tag{IV.14}$$

[1] G. Ungeheuer, *Elemente einer akustischen Theorie der Vokal-artikulation* (Springer-Verlag 1962).

whereas (IV.13) gives rise to

$$U = Su = S\frac{\delta\theta}{\delta x} \tag{V.16}$$

The expressions (V.13), (IV.14) and (V.11) make it possible to treat the exponential horn as an acoustical four-terminal network, to calculate its general circuit parameters and, finally, to find a simple formula for its formants.

Though p and U are, of course, functions of time, it is sufficient to deal only with their amplitudes, which are merely functions of x. We may rightly do so because only sinusoidal functions of time are taken into consideration. Although we indicated the amplitude of the velocity potential $\theta(x,t)$ by the separate symbol $\phi(x)$, we shall simply denote the amplitude of $p(x,t)$ as $p(x)$ or even as p. We shall also denote the amplitude of $U(x,t)$ by $U(x)$ or simply by U.

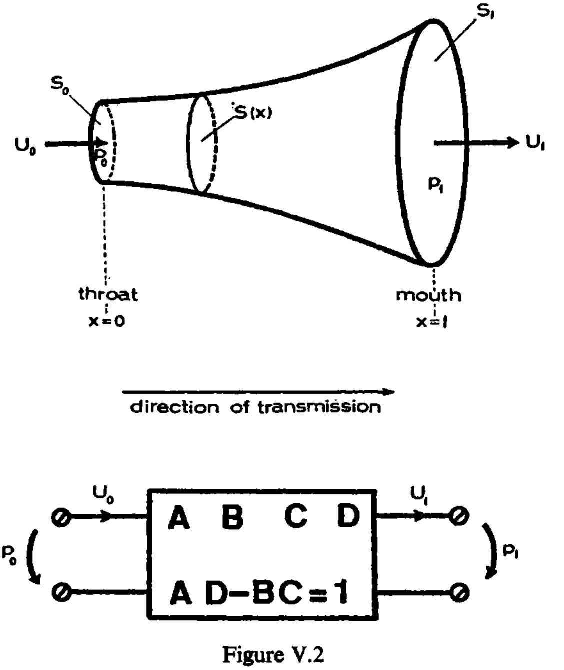


Figure V.2
The general horn seen as a four-terminal network.

As already stated in Chapter III, and again illustrated in figure V.2, the general circuit parameters *A*, *B*, *C* and *D* are defined as follows

$$p_0 = Ap_l + BU_l \tag{III.9}$$

$$U_0 = Cp_l + DU_l \tag{III.10}$$

Still keeping the same direction of transmission, namely from $x = 0$ to $x = l$, we can rewrite these equations as follows

$$p_l = Dp_0 - BU_0 \tag{V.17}$$

$$U_l = -Cp_0 + AU_0 \tag{V.18}$$

It is in this form that the general circuit parameters will come to the fore in our calculation.

As we confine ourselves to the amplitudes, we can derive from (IV.14)

$$p(x) = -j\omega\varrho_o\phi(x) \tag{V.19}$$

and from (V.16)

$$U(x) = S\frac{d\phi}{dx} \tag{V.20}$$

From (V.13) we can derive

$$\phi(x) = A_1\varepsilon^{b_1x} + A_2\varepsilon^{b_2x} \tag{V.21}$$

from which equation we can determine A_1 and A_2 by demanding that for $x = 0$

$$p(x) = p_o \tag{V.22}$$

and

$$U(x) = U_o \tag{V.23}$$

We finally find, after having applied (V.19) and (V.20):

$$A_1 = -\frac{b_2\dfrac{p_o}{j\omega\varrho_o} + \dfrac{U_o}{S_o}}{b_2 - b_1} \tag{V.24}$$

$$A_2 = \frac{b_1 \dfrac{p_o}{j\omega\varrho_o} + \dfrac{U_o}{S_o}}{b_2 - b_1} \tag{V.25}$$

The following step will be to introduce (V.24) and (V.25) into (V.21), then to apply (V.19) and (V.20), next to put $x = l$, so that we finally find p_l and U_l as functions of p_o and U_o. As a matter of fact we shall find nothing but the equations (V.17) and (V.18) with:

$$A = \frac{S_l}{S_o} \frac{b_2 \varepsilon^{b_2 l} - b_1 \varepsilon^{b_1 l}}{b_2 - b_1} \tag{V.26}$$

$$B = \frac{j\omega\varrho_o}{S_o} \frac{\varepsilon^{b_2 l} - \varepsilon^{b_1 l}}{b_2 - b_1} \tag{V.27}$$

$$C = j \frac{S_l}{\omega\varrho_o} b_1 b_2 \frac{\varepsilon^{b_2 l} - \varepsilon^{b_1 l}}{b_2 - b_1} \tag{V.28}$$

$$D = \frac{b_2 \varepsilon^{b_1 l} - b_1 \varepsilon^{b_2 l}}{b_2 - b_1} \tag{V.29}$$

These are the general circuit parameters we are looking for. It is easy to verify that indeed

$$AB - CD = 1 \tag{V.30}$$

because from the expressions (V.11), V.14) and (V.15) we may derive that

$$\frac{S_l}{S_o} = \varepsilon^{-ml} \quad \text{(V.31) and } \varepsilon^{(b_1 + b_2)l} = \varepsilon^{ml} \tag{V.32}$$

We are now in a position to derive the formant formula of the exponential horn. If we suppose the radiation impedance of the mouthopening to be very low and the impedance of the throat to be very high, the formant frequencies are given by (see Chapter III)

$$D = 0 \tag{V.33}$$

or, taking into account (V.29):

$$b_2 \varepsilon^{b_1 l} - b_1 \varepsilon^{b_2 l} = 0 \tag{V.34}$$

When we finally apply (V.14) and (V.15) we ultimately get:

$$\tan\sqrt{\frac{\omega^2}{c^2}-\frac{m^2}{4}}\,l = -\frac{2}{m}\sqrt{\frac{\omega^2}{c^2}-\frac{m^2}{4}} \tag{V.35}$$

or, applying (V.12)

$$\tan\sqrt{\left[\frac{\omega l}{c}\right]^2-\left[\tfrac{1}{2}ln\frac{S_l}{S_o}\right]^2} = -\frac{1}{\tfrac{1}{2}ln\frac{S_l}{S_o}}\sqrt{\left[\frac{\omega l}{c}\right]^2-\left[\tfrac{1}{2}ln\frac{S_l}{S_o}\right]^2} \tag{V.36}$$

Several interesting facts are shown by (V.34) and (V.35).

In the first place the much derided, paradoxical cut-off frequency comes to the fore as a true formant, as the frequency

$$\omega = \frac{|m|\,.\,c}{2} = \frac{c}{2l}\left|ln\frac{S_l}{S_o}\right| \tag{V.37}$$

obeys the formant formulas.

In the second place, as can be expected, the formant formula (V.36) demonstrates the laws of proportional growth: the areas S_1 and S_0 appear as a quotient so that proportional transversal growth has no influence on the formant frequencies, whereas the frequency ω always appears in the form of the 'package':

$$\frac{\omega l}{c},$$

in that way supporting the law of proportional axial growth.

In this monograph we are only interested in the exponential horn as a means for modifying the shape of the tubes of the twin-tube model. As a model for the complete vocal tract the exponential horn with its parameter m can only produce a very limited vocal gesture[2].

For $m = 0$ the formant formula folds down to simply:

$$\cos\frac{\omega l}{c} = 0 \tag{V.38}$$

[2] The terminology: vocal gesture was introduced by G.E. Peterson, see *Bell Laboratories Record*, XXIX, No 11 (Nov. 1951).

with the roots for $F = \frac{\omega}{2\pi}$:

$$F_1 = \frac{c}{4l},\ F_2 = 3\frac{c}{4l},\ F_3 = 5\frac{c}{4l} \text{ etc.} \tag{V.39}$$

These are the well-known formants of the organ pipe closed at one end.

VI

THE METHOD OF THE LOSS-FREE TWIN-TUBE RESONATOR

As Andrew M. Gleason has so ably said[1]: "When a mathematician meets a problem he cannot solve, like any other scientist he tries to solve instead some related problem which seems to contain only part of the difficulties of the original." Because we could not succeed in finding a general solution for an arbitrarily shaped tube we must fall back on a model that is composed of a, preferably small, number of simple, comparatively problemless, tubes. This method is referred to as 'plumbing' in our laboratory. The behaviour of those tubes in tandem can be predicted via their respective general circuit parameters, as already explained in Chapter III. Of course the question is: how few tubes and how simple?

A model consisting of one single tube with a constant cross-area and the same length as the vocal tract is not very realistic because it can only produce the neutral vowel [ə]. Because the length of the tube is its only relevant parameter, it can only produce formant frequencies that are odd multiples of the first formant.

But already the twin-tube resonator, depicted in figure VI.1, is able to generate the complete gamut of formant positions the real vocal tract has at its disposal.

The use of tubes with constant cross-areas as building blocks has several advantages. One hardly suffers any mental agony when supposing that the streamlines in the tube are parallel with its axis. Furthermore the waves in the tube travel at the ordinary velocity of sound *c*.

[1] Andrew M. Gleason (Harvard), *Science*, Volume 145, Number 3631 (31 July 1964), p 451.

Because in this case $m = 0$ and $S(x) = S$, the general circuit parameters of the tube become very simple. As can easily be derived from the formulas (V.26), (V.27), (V.28) and (V.29), a tube with the cross-area S and the length l has the following parameters

$$A = D = \cos\frac{\omega l}{c} \qquad B = j\frac{\varrho_0 c}{S}\sin\frac{\omega l}{c} \qquad C = j\frac{S}{\varrho_0 c}\sin\frac{\omega l}{c} \tag{VI.1}$$

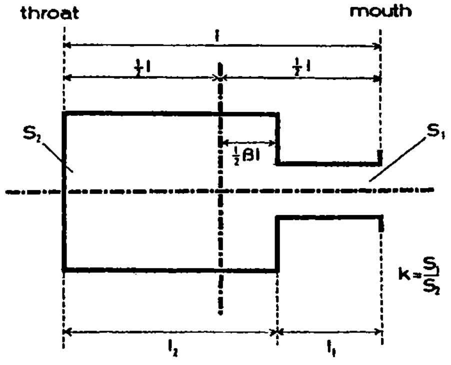


Figure VI.1
The three parameters of the twin-tube resonator:
the total length l
the eccentricity β
the constriction factor $k = \dfrac{S_1}{S_2}$

According to (III.21), two tubes in tandem have the following D parameter:

$$D = D_1 D_2 + C_1 B_2 \tag{III.21}$$

with tube 1 at the sending end.

Using (VI.1) and noticing that the indices 1 and 2 must be interchanged because, adhering to the already existing tradition in literature, we call l_1 the length of the tube at the mouthside, we have

$$D = \cos\frac{\omega l_1}{c}\cos\frac{\omega l_2}{c} - \frac{S_2}{S_1}\sin\frac{\omega l_1}{c}\sin\frac{\omega l_2}{c} \tag{VI.2}$$

The formants of the loss-free twin-tube are found by putting

$$D = 0 \tag{VI.3}$$

or

$$\tan\frac{\omega l_1}{c}\tan\frac{\omega l_2}{c} = \frac{S_1}{S_2} \tag{VI.4}$$

This formula, and its direct graphical interpretation via tan and cot functions, is already known in literature[2 3] but, in our opinion, its possibilities have not been fully explored. Its graphical interpretation, for instance, is greatly facilitated by the introduction of the new variable β, see figure VI.1:

$$l_1 = \tfrac{1}{2}l - \tfrac{1}{2}\beta l \tag{VI.5}$$

$$l_2 = \tfrac{1}{2}l + \tfrac{1}{2}\beta l \tag{VI.6}$$

Apparently $-1 < \beta < 1$.

Using the general theorem

$$\tan x \,.\, \tan y = \frac{\cos(x-y) - \cos(x+y)}{\cos(x-y) + \cos(x+y)} \tag{VI.7}$$

and calling

$$\frac{S_1}{S_2} = k \tag{VI.8}$$

the constriction factor, or opening ratio, we can transform (VI.4) as follows

$$\boxed{\cos\frac{\omega l}{c} = \frac{1-k}{1+k}\cos\beta\frac{\omega l}{c}} \tag{VI.9}$$

It is quite easy to see that the formants of the twin-tube model obey the TWO LAWS OF PROPORTIONAL GROWTH:

[2] G. Fant, *Acoustic theory of speech production* (Mouton, The Hague 1961).
[3] J.L. Flanagan, *Speech Analysis, Synthesis and Perception* (Springer Verlag-Berlin-Heidelberg-New York 1965).

1. multiplication of all cross-areas by the same factor cannot make itself felt in the formant formula (proportional TRANSVERSAL growth);
2. when all axial dimensions of the tube are multiplied by the same factor a, the eccentricity β does not change; consequently the formant will change inversely proportional to a, because ω only appears in the product ωl (proportional AXIAL growth).

Formula (VI.9) is extremely inviting for the construction of a nomogram. It presents the formant frequencies as the points of intersection of two cosines. As the parameters k and β only operate in the right-hand member it is advantageous to keep the parameter l at a constant value (preferably its average value) and deal with its variations later on. In this way the left-hand member becomes a stationary curve in our nomogram, to be intersected by different types of right-hand members.

Let us first investigate the character of the parameter β, called the ECCENTRICITY because it marks the position of the constriction relative to the centre of the vocal tract.

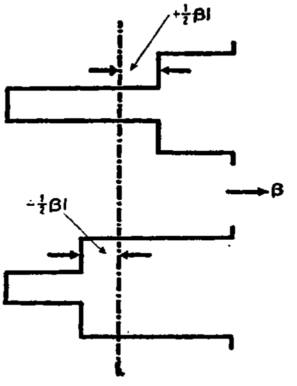


Figure VI.2

Two geometrically different twin-tubes with the same formants (compensatory mechanism).

The first thing to strike us is that the sign of β cannot make itself felt in (VI.9). Consequently the two twin-tubes depicted in figure VI.2 have the same formants because they only differ in the sign of

β. So the twin-tube model houses a compensatory mechanism which applies to *all* formants, F_1, F_2, F_3 *etc.* It is not necessary, of course, to formulate this mechanism via the eccentricity β. Reversal of the sign of β in (VI.5) and (VI.6) comes down to interchanging l_1 and l_2 which, as (VI.4) shows, does not influence the formants as long as S_1 and S_2 are not interchanged.

The twin-tube method is based on the acoustical importance of the joint of the two tubes, the only place between throat and lips where the cross-area makes a step. The joint may be rightly called the point (or place) of articulation.

In formula (VI.9) the factor

$$\frac{1-k}{1+k} = \frac{S_2 - S_1}{S_2 + S_1} = r \qquad \text{(VI.11)}$$

comes to the fore as a typical reflection factor. We are pursuaded to see the reflection-free case ($S_1 = S_2$, corresponding with only ONE tube with constant cross-section) as a sort of basic situation with its characteristic (though neutral) formant pattern and to regard deviations from that pattern as the result of reflections at the joint when we take $k \neq 1$.

It is also possible, however, to see only the mouth and the throat as 'official' reflections and to treat the joint differently. At the joint the volume-velocity U and the sound pressure p remain continuous but there is a step in the particle velocity u. The streamlines in tube 2 are abruptly 'squeezed' into the smaller tube 1.

At the joint

$$S_2 u_2 = S_1 u_1 \qquad \text{(VI.12)}$$

or

$$u_1 = \frac{S_2}{S_1} u_2 \qquad \text{(VI.13)}$$

This step-like behaviour of the particle velocity can be described as a rapid in a river. We shall point out in Chapter VIII how the rapid plays a role in the concept of the so-called fractured stationary waves we can add to the twin-tube resonator.

An interesting class of vowels produced by the twin-tube is represented by $\beta = 0$, in other words the joint is exactly at the centre of the model. In that case the formant formula boils down to:

$$\cos\frac{\omega l}{c} = \frac{1-k}{1+k} \tag{VI.14}$$

As is shown in figure VI.3 the formants of these vowels are found by intersecting the cosine with horizontal lines.

A few typical examples are presented.

For $k = 1$, in other words, when $S_1 = S_2$, the twin-tube degenerates into one single tube. The formants are found as the points of intersection between the cosine and the f-axis. These zero-crossings are given by

$$\cos\frac{\omega l}{c} = 0 \tag{VI.15}$$

which yields the frequencies

$$F_1 = \tfrac{1}{4}\frac{c}{l}, F_2 = \tfrac{3}{4}\frac{c}{l}\ etc. \tag{VI.16}$$

These are, by the way, the well-known resonances of the organ pipe closed at one end, which form a series of odd multiples of one fundamental resonance frequency, namely

$$F_1 = \tfrac{1}{4}\frac{c}{l} \tag{VI.17}$$

Taking $l = 17.5$ cm as an average value for the real vocal tract and $c = 35\,000$ cm/sec for the warm air leaving the throat, we find $F_1 = 500$ Hz(c/s) and $F_2 = 1500$ Hz which is in good agreement with the values found by Peterson and Barney for the neutral vowel [ə] in their classical measurements[4] we shall quote more than once in this monograph.

[4] G. E. Peterson and H. L. Barney, "Control Methods Used in a Study of Vowels", *JASA*, Vol. 24, Nr 2 (March 1952), pp. 175-184.

When $k = 8$ the line will be well below the f-axis because then

$$\frac{1-k}{1+k} = -\tfrac{7}{9} \qquad \text{(VI.18)}$$

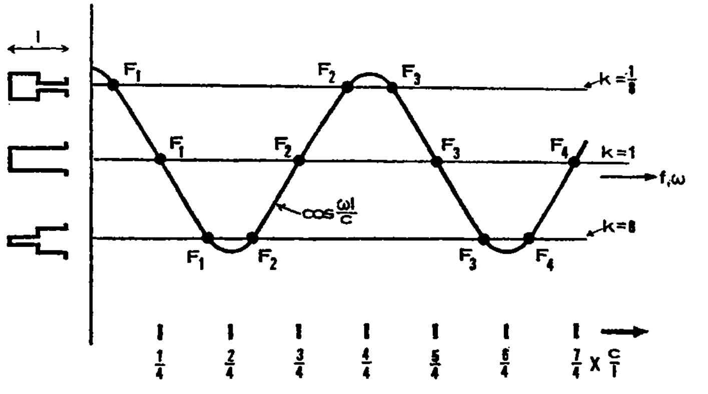


Figure VI.3

Nomogram for the formants of the twin-tube resonator for $\beta = 0$ ($\omega = 2\pi f$).

The formants F_1 and F_2 are close together at both sides of the frequency

$$F = \tfrac{1}{2}\frac{c}{l} \qquad \text{(VI.19)}$$

For $l = 17.5$ cm this centre-frequency corresponds with 1000 Hz. Therefore the twin-tube with

$$\beta = 0,\ k = 8,\ l = 17.5 \text{ cm}$$

could pose as a model for [a] in Dutch.

When, on the other hand, we try $k = \frac{1}{8}$, the line of intersection will be well above the f-axis because now

$$\frac{1-k}{1+k} = +\tfrac{7}{9} \qquad \text{(VI.20)}$$

This configuration is characterized by a low value of F_1 (a typical

Helmholtz resonator!) and a twin-formant around the centre-frequency

$$F = \frac{c}{l} \tag{VI.21}$$

Depending on the value of l there are several possibilities. For instance, for $l = 17.5$ cm the centre-frequency lies at 2000 Hz. As can be read form the nomogram in that case $F_1 = 217$ Hz. Consequently, for $\beta = 0$, $k = \frac{1}{8}$, $l = 17.5$ cm the twin-tube model generates something between [ü] and [i] for the average Dutch talker. For $l = 14.5$ cm, however, the centre-frequency is located at 2414 Hz whereas $F_1 = 262$ Hz. So, in the case of $\beta = 0$, $k = \frac{1}{8}$, $l =$ 14.5 cm the twin-tube model produces [i] in Dutch.

Let us now try some other values of β, keeping k at the value $k = \frac{1}{8}$.

For $\beta = +\frac{1}{3}$, see figure VI.4, the length of the front tube is $l_1 = \frac{1}{3}l$. In order to find the formants we must determine the points of intersection between $\cos\frac{\omega l}{c}$ and $\frac{7}{9}\cos\frac{1}{3}\frac{\omega l}{c}$. There is a F_1 well below $\frac{1}{4}\frac{c}{l}$ and a second formant F_2 at the frequency

$$F_2 = \frac{3}{4}\frac{c}{l} \tag{VI.22}$$

For l = 17.5 cm we have $F_2 = 1500$ Hz. We can read F_1 from the picture, which yields $F_1 = 200$ Hz. This combination is certainly acceptable for [ü] in Dutch. We might remark in passing, that for $\beta = +\frac{1}{3}$ (and also for $\beta = -\frac{1}{3}$) variation of k does NOT influence the position of F_2. As we shall see in Chapter VIII the point of constriction is located in a loop or node in this case.

For $\beta = +\frac{7}{9}$ we have a very short tube in front. Reading both formants from the nomogram, we get $F_1 = 0.139\frac{c}{l}$ and $F_2 = 0.57\frac{c}{l}$.

For $l = 22$ cm, this comes down to $F_1 = 220$ Hz and $F_2 = 900$ Hz, which acceptably describes [u] in Dutch.

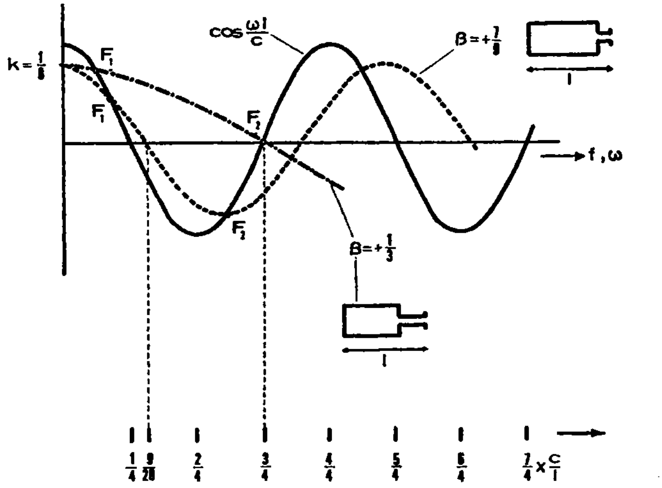


Figure VI.4

The formants of the twin-tube model for $k = \frac{1}{8}$, $\beta = +\frac{1}{3}$ and $\beta = +\frac{7}{9}$.

The few examples presented thus far in this chapter demonstrate the flexibility of the twin-tube model with its parameters k, β and l. By suitable choice of these three basic parameters the twin-tube will generate any combination of F_1 and F_2 the real vocal tract is able to produce.

In Chapter X we shall show how the 3 parameters can be deduced from the real vocal tract, among other things via radiograms.

The k-parameter has three interesting regions:

$k > 1$, describing the vowels traditionally called open
$k = 1$, pertaining to the neutral vowel
$k < 1$, referring to the close vowels.

In words:

in open vowels the widest tube is in front
in close vowels the narrowest tube is in front
in the neutral vowel there is no clear distinction between the wide and the narrow tubes.

We are not sure, however, that every phonetician will use the notions 'open' and 'close' in the above-mentioned sense.

It is a pity that the hope that the β-parameter is in some elegant way related to the traditional back-front opposition is false.

The unambiguously defined β-parameter indicates the position of the acoustically relevant, step-like constriction. For instance, for [u], with $\beta = +\frac{7}{8}$, the real place of articulation is in FRONT, at the lips. Nevertheless, [u] is traditionally labelled as a back vowel because the hump of the tongue is at the back of the mouth (which is, after all, only half-way the vocal tract). The constriction of the vocal tract at the hump of the tongue is not step-like so that it has only a secundary acoustical importance. In articulating [u], space is created just behind the important lip-constriction. Also, the vocal tract is widened at the throat-side. Therefore the hump of the tongue in [u] should be considered as an acoustically harmless 'parking place' for tongue matter.

The l-parameter is not only influenced by the anatomical constraints of the speaker, which depend on age and sex and individual variations. A speaker is able to control the axial length l of his vocal tract by manipulations with his lips, tongue (tip and hump) and larynx. For instance, in [u] the vocal tract begins at the outside of the pouted lips, whereas in [i] it begins just behind the teeth. The role of l comes to the fore in an interesting way in the vector representation of the formants of the twin-tube, which is discussed in the next chapter.

VII

VECTOR REPRESENTATION OF THE TWIN-TUBE FORMANTS

In this chapter we shall confine ourselves to the vector representation of merely the first two formants F_1 and F_2. Perhaps needless to say we are fully aware of the fact that this procedure might not be harmless in all cases, in spite of its apparent advantages.

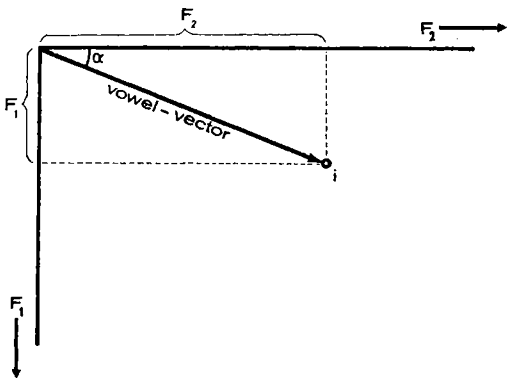


Figure VII.1

Definition of the vowel-vector.
The choice of the axes for F_1 and F_2 harmonizes with the shape of Hellwag's vowel-triangle.

When, in general, two numbers are characteristic of a certain notion, it is possible to represent these numbers as a point in a plane, having these two numbers as co-ordinates. As shown in fig. VII.1, this formal procedure may be applied to the formants F_1 and F_2 of, for instance the vowel [i]. We are fully justified calling this point [i] because its co-ordinates represent the two formants that are thought to be characteristic of the vowel [i] in question.

The line connecting the point [i] with the origin is defined as the vowel vector [i]; its angle α is called its argument. Evidently

$$\tan\alpha = \frac{F_1}{F_2} \qquad \text{(VII.1)}$$

It is interesting to compare the formants of vowels with the same argument, as is done in figure VII.2. Because

$$\tan\alpha = \frac{F_1}{F_2} = \frac{F_1^*}{F_2^*} \qquad \text{(VII.2)}$$

the formants of [i*] can be deduced from those of [i] by multiplying them by the same factor. This means that the vocal tract which produces [i] has been subjected to proportional axial growth.

IN PROPORTIONAL AXIAL GROWTH THE FORMANTS SLIDE ALONG LINES THROUGH THE ORIGIN, IN OTHER WORDS ALONG THE ARROWS OF THE VOWEL VECTORS.

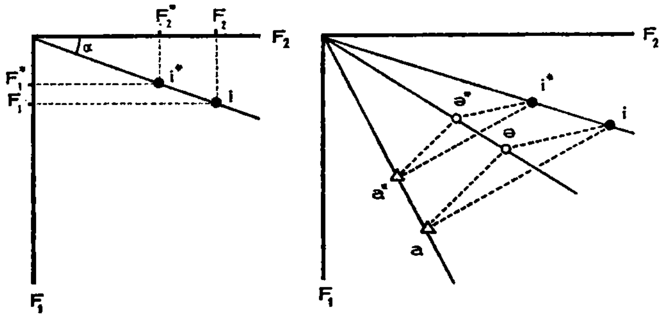


Figure VII.2

The formant shifts in the vowel diagram due to proportional axial growth.

When, see fig. VII.2, [a], [ə] and [i] represent three vowels produced by different configurations of the same vocal tract, proportional axial growth of that tract will yield the formant positions [a*], [ə*] and [i*]. Because all formants are divided by the same factor the lines ə*-i* and ə-i will be parallel. The same can be said of the lines a*-ə* and a-ə, and of the pair a*-i* and a-i.

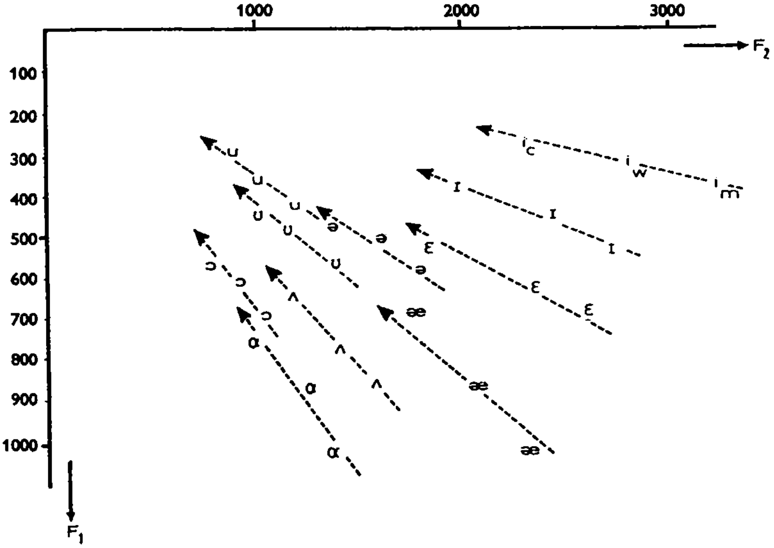


Figure VII.3

The formants of men, women and children. The averages have been taken from Table II given by Peterson and Barney in their paper 'Control methods used in a study of the vowels', *JASA* Volume 24, Number 2 (March 1952). The formants frequencies F_1 and F_2 have been plotted on linear scales. The children have the highest formants, the men the lowest. The women are in between.

We discovered an interesting application of figure VII.2 by studying and interpreting the results of the well-known measurements of Peterson and Barney[1]. In a separate table they present, among other things, the average values of the first two formants of a large group of talkers pronouncing the same vowels in isolated words. They also give separate averages for children, women and men. By plotting these averages on LINEAR frequency scales we could

[1] *l.c.*

produce figure VII.3. There is a strong tendency in phonetics, especially in the United States, to use non-linear frequency scales for plotting formant positions in vowel diagrams. For instance, one uses a linear scale for F_1 and the so-called mel-scale for F_2. We shall not, in this monograph, discuss the possible worthwhileness of logarithmic scales or the mel-scale for the interpretation of formant frequencies by the nervous system. Graphical interpretation of formulas is greatly furthered by sticking to linear scales as long as possible. This principle is illustrated by figure VII.3 where the corresponding vowels of children, women and men show the striking tendency to lie on straight lines through the origin. This proves that PROPORTIONAL AXIAL GROWTH IS AT WORK HERE. For the sake of completeness we here repeat that proportional transversal growth does not influence the formant positions. As to the accuracy with which the points fall on the lines, we must bear in mind that the formants have been measured by means of the spectrograph. A spectrograph can only measure in steps equal to the fundamental frequency of the vocal cords. Consequently, the formants have been measured by Peterson and Barney in steps of, on the average, 264 Hz for children, 223 Hz for women and 132 Hz for men.

For the sake of clarity we have selected the vowels [u], [i] and [a], representing the corner-points of Hellwag's classical vowel-triangle (1781), and shown them separately in figure VII.4. As we expected, the three triangles have the same shape. This fact throws an interesting light on several problems keeping linguists and phoneticians busy.

While learning to speak the child does not (and indeed cannot) reproduce the absolute values of the vowel-formants of the adult speakers in its environment. In the long run the child learns to master the muscular activity necessary for producing the same number of perceptively different vowel sounds used for coding purposes by adult speakers of the same language in situations where the need for utmost clarity is greatest. Gradually the child, in his own language, learns to choose the normalized number of perceptually different tongue positions realizable within the limiting boundaries of the lips, the teeth, the back and the bottom of the mouth *etc.* He uses and needs his ear in this process but certainly does not match the absolute values of his formants to an example.

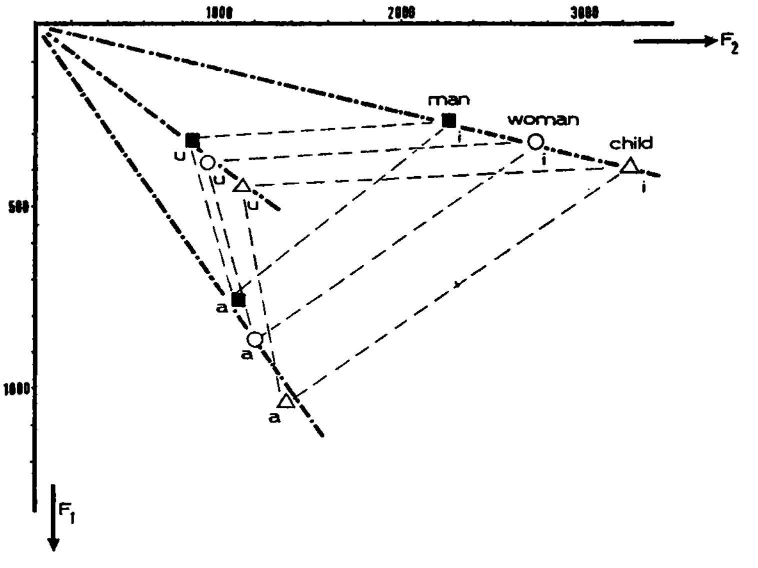


Figure VII.4

The evidence of proportional axial growth in the vowel triangles of children, women and men.

Once settled in a certain period of life the neurological programs for controlling the articulatory muscles do not change materially as the child grows up. Apparently the tongue maintains the same relative position in the vocal tract for each vowel. Lengthening of the vocal tract by growth makes the formants shift downwards and there are no compensatory attemps on the part of the talker to make them stay in their 'youthful' positions. Obviously the often postulated auditory feed-back mechanism for regulating the articulatory movements is not at work here.

Every talker has his own vowel system. The range of his formant-frequencies is limited by the dimensions of his vocal tract. The listener has to cope with the difficulty that the absolute values of the formants of the vowels differ from talker to talker. He is helped, among other things, by the fact that in a given language all talkers have the same number of vowels. A listener is able to adapt himself in a surprisingly short time to the formant positions of a talker.[2] When he is not able to do so, for instance when isolated words of unknown talkers are presented to him, his identifications of the intended vowels become rather faulty.[3]

When we admit that the exact acoustic imitation of the vowels is not the aim in learning to speak, we can understand why the inherited shapes of parts of the vocal tract, like the hard palate, can withstand supposed 'compensatory' actions.

The vowel-diagram of an individual talker need not have exactly the same shape as the average diagrams presented in fig. VII.3 and fig. VII.4, but the sequence of the constituent vowels with respect to F_1 and F_2 will be the same which is sufficient for the listener who knows the language in question.

In phonemics one tries to find 'invariants', unambiguous descriptions on some level of the phonemes. It has become clear by now that one should not look for these invariants on the acoustic level. The place where to find them is the articulatory level.

From the figures VII.3 and VII.4 it is apparent that the average child, the average woman and the average man give essentially the same commands to their articulators when producing a certain vowel, in spite of the fact that this sameness gets lost on the level of the absolute values of the formants. These commands, given subconsciously by the talker, may be of the following nature:

form a very narrow tube between the tongue and the palate, beginning just behind the upper teeth and ending near the centre of the vocal tract, at the same time creating a wide back cavity. The reader will have noticed that this is the command for [i]. In terms of the twin-tube parameters: adjust k, β and l to the appropriate values for [i].

[2] Peter Ladefoged and D. E. Broadbent, 'Information conveyed by vowels', *JASA*, Vol. 29, No. 1 (January, 1957), pp. 98-104.
[3] Peterson and Barney, *l.c.*

The parameters k and β are true 'shape' parameters. The parameter l is of a hybrid nature. It is consequent to split l up in the following manner:

$$l = \gamma\, l_o \qquad \text{(VII.3)}$$

where γ is a true (dimensionless) shape parameter and l_o represents the average length of the vocal tract of the talker whose vowels we are investigating.

From fig. VII.3 we derive

$$\underset{\text{average child}}{l_o} : \underset{\text{average woman}}{l_o} : \underset{\text{average man}}{l_o} = 1 : 1.16 : 1.35$$

So, the parameters k, β and γ merely determine the shape of a vocal tract, leaving it to the parameter l_o to decide at what absolute frequencies the formants will be located.

It is possible now to give a modern content to the, up to now, rather loosely defined square 'phonetic' brackets [ ]. For instance, for the vowel [i] we may define:

$$[\mathrm{i}] = [\mathrm{k}_i, \beta_i, \gamma_i] \qquad \text{(VII.4)}$$

Likewise we have

$$[\mathrm{a}] = [\mathrm{k}_a, \beta_a, \gamma_a] \qquad \text{(VII.5)}$$

and

$$[\mathrm{u}] = [\mathrm{k}_u, \beta_u, \gamma_u] \qquad \text{(VII.6)}$$

etc., *etc.* Between the square brackets are the true invariants.

We shall now, as illustrated in figure VII.5, represent the first two formants of the twin-tube in a vector diagram. For that purpose we make use of the examples given in the figures VI.3 and VI.4 of the foregoing chapter. In the nomogram we introduce l_o instead of l so that the role of the parameter γ comes to the fore.

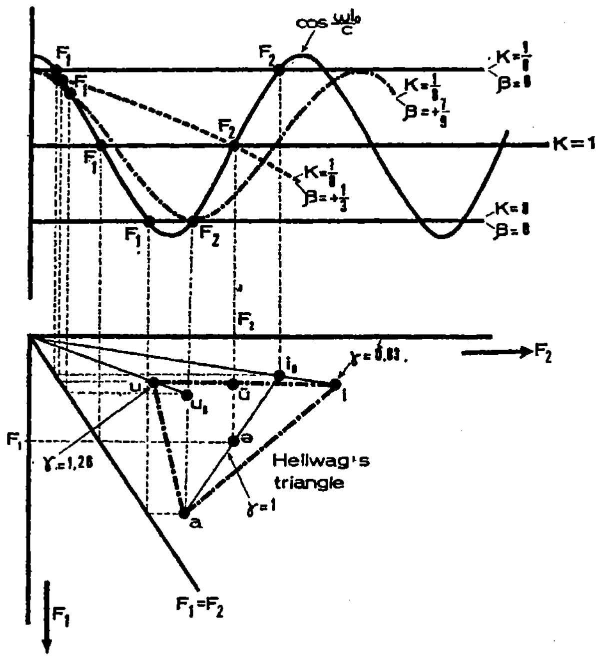


Figure VII.5

The vector representation of the formants of the twin-tube resonator leading to the vowel triangle postulated by Hellwag in his 1781 thesis *De formatione loquelae.*

For the benefit of the reader the values of the parameters k, β and γ of our examples have been collected in the following table.

TABLE VII.1

vowel	k	β	γ
[ə]	1	?	1
[a]	8	0	1
[u]	$\frac{1}{8}$	$+\frac{7}{9}$	1.26
[i]	$\frac{1}{8}$	0	0.83
[ü]	$\frac{1}{8}$	$+\frac{1}{3}$	1

The vector construction can best be illustrated following the simple case of [ə] with $k = 1$ and $\gamma = 1$. In that case the twin-tube has the length l_o which serves as a description of the anatomic limitations of the vocal tract in question.

The formants F_1 and F_2 come to the fore as the first two zero-crossings of $\cos \frac{\omega l_o}{c}$. The values of F_1 and F_2 must be plotted on the corresponding axes for F_1 and F_2 depicted below the nomogram. The plotting of F_2 comes down to simply projecting the second zero-crossing on the horizontal F_2-axis. The value of F_1 cannot be directly projected on the vertical F_1-axis. This is easily effected, however, by projecting F_1 on the line $F_1 = F_2$ drawn through the origin, first. In that way it is even possible to have different scales for F_1 and F_2, which is useful for giving the intended vowel diagram a suitable shape. The resulting point in the $F_1 F_2$ plane can now be called [ə].

For [a] we suppose that the vocal tract also has the length l_o, so that we may take $\gamma = 1$. Furthermore we have $k = 8$.

For [ü] we again put $\gamma = 1$, but now $k = \frac{1}{8}$ and $\beta = +\frac{1}{3}$. In this case F_2 coincides with the second zero-crossing of $\cos \frac{\omega l_o}{c}$. The first formant assumes a low value.

The construction of [i] is slightly more complicated. In that case the length of the vocal tract is less than l_o; as radiograms show, we have, for instance, $l = 0.83\, l_o$ or $\gamma = 0.83$. We first construct the vector [i_o], corresponding with $\gamma = 1$, $\beta = 0$ and $k = \frac{1}{8}$. We then subject the twin-tube to proportional axial shrinking by multiplying all axial dimensions by the factor 0.83. Consequently ALL formants are divided by the factor 0.83; in other words [i_o] slides along a line through the origin until its arrow is lengthened by a factor $\frac{1}{0.83}$. We may then call the new vector [i].

For the construction of [u] a similar construction has to be applied. In [u] the length of the vocal tract is greater than l_o. For instance, $\gamma = 1.26$. We now have to shorten the arrow of [u_o] by dividing it by 1.26. We call the resultant vector [u].

dignoſcamus, ad quam partem nixus factus fuerit. Articulatio copulat literas cum literis, vocales & conſonantes cum vocalibus & conſonantibus, ſtruit diphthongos, & complicationes conſonantium, & totas ſyllabas perficit: in ſyllabis autem terminatur, dum quævis nova ſyllaba filum ejus a præcedente inceptum deſerat, novumque incipiat. Unica ſyllaba unicam ſaltem admittit vocalem vel diphthongum, ſæpe autem integra utrâve abſolvitur nuda.

§ 56.

Ut diſcerem, quid ſingularum literarum naturam definiat, varia tentavi, comparavi literas cum literis multiplici ratione, earum inquirens ſimilitudines & differentias, attente obſervavi, quænam organa, & qua eadem lege agant in quavis data litera; quid in cujusvis literæ productione ſit neceſſarium, conſtans, eſſentiale, quid indifferens, mutabile, accidentale; examinavi ſingulas & combinatas, & comparatas ad auditum, ad viſum, ad tactum, & ad ſenſum, qui ipſis loquelæ inſtrumentis ob nerves ineſt. Inſuper aliquorum circa idem objectum cogitata & obſervata, quoad eorum copiam nanciſci potui, non præterivi, & omnino quædam eorum meis, ſed tamen quoque meorum nonnulla illis recentiora, veriora, certiora mihi viſa ſunt, multa vero meorum ex illis confirmata lætus deprehendi. Quæ demum hiſce ſtudiis obtinuerim, fideli nunc commentario exponam, primo de vocalibus, dein de conſonantibus diſſerens.

§ 57.

Princeps vocalium, reliquarum baſis, vel in ſcala poſitarum centrum eſt *a*: ex hac duplex aſcendit ſcala, in gradus extremos *i* & *u* terminata: gradibus his extremis & homologis inferioribus termini interjacent intermedii. Graduum & terminorum intermediorum ad baſin relatio ſub hoc ſchemate concinno poteſt repræſentari:

u ü i
o ö e
å ä
a

D Vo-

Figure VII.6

Photographic reproduction of page 25 of Hellwag's thesis on which appears his classical vowel triangle.

The points *u*, *ü*, *i* and *a* in figure VII.5 delimit a triangle, standing on its top *a* and having *u*, *ü*, *i* for basis. The internal point *ə* has the character of a sort of centre of gravity.

Hellwag's vowel triangle (1781), depicted in figure VII.6, has the same top and the same basis. The other vowels of Hellwag can be constructed too in figure VII.5 by choosing suitable values for k, β and γ. The neutral vowel [ə] is not mentioned by Hellwag, but we shall not tackle that problem here.

It is highly rewarding that the twin-tube parameters lead to the same systematic arrangement of the vowels as Hellwag's triangle. This certainly makes the twin-tube an attractive model for linguistic use.

In Chapter X we shall show how to estimate the twin-tube parameters by studying actual radiograms and other data.

VIII

STATIONARY WAVE-PATTERNS IN THE TWIN-TUBE RESONATOR

When, in a tube with constant cross-area, two equally strong sine waves having the same frequency travel in opposite directions, a stationary wave pattern will ensue:

$$A\varepsilon^{j\omega(t+\frac{x}{c})} + A\varepsilon^{j\omega(t-\frac{x}{c})} \qquad \text{(VIII.1)}$$

or

$$2A\cos\frac{\omega x}{c}\varepsilon^{j\omega t} \qquad \text{(VIII.2)}$$

This is the complex notation for

$$2A\cos\frac{\omega x}{c}\cos\omega t \qquad \text{(VIII.3)}$$

This formula tells us that there is a sinusoidal vibration with the frequency ω in every point of the tube, but that its amplitude is a function of x.

At the points where $\cos\frac{\omega x}{c} = 0$, there is no vibration at all. These points are called the nodes (see figure VIII.1).

At the points where $\cos\frac{\omega x}{c} = \pm 1$ the amplitude shows a maximum. These points are called the loops or anti-nodes.

The stationary wave-pattern of figure VIII.1 is directly applicable to the case of the single tube, [ə], which in Chapter VI came to the fore as a degeneration of the twin-tube resonator ($k = 1$). According to (VI.16) in this case there are the following formants

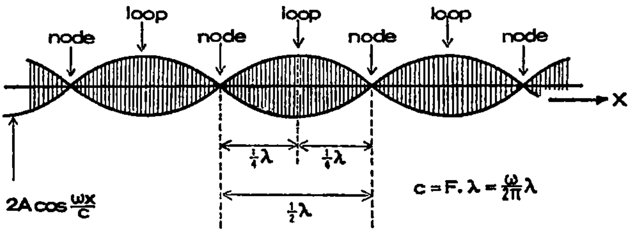


Figure VIII.1

Amplitude envelope of stationary wave. The distance between two adjacent nodes (or loops) is defined as half a wave-length. The wave-length is indicated by the symbol λ.

$$F_1 = \tfrac{1}{4}\frac{c}{l} \text{ and } F_2 = \tfrac{3}{4}\frac{c}{l} \qquad \text{(VI.16)}$$

Because

$$c = F\,.\,\lambda = \frac{\omega}{2\pi}\lambda \qquad \text{(VIII.4)}$$

we have

$$l = \tfrac{1}{4}\lambda_1 \quad \text{and} \quad l = \tfrac{3}{4}\lambda_2 \qquad \text{(VIII.5)}$$

Evidently the single tube 'contains' parts of the stationary wave-pattern of figure VIII.1.

In the upper part of figure VIII.2 the case of [ə] has been depicted.

For the single tube with its constant cross-area we need not distinguish between the volume velocity and the particle velocity.

The velocity always has a node at the throat and a loop in the mouth opening.

The sound pressure always has a loop at the throat and a node in the mouth opening.[1]

In the case of the second formant there is an additional node in the tube, located between the throat and the mouth-opening.

The stationary wave-pattern in the single tube permits us to represent the acoustic phenomena in the tube as follows. The sound waves travel to and fro between the throat and the mouth opening at the speed c. There is a 'hard' reflection at the throat and a 'soft'

[1] See Chapter III. The pressure node is there by definition.

reflection in the mouth opening. As a matter of fact the two waves travelling in opposite directions are each other's reflections.

In a hard reflection a wave returns immediately and with the same amplitude. In a soft reflection a wave also returns with the same amplitude but after a delay of $\frac{1}{2}T$ seconds, given by

$$\tfrac{1}{2}T = \frac{1}{2F} = \frac{\pi}{\omega} \tag{VIII.6}$$

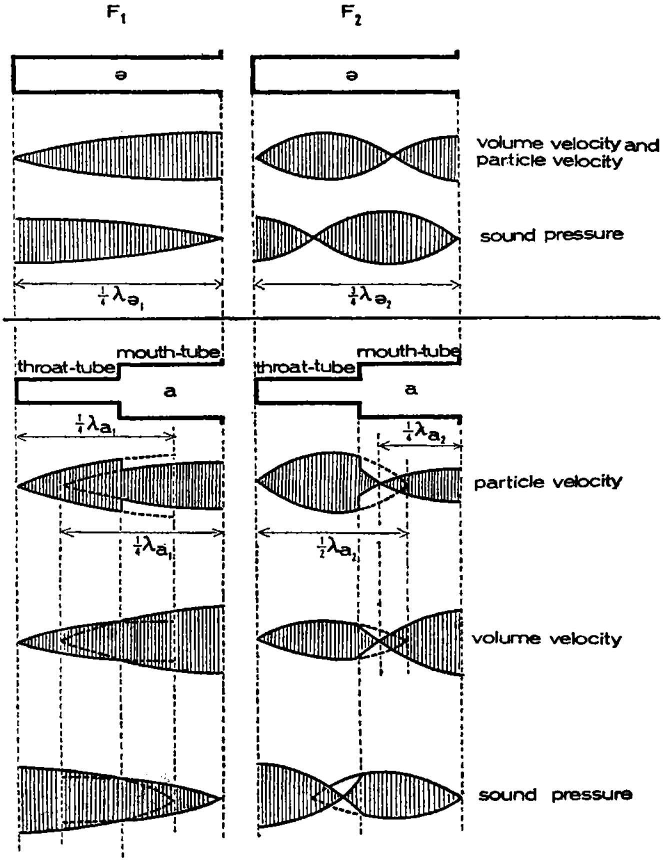


Figure VIII.2
Fractured stationary wave-patterns in all cases but [ə].

It takes a wave $\frac{l}{c} = \frac{1}{4F} = \frac{T}{4}$ seconds (for the first formant) to travel the length of the tube. During one complete vibration, in other words during the interval T, the wave travels twice the length of the tube and only once suffers a delay of $\frac{1}{2}T$ at the soft reflection in the mouth opening. Accordingly, $2 \times \frac{T}{4} + \frac{1}{2}T = T$. The hard reflection does not produce a delay.

In figure VIII.3 a model of the soft reflection in the mouth-opening is given. It is composed of two equally heavy billiard balls suspended on strings. The black ball is lifted and then set free at the time t_1. At the time t_2 it has gained speed; at the moment t_3 it collides with the white ball, transferring almost at once its full kinetic energy to the white ball. The black ball remains at rest now whereas the white ball proceeds on its own. This process is referred to by billiard players as 'stun and stab'. The white ball continues its circular path until its kinetic energy has been transferred completely into potential energy at the time t_5. It returns, gains speed at t_6 and at the time t_7 collides with the black ball. Now the black ball takes over the complete kinetic energy of the white one. The net result of this interesting procedure is that the black ball has been reflected with a time delay equal to one half of the period of the pendulum.

The hard reflection of the sound waves at the throat can also be illustrated by a pendulum-model, see figure VIII.4, in which a ball hits a solid wall and is immediately reflected.

Things become slightly more complicated when the single tube is pinched so that it changes into a twin-tube. For the CALCULATION of the formants we still have the twin-tube formula, either in its traditional form:

$$\tan\frac{\omega l_1}{c} \cdot \tan\frac{\omega l_2}{c} = \frac{S_1}{S_2} = k \qquad \text{(VIII.7)}$$

or in its eccentricity transformation:

$$\cos\frac{\omega l}{c} = \frac{1-k}{1+k}\cos\beta\frac{\omega l}{c} \qquad \text{(VIII.8)}$$

For the sake of completeness we add, that this formula can also be written as:

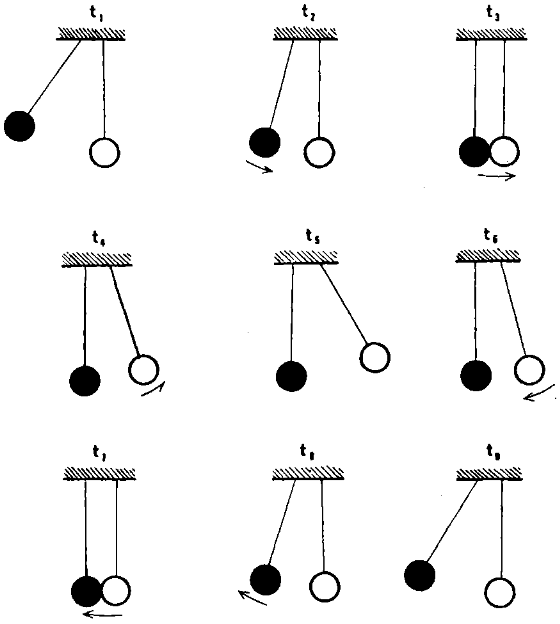


Figure VIII.3

Pendulum-model of the soft, delayed reflection of the sound waves in the mouth-opening.

$$\cos \quad \omega \frac{l_1 + l_2}{c} = \frac{1-k}{1+k} \cos \quad \omega \frac{l_1 - l_2}{c} \qquad \text{(VIII.9)}$$

If we should want to describe or see the formant positions as WAVE LENGTHS, we may write the formant formula as follows:

$$\cos 2\pi \frac{l}{\lambda} = \frac{1-k}{1+k} \cos \beta \, 2\pi \frac{l}{\lambda} \qquad \text{(VIII.10)}$$

$$\cos 2\pi \frac{l_1 + l_2}{\lambda} = \frac{1-k}{1+k} \cos 2\pi \frac{l_1 - l_2}{\lambda} \qquad \text{(VIII.11)}$$

Though, as we have already shown, there is an elegant way of graphically solving the formant formulas, we may ask ourselves whether it is still worthwhile to think in the simple terms of stationary waves in order to achieve a better physical insight into the birth of the formants. The waves keep running to and fro in the tube but they are reflected at three places instead of only at two, namely, at the throat, in the mouth-opening and at the JOINT between the two tubes.

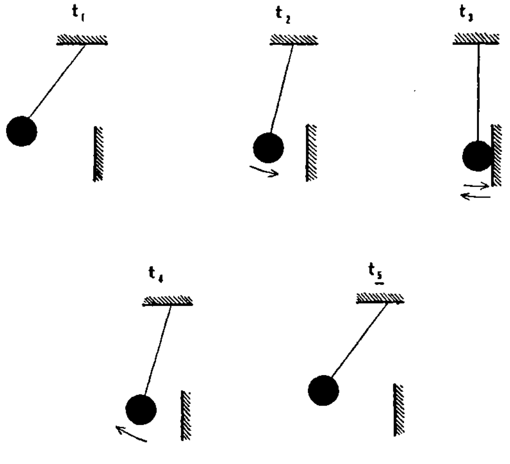


Figure VIII.4

Pendulum-model of the hard, non-delayed reflection of the sound waves at the throat.

Once more the net result, however, is a stationary wave pattern in each of the two tubes. These patterns, of course, have the same frequency but different amplitudes now. No extra nodes or loops are introduced by pinching the [ə] tube, but the distance between a node and a loop is not necessarily equal to $\frac{1}{4}\lambda$ now. We shall illustrate this behaviour for a case with $\beta = 0$ and $k > 1$, the [a] like vowel represented in the bottom part of figure VIII.2.

As explained in Chapter VI, at the joint in the twin-tube where the cross-area makes a step, the volume velocity remains continuous whereas the particle velocity makes a step in the opposite direction:

when the cross-area steps down the particle velocity goes up. This is the typical behaviour of a rapid in a stream.

Though the volume velocity remains continuous its derivative with respect to x makes a step. The same can be said of the sound pressure.

Let us discuss F_1 first. Though the quarter wave-length, $\frac{1}{4}\lambda$, does no longer 'fit' into the twin-tube, because the tube is too long for that, it still plays an interesting part. The throat-tube contains part of an, in this case, undersized $\frac{1}{4}\lambda$ pattern that cannot reach the mouth-opening. On the other hand the mouth-tube contains part of a likewise undersized $\frac{1}{4}\lambda$ pattern that cannot reach the throat. These united patterns, however, 'fill' the whole twin-tube because they have DIFFERENT amplitudes and are linked together at the joint by the condition of continuity. In that way the wave-pattern, as it were, skips part of the too long twin-tube.

In the case of F_2 the twin-tube, being too short, cannot contain three quarter wave-lengths. The throat-tube 'houses' part of an oversized $\frac{3}{4}\lambda$ pattern that tries to protude from the mouth. On the other hand the mouth-tube holds part of a likewise oversized $\frac{3}{4}\lambda$ pattern that tries to pierce the throat. Again the united patterns fill the whole twin-tube.

The concept of the stationary waves is, as it were, saved by the inequality of the amplitudes of the wave patterns in the two tubes. It is possible to calculate the ratio of these amplitudes for the general case of the twin-tube. As the twin-tube model is meant to be a formant generator and not a good imitation of the physical processes going on in the real vocal tract, we shall not overload this monograph with these calculations, the linguistic importance of which must certainly be doubted.

In Chapter VI, see figure VI.4, we drew attention to the fact, that for $\beta = \frac{1}{3}$ (and also for $\beta = -\frac{1}{3}$) the position of F_2 was independent of the value of the constriction factor k. So, starting from [ə], it is not possible to alter F_2 by pinching the tube at the position $l_1 = \frac{1}{3}l$ or at $l_1 = \frac{2}{3}l$. As can be seen from figure VIII.2 we cannot alter F_2 when the place of articulation is in a loop or a node.

As figure VIII.5 clearly shows, in the twin-tube model the Helmholtz resonances come to the fore as the FIRST formants of the vowels with a low value of k. As a practical example the value $k = \frac{1}{8}$ has been demonstrated.

We want to show first of all, that the formant formula of the twin-tube model is in agreement with the classical formula of the frequency of the Helmholtz resonator. As we have neglected the end correction in the twin-tube, we must likewise neglect this correction in the Helmholtz formula[2], which then may be written as follows

$$F = \frac{c}{2\pi}\sqrt{\frac{S_1}{l_1 V}} \tag{VIII.12}$$

with: l_1, the length of the neck
S_1, the cross-area of the neck
V, the volume of the cavity.

When, as is the case for the twin-tube, the cavity has a constant cross-area, we may write

$$V = l_2 S_2 \tag{VIII.13}$$

Furthermore, it can be directly derived from (VI.5) and (VI.6) that the product

$$l_1 l_2 = \tfrac{1}{4} l^2 (1-\beta^2) \tag{VIII.14}$$

Combination of (VIII.12), (VIII.13) and (VIII.14) yields:

$$F = \frac{c}{\pi} \cdot \frac{2}{l}\sqrt{\frac{k}{1-\beta^2}} \tag{VIII.15}$$

When we develop the formant formula (VI.9) into a series, which makes sense for low values of ω, and only take into consideration the first two terms of the series for the cosines, we get

$$1 - \tfrac{1}{2}\left(\frac{\omega l}{c}\right)^2 = \frac{1-k}{1+k}\left[1 - \tfrac{1}{2}\left(\frac{\beta\omega l}{c}\right)^2\right] \tag{VIII.16}$$

[2] See for instance: J. W. S. Rayleigh, *The Theory of Sound*, II, p. 491 and E. Skudrzyk, *Die Grundlagen der Akustik*, p. 350, *etc.*

This can be written as

$$F_1 = \frac{c}{\pi} \cdot \frac{2}{l} \sqrt{\frac{k}{1 + k - \beta^2(1 - k)}} \qquad \text{(VIII.17)}$$

By comparing (VIII.15) and (VIII.17) we see that these formulas are in fair agreement provided $k \ll 1$, which indeed is the case for the vowels in question.

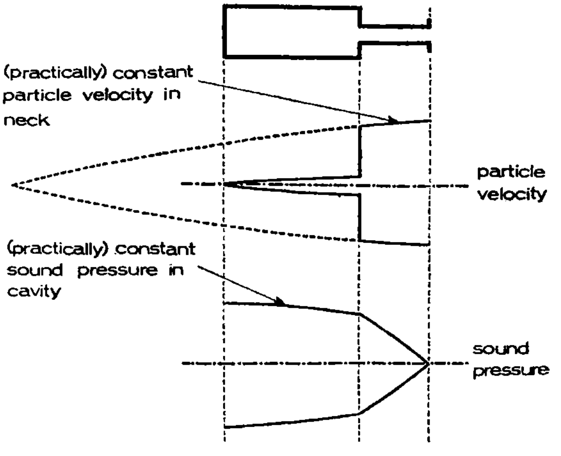


Figure VIII.5

Fractured stationary wave patterns in the Helmholtz-variety of the twin-tube model.

The classical formula's for the frequency of the Helmholtz resonance have been derived assuming that the air in the neck of the resonator moves to and fro as a lumped mass. The wave pattern in figure VIII.5 shows that the particle velocity is practically constant indeed in the mouth tube which forms the neck of the resonator.

Another assumption in the classical approach is that the air in the cavity behaves like an elastic cushion, as a stiffness without mass effects. The pattern shows the particle velocity is very low indeed in the cavity as a result of the step at the point of constriction. This is in harmony with the image of a rapid we have formed of the joint.

Finally, it is supposed that the pressure is constant throughout the cavity. This assumption is supported by the shape of the pressure pattern which comes to the fore in figure VIII.5.

An interesting case is represented by $k \to 0$ or $k \to \infty$, two conditions we shall describe as saturation. For $k \to 0, \frac{1-k}{1+k} \to 1$ whereas for $k \to \infty, \frac{1-k}{1+k} \to -1$.

In the case of close saturation ($k \to 0$) the formant formula in its (VIII.9) version simply boils down to

$$\cos\frac{\omega l}{c} = \cos\frac{\omega(l_1 - l_2)}{c} \qquad \text{(VIII.18)}$$

By eliminating l_2 we arrive at:

$$\cos\frac{\omega l}{c} = \cos\frac{\omega l}{c}\cos\frac{2\omega l_1}{c} + \sin\frac{\omega l}{c}\sin\frac{2\omega l_1}{c} \qquad \text{(VIII.19)}$$

By eliminating l_1 we get:

$$\cos\frac{\omega l}{c} = \cos\frac{\omega l}{c}\cos\frac{2\omega l_2}{c} + \sin\frac{\omega l}{c}\sin\frac{2\omega l_2}{c} \qquad \text{(VIII.20)}$$

Expression (VIII.19) is true provided

$$\cos\frac{2\omega l_1}{c} = 1 \text{ and } \sin\frac{2\omega l_1}{c} = 0 \qquad \text{(VIII.21)}$$

which leads to the condition that

$$\frac{2\omega l_1}{c} = n_1 2\pi \qquad \text{(VIII.22)}$$

where n_1 is an integer ($n_1 = 0, 1, 2, 3,$ *etc.*).

As

$$c = \frac{\omega}{2\pi}\lambda \qquad \text{(VIII.23)}$$

we finally have

$$l_1 = n_1 \frac{\lambda}{2} \tag{VIII.24}$$

Likewise, by solving (VIII.20) we get

$$l_2 = n_2 \frac{\lambda}{2} \tag{VIII.25}$$

where n_2 is an integer too.

Summarizing, in the case of close saturation the formants come to the fore as the $\frac{1}{2}\lambda$, $2.\frac{1}{2}\lambda$, $3.\frac{1}{2}\lambda$, *etc.* resonances of BOTH tubes.

In the case of open saturation ($k \to \infty$) the formant formula reduces to

$$\cos\frac{\omega l}{c} = -\cos\frac{\omega(l_1 - l_2)}{c} \tag{VIII.26}$$

which ultimately leads to

$$l_1 = m_1 \frac{\lambda}{4} \tag{VIII.27}$$

and

$$l_2 = m_2 \frac{\lambda}{4} \tag{VIII.28}$$

where m_1 and m_2 are odd integers (= 1, 3, 5, *etc.*).

Summarizing, in the case of open saturation the formants come to the fore as the $\frac{1}{4}\lambda$, $3.\frac{1}{4}\lambda$, $5.\frac{1}{4}\lambda$, *etc.* resonances of both tubes.

In practice the vocal tract is never completely saturated because in vowels the airstream may never entirely be blocked. Consequently, in vocal tracts approaching close saturation F_1 will not be equal to zero but will reach its lowest practical value instead. The second formant, F_2, will come close to the $\frac{1}{2}\lambda$ resonance of the longest tube. Normally this will be the $\frac{1}{2}\lambda$ resonance of the back cavity but this is not necessary. In vocal tracts approaching open saturation F_1 will be close to the $\frac{1}{4}$ wave length resonance of the longest tube. The second formant will be supplied by the tube having a length suitable for producing the next higher frequency.

IX

THE VOWEL TRIANGLE AS A LIMITING CONTOUR OF THE FORMANTS OF THE VOCAL TRACT

In Chapter VII we have seen that the vector diagram formed by [u], [i], [ɑ] (or [a]) and [ə] of a talker has the shape of a triangle with [ə] in the middle. When we introduce in this diagram, as done in figure IX.1, the other vowels, we notice that they all remain within a contour, which has the shape of a crude triangle also. The figure represents the 10 vowels of the average male speaker of the Peterson and Barney experiment, with [a] and [ü] thrown in as additional fillers. We shall show now how the twin-tube method is able to predict the necessary presence of the three sides of the triangular contour.

The position of the upper side is fixed by the fact that the lowest practicable values of k (around $k = \frac{1}{8}$) produce Helmholtz resonances at practically the same low value of F_1; the point of intersection between $\cos\frac{\omega l}{c}$ and $\frac{1-k}{1+k}\cos\beta\frac{\omega l}{c}$ is confined to the limited low frequency region, see figure VII.5.

The left-hand side is a self-evident boundary. It represents the line $F_1 = F_2$. As we call the lowest formant F_1 and the highest formant F_2, the region between the line $F_1 = F_2$ and the F_1-axis is a forbidden zone because in that domain $F_2 < F_1$. The extent to which the formant positions do approach the boundary is determined by the maximal length a talker can give to his vocal tract by lowering his larynx, protruding his lips and moving backwards the hump of his tongue. In these actions he is limited by anatomical restrictions. Also the highest feasible value of k (around $k = 8$)

plays a role. For instance, the line $F_1 = F_2$ is reached for $\beta = 0$, $k = \infty$.

The right-hand side of the triangle plays an interesting part too. As figure VII.5 clearly shows, for a given length of the twin-tube F_2 reaches a maximum for $\beta = 0$, that is for a twin-tube with the joint exactly in the middle. The only means of further increasing F_2

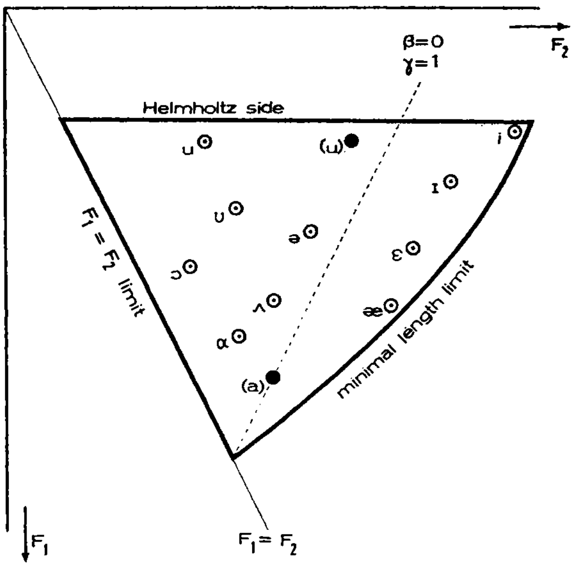


Figure IX.1

The triangular contour of the vowel inventory of a talker as predicted by the twin-tube model.

consists in making the vocal tract shorter. As already mentioned shortening of the vocal tract is effected by eliminating the influence of the lips, by keeping the hump of the tongue low and by pulling up the larynx. As these activities are limited by the anatomical restrictions of the talker the formant positions cannot fall too far to the right side of the dotted line in figure IX.1. In that way the limiting contour retains a triangular shape.

HELLWAG, in 1781, did not, and indeed could not base his famous triangle on acoustical calculations. It looks more as if he were inspired by the anatomical restrictions of the human vocal apparatus which, as we have seen, have their important say in the theory of the twin-tube model.

The success of the twin-tube method in predicting the triangular shape of man's vowel inventory certainly proves that this method is of practical value for linguistics.

In the vector representation of the vowel formants we have, in this monograph, confined ourselves to depicting F_1 and F_2 only. We wish to state explicitly that this restriction has nothing to do with the well-known hypothesis that vowels can be presented adequately (whatever this may mean) by two-formant approximations. The twin-tube model is NOT a two-formant generator; it generates all formants including the much disputed F_3 though accuracy cannot be expected. The twin-tube model is not connected with problems of perception or interpretation by the nervous system. The task of the model is to generate formants at the frequency locations to be expected on the basis of visible speech patterns of vowels[1] produced by real vocal tracts. For instance, it predicts the twin-formants at the appropriate positions: the well-known $F_2 F_3$ twin for [i]-like vowels and the $F_1 F_2$ twin of the [a]-type, as is very clearly shown in figure VI.3.

It remains an interesting fact, however, that Hellwag's vowel triangle can be correlated with the $F_1 F_2$ vector diagram, independent of the possible influence of F_3 on things.

The role of F_3 in perception, or rather in phonemic interpretation, is still under discussion. A promising, though not exactly problemless, approach seems to be the study of two-formant synthetic vowels.

As early as 1952 a group of investigators at the Haskins-laboratories[2] published some interesting results of their experiments

[1] R. K. Potter, G. A. Kopp and H. C. Green, *Visible Speech* (D. van Nostrand, Inc. 1947).

[2] P. Delattre, A. M. Liberman, F. S. Cooper and L. J. Gerstman, "An experimental study of the acoustic determinants of vowel color: observations on one- and two-formant vowels synthesized from spectrographic patterns", *Word*, Vol. 8, no 3 (1952).

with two-formant synthetic vowels. Using the well-known, if not famous, pattern playback, they studied one- and two-formant synthesized vowels and found that, as a rule, F_2 of a synthetic vowel was higher than F_2 of a vowel produced by a real vocal tract. They suggested, though not in the same words, that the F_2 of a synthesized two-formant vowel had to play the role of a combined F_2F_3 of a 'natural' vowel. Though we do not pretend to contradict that statement we should like to add a possible additional reason for arriving at higher second formants in synthesized vowels.

During experiments on artificial two-formant vowels in the Institute of Phonetic Sciences of the University of Amsterdam, the following experience was gained. When subjects were invited to produce the vowel system of their native language by adjusting the formant controls of a two-formant generator[3] they tried to use to advantage the entire F_2 range of the apparatus. They created a system of optimally contrasting formant positions, they could not even realize with their own vocal tract: for instance, F_2 of [i] corresponded with the highest F_2 obtainable with the apparatus whereas F_2 of [u] was the lowest F_2 of the machine. The subjects did not disclose a tendency to copy the absolute frequency positions of their own vowels. The obvious conclusion is: the subjects follow their ingrained urge to fully exploit the frequency range of the formant generating system they control, be that system their own vocal tract or some sort of formant synthesizer.

In our opinion the further study of this phonemic urge is more important to LINGUISTICS than experiments on the matching of, for instance, 'vowel colour'.

[3] J. G. Blom and J. Z. Uys, "Some Notes on the Existence of a 'Universal Concept' of Vowels", *Phonetica* (1966), 15, pp. 65-85.

X

APPLICATION OF THE TWIN-TUBE METHOD TO THE VOWEL SYSTEM OF DUTCH

The opening of this chapter is a summary, meant as a caution for the benefit of the reader. The method of the loss-free twin-tube resonator as we use it in this monograph must not be seen as a precision method of predicting the absolute values of the formant positions of a particular vocal tract. In the process of speaking and hearing the role of the absolute values of the formant frequencies must be a minor one for several reasons. The anatomical differences between talkers intending the same vowel, as well as the talkers' individual ways of controlling their articulatory muscles cause their formant frequencies to spread largely. On top of that

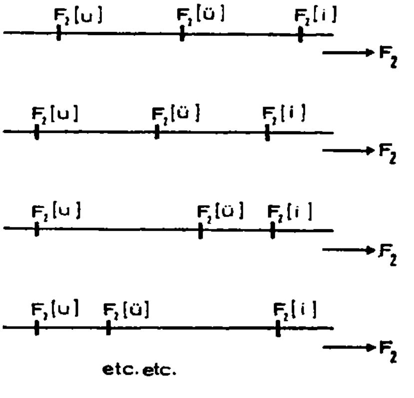


Figure X.1

Showing how the phonemic contrast F_2[u] < F_2[ü] < F_2[i] is independent of the absolute values of the formant positions.

there is a considerable overlap of the formants of adjacent vowels. On the perception side, the Peterson and Barney experiment[1] has shown that a listener is not able by far to correctly identify, that is, in agreement with the talker's intention, the vowels in a series of isolated words, randomly composed of words uttered by a large number of DIFFERENT talkers. The successful identification by a listener of isolated words uttered by ONE AND THE SAME talker is no doubt due to the listener's interpretation of the talker's complete vowel system (which he can get to know in an often surprisingly short time) and NOT to his classification of the separate vowels solely on the basis of their own absolute formant frequencies. The success of the mechanism of speech and hearing must be attributed to the interpretation by the listener of vowel formant SEQUENCES instead of absolute vowel formant positions. An illustration of this principle is given in figure X.1, showing how the contrast between the second formants of, for instance, the vowels [u], [ü] and [i] in Dutch is 'carried' by the sequence of those formants.

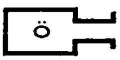
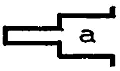
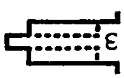
We may formulate our view as follows: the vowel triangle is a conventional two-dimensional structure of phonemically relevant formant sequences. Consequently, for linguistic, that is phonemic purposes, one needs a model that is able to predict the relevant formant sequences (and NOT the absolute formant positions) on basis of parameters that are known to be active in real speech production. Such a model is the loss-free twin-tube resonator with its parameters

k, the constriction factor or opening ratio
β, the eccentricity
γ, the length modulation factor.

In this chapter our aim is to find a set of three parameters for each of the vowels of Dutch. We must bear in mind that no absolute values are called for: it is the SEQUENCE of the corresponding formant positions that counts. Consequently only the correct sequence of the parameters is of importance.

[1] *l.c.*

TABLE X.1

No.	Configuration	β	γ	k
1		$\frac{7}{9}$	$\frac{5}{4}$	$\frac{1}{8}$
2		$\frac{7}{9}$	$\frac{5}{4}$	$\frac{1}{4}$
3		$\frac{7}{9}$	$\frac{5}{4}$	$\frac{1}{2}$
4		0	$\frac{5}{4}$	8
5		$\frac{1}{3}$	1	$\frac{1}{8}$
6		$\frac{1}{3}$	1	$\frac{1}{4}$
7		?	1	1
8		0	1	8
9		0	$\frac{4}{5}$	$\frac{1}{8}$
10		0	$\frac{4}{5}$	$\frac{1}{4}$
11		0	$\frac{4}{5}$	$\frac{1}{2}$
12		$\frac{7}{9}$	1	8

In Table X.1 the normalized values of β, γ and k have been presented. They have been chosen in such a way that their correct sequence has been taken into account. Moreover, pains have been taken to ensure that the set of parameters can be related to a real vocal tract in spite of the fact that their values have been streamlined a little bit in order to arrive at round figures.

The natural limitations of the human anatomy have been, as it were, incorporated in the Table. This especially applies to the range of the parameter γ, the modulation of the length of the vocal tract as measured along its axis between the position of the vocal cords and the free end of the mouth tube.

It is interesting to notice how the length of the vocal tract is influenced by the way in which the mouth tube is being formed.

For instance, in the vowels [i], [I] and [e] (the numbers 9, 10 and 11 of Table X.1) the blade of the tongue plays the dominant role: the lips are inactive. Consequently the vocal tract may be thought to begin BEHIND the teeth. The hump of the tongue is in a low position whereas the position of the throat is high. In other words, there are three reasons why the AXIS of the vocal tract is short. We propose $\gamma = \frac{4}{5}$ as a normalized value in this case.

In the vowels [u], [o], [ɔ] and [ɑ] (the numbers 1, 2, 3 and 4 of Table X.1) the protruded lips are highly involved in the creation of the mouth tube. The blade of the tongue is inactive. The hump of the tongue has a high position whereas the position of the throat is low in this case. Now there are reasons for having a long vocal tract. We propose $\gamma = \frac{5}{4}$ here as a normalized value.

In the vowels [ü], [ö] and [ə] (numbers 5, 6 and 7) the mouth tube is the result of the joint efforts of the lips and the blade of the tongue. The hump of the tongue is low now whereas the throat is not excessively lowered. Therefore the length of the vocal tract is about average so that we propose $\gamma = 1$ for this series. Also in [a] and [ɛ] (numbers 8 and 12) no extreme shortening or lengthening of the axis takes place so that we may group these vowels under $\gamma = 1$ too.

The order of magnitude of the length of the vocal tract can be derived easily from an X-ray photograph by tracing the axis of the tract with the aid of a curvimeter as used in reading maps.

When the formant positions of a vowel are known, the determination of the corresponding twin-tube parameters is highly facilitated by the nomogram shown in figure X.2. For the sake of simplicity the length of the vocal tract has been normalized to the average value l_o. We shall free ourselves from this restriction later on.

Using the procedure illustrated in figure VII.5 we have constructed three lines corresponding with the values $\beta = 0$, $\beta = \frac{1}{3}$ and $\beta = \frac{7}{9}$. The points on these lines represent different values of k, namely, $k = \frac{1}{8}$, $k = \frac{1}{3}$, $k = \frac{3}{5}$, $k = 1$ and $k = 8$. The three lines meet in the point $k = 1$, the neutral position characterized by $F_1 = \frac{c}{4l_o}$ and $F_2 = \frac{3c}{4l_o}$.

Values for γ may be introduced in the nomogram by dividing both co-ordinates of a point by γ, that is by making a point slide along the arrow of its vowel vector (see Chapter VII). In figure X.2 an example has been given for the set of values $\beta = 0$, $k = \frac{1}{8}$, $\gamma = \frac{4}{5}$. The principle of sliding may be applied to other points too, provided we keep the variations of γ within the anatomical limitations of a real vocal tract.

The nomogram brings to the fore the role the parameter β plays in the contrast between the open vowels ($k = 8$). As the nomogram holds good for all languages it is interesting to note the mutual formant positions of English [ɑ], [æ] and [ɛ] in, for instance, figure VII.3. The shape of the diagram formed by those vowels is in agreement with the shape of the diagram formed by the points

$$\gamma = \tfrac{5}{4} \quad \beta = 0 \quad k = 8$$
$$\gamma = 1 \quad \beta = \tfrac{1}{3} \quad k = 8$$
$$\gamma = 1 \quad \beta = \tfrac{7}{9} \quad k = 8$$

in our nomogram displayed in figure X.2. (In Dutch the contrast between [æ] and [ɛ] does not function phonemically.)

The vowels of Dutch represented in Table X.1 may also be grouped in the Hellwag-fashion as shown in figure X.3. This arrangement is in agreement with the corresponding vector diagrams based on F_1 and F_2. Moreover it is very helpful for memorizing the vowel inventory.

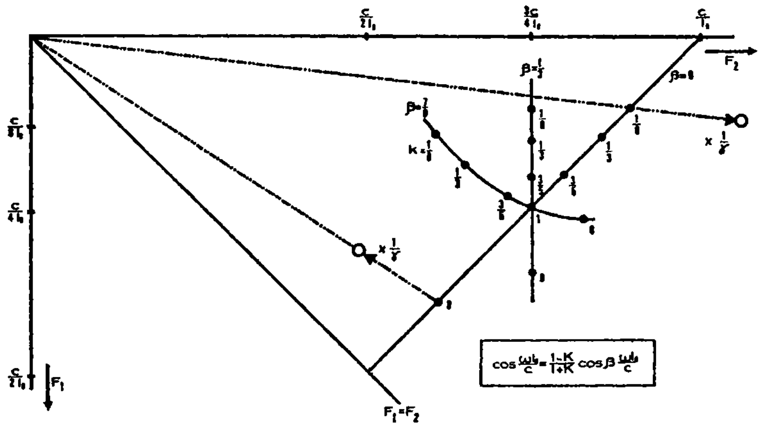


Figure X.2

Nomogram for facilitating the estimation of the normalized values of the twin-tube parameters. The three drawn lines represent three values of β. The points on these lines represent some values of k. Values for γ may be introduced by DIVIDING both coordinates of a point by γ, as is shown for the examples $\beta = 0$, $k = \frac{1}{8}$, $\gamma = \frac{4}{5}$ and $\beta = 0$, $k = 8$, $\gamma = \frac{5}{4}$.

In the following figures we have given inside views of the vocal tract in the midsagittal plane for some characteristic vowels of one and the same Dutch speaker. The drawings have been adapted by Mr G. van Gelder, speech therapist, from X-ray photographs kindly made for me by Prof Dr B.G. Ziedses des Plantes, professor of radiology, director of the X-ray laboratory of the University of Amsterdam.

In every picture the axis of the vocal tract has been indicated, together with the point of articulation according to the normalized value of the twin-tube parameter β for the vowel in question. Also the length of the tract, as measured in cm along the axis by means of a curvimeter, has been indicated.

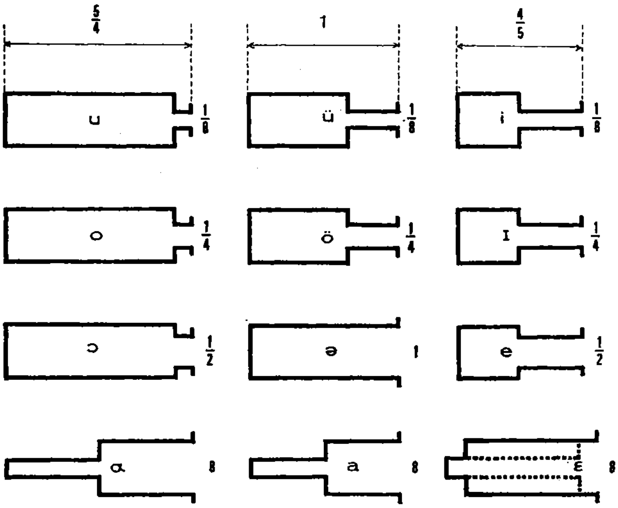


Figure X.3

Hellwag – arrangement of the vowel phonemes of Dutch. Only the values of the parameters γ and k have been indicated.

It stands to reason that views in the midsagittal plane do not permit us to determine the k-parameter with a high degree of accuracy. One should use cross-sections perpendicular to the axis for that purpose or take refuge in conclusions from, for instance, palatograms or measured formant frequencies. It is very helpful, however, that in the chosen examples k is either very high, namely equal to 8, or very low, namely equal to $\frac{1}{8}$.

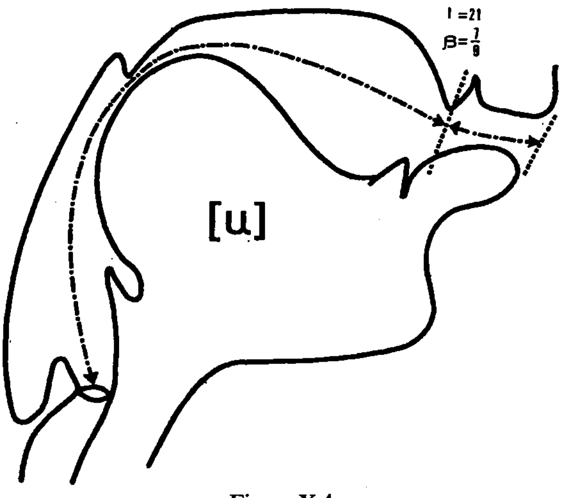


Figure X.4
The vowel [u] in Dutch
The mouth tube is formed solely by the lips. The wide throat tube resembles a diabolo. The vocal tract is extremely long.

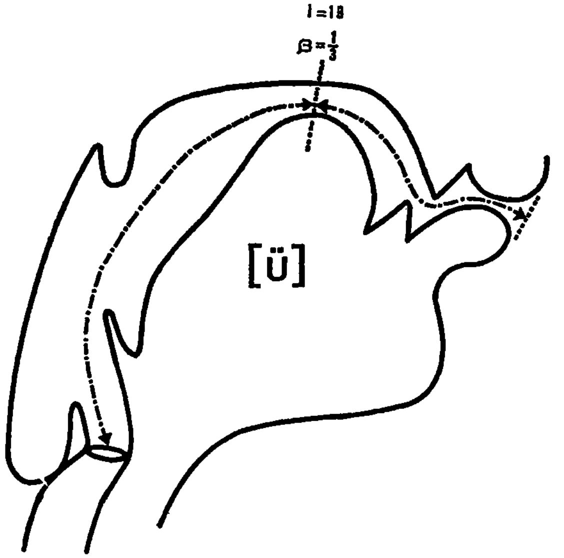


Figure X.5
The vowel [ü] in Dutch
The mouth tube is formed by the lips and the tongue. The length of the tract is about average.

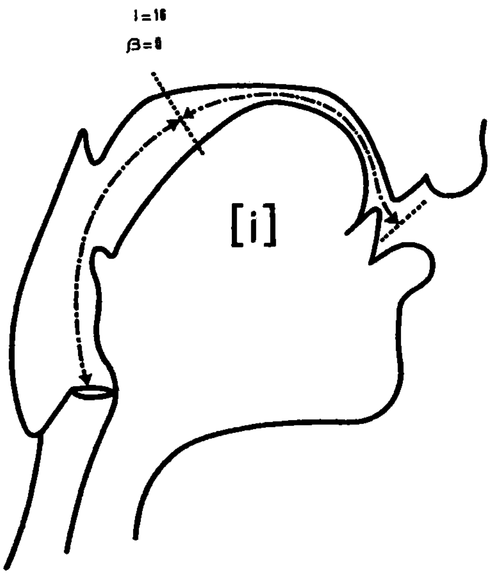


Figure X.6

The vowel [i] in Dutch

The mouth tube is formed solely by the tongue. The vocal tract is extremely short: note the high position of the vocal chords.

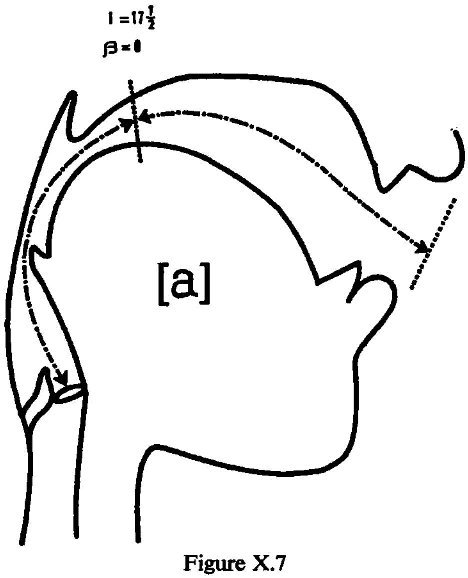


Figure X.7

The vowel [a] in Dutch

The length of the vocal tract is about average. Note the narrow throat tube.

XI

THE FORMANTS OF RESONATORS CONTAINING EXPONENTIAL HORNS

The model of the twin-tube resonator comprises two tubes with uniform cross-areas. Though this simple model suffices for the prediction of the, phonemically relevant, mutual formant contrasts of most vowels, it certainly cannot illustrate everything. The twin-tube method can be successfully applied only to sustained vowels and vowels in carefully pronounced isolated words. It fails, however, to illustrate dynamic processes like, for instance, diphthongs and vowels in free running speech for reasons we shall bring to the fore later on. The cause of this failure must not be sought in the twin-tube method itself. The real cause lies in the fundamentally different ways of production of isolated vowels on the one hand and of diphthongs and free running speech on the other hand. Therefore we shall, in this chapter, discuss another formant generating model that is better geared to the needs of dynamic processes.

The study of X-ray photographs of the real vocal tract provides a cue for modifying the twin-tube model. One notices that parts of the vocal tract may roughly have the shape of an hour-glass in spite of the fact that they are being approximated by tubes with a uniform cross-area. Two questions arise

(a) how far may one constrict a tube without endangering the twin-tube model as a method for predicting the formant contrasts between the vowels

(b) does excessive constriction possibly have a function in speech production.

We shall try to answer both questions by calculation. A simple way of calculating a constricted part of the vocal tract is to treat it as an exponential horn. As we have seen in Chapters IV and V, it is possible to solve Webster's horn equation for the exponential horn; the values of the general circuit parameters of this type of horn are given by the equations (V.26), (V.27), (V.28) and (V.29). For the purpose of calculating the exponential diabolo, see figure XI.2, it is not necessary to fully elaborate these equations by filling in b_1 and b_2. For the time being we may keep them simply. It is convenient, however, to denote the total length of a diabolo by the symbol l. For that reason we shall denote the length of the exponential horn by $\frac{1}{2}l$, so that we arrive at the following expressions for the general circuit parameters of the exponential horn.

$$A = \varepsilon^{\frac{1}{2}ml} \frac{b_2 \varepsilon^{\frac{1}{2}b_2 l} - b_1 \varepsilon^{\frac{1}{2}b_1 l}}{b_2 - b_1} \tag{XI.1}$$

$$B = \frac{j\omega\varrho_o}{S_o} \cdot \frac{\varepsilon^{\frac{1}{2}b_2 l} - \varepsilon^{\frac{1}{2}b_1 l}}{b_2 - b_1} \tag{XI.2}$$

$$C = j \frac{S_o \varepsilon^{\frac{1}{2}ml}}{\omega\varrho_o} b_1 b_2 \frac{\varepsilon^{\frac{1}{2}b_2 l} - \varepsilon^{\frac{1}{2}b_1 l}}{b_2 - b_1} \tag{XI.3}$$

$$D = \frac{b_2 \varepsilon^{\frac{1}{2}b_1 l} - b_1 \varepsilon^{\frac{1}{2}b_2 l}}{b_2 - b_1} \tag{XI.4}$$

The reader will have noticed that we have also changed S_l into $S(\frac{1}{2}l) = S_o \varepsilon^{\frac{1}{2}ml}$.

We shall now calculate the general circuit parameters of a series arrangement of two exponential horns which are each other's images as depicted in figure XI.2.

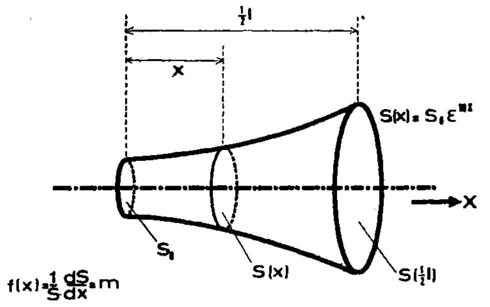


Figure XI.1
The exponential horn and its characteristics.

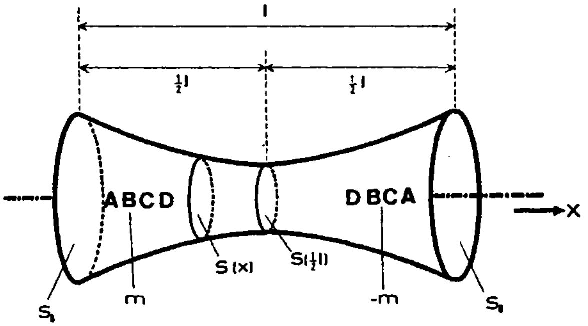


Figure XI.2
Series arrangement of an exponential horn and its image (so-called exponential diabolo).

Note: in this case $m < 0$ because we have defined:
$$S(x) = S_o \varepsilon^{mx}.$$

When a four-terminal network is reversed its parameters A and D change places whereas its parameters B and C remain the same. This principle is illustrated in figure XI.3. Making use of the formula's (III.18), (III.19), (III.20) and (III.21) we arrive at

$$A_t = A_1 A_2 + B_1 C_2 = 2AD - 1 = 2BC + 1 \qquad \text{(XI.5)}$$
$$B_t = A_1 B_2 + B_1 D_2 = 2AB \qquad \text{(XI.6)}$$
$$C_t = A_2 C_1 + C_2 D_1 = 2CD \qquad \text{(XI.7)}$$
$$D_t = D_1 D_2 + C_1 B_2 = 2AD - 1 = 2BC + 1 \qquad \text{(XI.8)}$$

being the general circuit parameters of the exponential diabolo.

Now we know its parameters we are able to use the diabolo as a building block for models of the vocal tract. Let us first consider a diabolo that represents the complete vocal tract. This case is illustrated in figure XI.4.

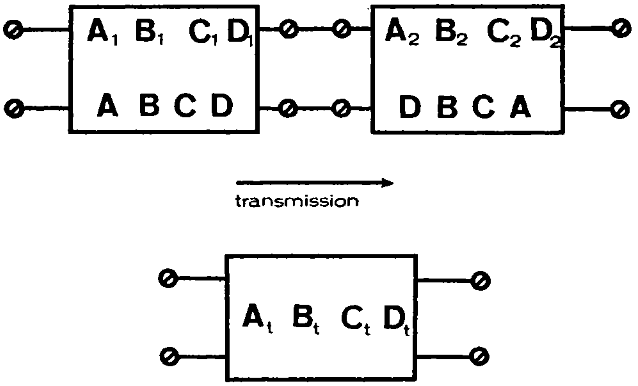


Figure XI.3

The general circuit parameters of the series arrangement of a tube and its image.

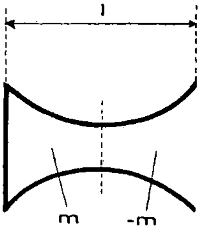


Figure XI.4

The formants of a diabolo closed at one end. They are given by $D_t = 0$.

We find the formants by putting $D_t = 0$ and making use of (XI.8), (XI.2) and (XI.3) which yields:

$$\cos\sqrt{\frac{\omega^2}{c^2} - \frac{m^2}{4}}\, l = \frac{m^2 c^2}{4\omega^2} \tag{XI.9}$$

In other words, the formants come to the fore as the point of intersection between the functions

$$\cos\sqrt{\frac{\omega^2}{c^2} - \frac{m^2}{4}}\, l \text{ and } y = \frac{m^2c^2}{4\omega^2}$$

It is very tempting, though not worthwhile, to try to perform a graphical solution of (XI.9) with the aim of describing the vocal gesture which is the result of pinching a tube with constant cross-area in the middle, so that an exponential diabolo comes into being. It is scarcely rewarding to do so because (XI.9) pertains to the stationary condition; its application to the rapid change in the formant positions measured, for instance, during the roll-off of a diphthong is rather risky. Things are still more complicated by the fact, that during the diphthong, at least in Dutch, the repetition frequency of the vocal cords diminishes appreciably. As experiments with vowel synthesizers show, interesting phase effects may result.

With a view to the foregoing, it is more realistic to use (XI.9) for describing the beginning and the end of a vocal gesture.

Starting from $m = 0$, the tube with constant cross-area, we have

$$\cos\frac{\omega l}{c} = 0 \qquad \text{(XI.10)}$$

with the well-known formants

$$F_1 = \frac{1}{4}\frac{c}{l} \qquad \text{(XI.11)}$$

and

$$F_2 = \frac{3}{4}\frac{c}{l} \qquad \text{(XI.12)}$$

When we pinch the tube in the middle, then $m \neq 0$, and there will appear a formant *below* $\frac{1}{4}\frac{c}{l}$, given by

$$\omega = \frac{mc}{2} \qquad \text{(XI.13)}$$

or

$$F = \frac{m\,c}{4\pi} \tag{XI.14}$$

This, after all is nothing but the paradoxical cut-off frequency of the exponential horn.

When we suppose that (XI.9) is able to describe the relatively stationary beginning and end positions of the gesture, we have proved that, in some way or another, F_1 has fallen down from the value $\frac{1}{4}\frac{c}{l}$ to the value $\frac{m}{4}\frac{c}{\pi}$, in accordance with what we measure in practice.

We shall not, in this monograph, attempt to predict by calculation the avalanche effect in the shift of F_1 in the diphthong variant of the twin-tube model. As I have explained elsewhere[2], it is possible to materialize our models, both of vowels and diphthongs, in hardware and to drive these by means of an artificial throat in the form of an acoustic siren. The dynamic exponential twin-horn of the diphthongs, for instance, may be realized as a water-filled rubber sleeve which produces a constriction around its middle when water is pumped into it. The sound curves produced with the aid of this gadget display the accelerated shift of F_1, accompanied by the (in Dutch) necessary decrease of the frequency of the air puffs (falling intonation).

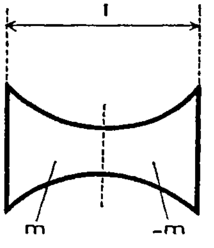


Figure XI.5

The formants of a diabolo closed at both ends. They are given by $C_t = 0$.

[2] H. Mol, The vowel siren as a tool in speech research, *Nomen, Leyden studies in linguistic and phonetics* (Mouton, The Hague 1969), pp. 104-113.

For cases of close saturation it is interesting to calculate the stationary waves in a diabolo closed at BOTH ends, see figure XI.5. In order to find the formants we must now make use of the C parameter of the diabolo. This can best be seen from (III.10). The condition that $p_l = 0$ in spite of the fact that $U_o = U_l = 0$ obviously requires that $C = 0$.

The C parameter of the diabolo, called C_t in this chapter, is given by (XI.7). We finds the formants by putting $C_t = 0$ and making use of (XI.3) and (XI.4) which yields:

$$\tan\sqrt{\frac{\omega^2}{c^2} - \frac{m^2}{4}}\,\tfrac{1}{2}l = -\frac{2}{m}\sqrt{\frac{\omega^2}{c^2} - \frac{m^2}{4}} \qquad \text{(XI.15)}$$

and

$$\sin\sqrt{\frac{\omega^2}{c^2} - \frac{m^2}{4}}\,\tfrac{1}{2}l = 0 \qquad \text{(XI.16)}$$

We see that expression (XI.15) is sensitive to the sign of m. Starting from $m = 0$ we have

$$\cos\frac{\omega l}{2c} = 0 \quad \text{and} \quad \sin\frac{\omega l}{2c} = 0 \qquad \text{(XI.17)}$$

These formulas together produce the well-known formants of the tube with constant cross-area closed at both ends:

$$F_1 = \frac{c}{2l},\ F_2 = 2\frac{c}{2l},\ F_3 = 3\frac{c}{2l},\ \text{etc etc.}$$

So the lowest formant is $F_1 = \dfrac{c}{2l}$, a half-wavelength mode of the tube.

It is apparent from (XI.15) and (XI.16) that in this case also the cut-off frequency $\omega = \dfrac{mc}{2}$ is a formant, but its frequency is so low that we need not take it into account for not too high values of m. We shall, therefore, concentrate on the way in which the original frequency $F_1 = \dfrac{c}{2l}$ behaves when the tube is made to shrink ($m < 0$)

or to bulge ($m > 0$) around its 'waist'. It is possible to derive from (XI.15) that

$$F_1 < \frac{c}{2l} \text{ for } m < 0$$

$$F_1 > \frac{c}{2l} \text{ for } m > 0.$$

Interestingly enough, this behaviour with respect to the sign of m may best be illustrated by the triplet-tube model depicted in figure XI.6. This model can be explained via the twin-tube model likewise shown in the figure. In the twin-tube model there is, by definition, a pressure node at the mouth-opening, meaning that the (radiation) impedance 'looking' into space is equal to zero.

Also, for all formant frequencies of the twin-tube, the impedance looking back into the mouth-opening is equal to zero as well, see Appendix § 2). Consequently, when we load a twin-tube by its image, its original formants, including F_1, will remain the same,

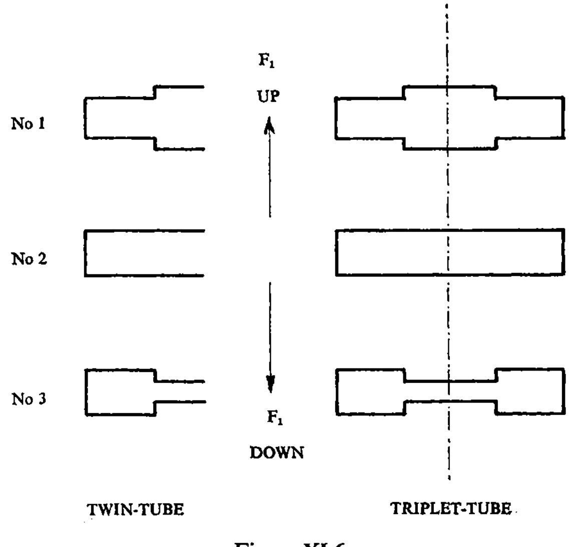


Figure XI.6
Paired configurations with the same first formant.

though higher modes are added to it. As F_1 of the open twin-tube No. 1 is higher than that of the neutral twin-tube No. 2, and F_1 of the closed twin-tube No. 3 is lower than that of the neutral tube, we may conclude the following:

F_1 of the double configuration No. 2 goes down by pinching the tube and goes up by making the tube bulge around its middle. Therefore the triplet-tube may be considered as a model for the exponential diabolo.

We shall now derive a general formula for the formants of a twin-tube resonator in which the throat tube has been replaced with a diabolo, see figure XI.7.

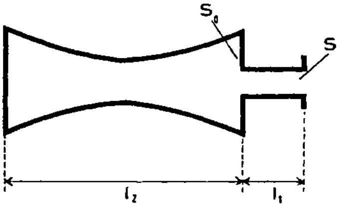


Figure XI.7
Back diabolo variant of the twin-tube resonator.

The formant formula is as follows:

$$D_2D_1 + C_2B_1 = 0 \qquad \text{(XI.18)}$$

By studying the formulas (XI.2), (XI.3) and (XI.4) we learn that C_2 is proportional to S_o and that B_1 is inversely proportional to S, whereas D_1 and D_2 do not contain the factors S and S_o. Consequently, in the special case of close saturation where $\frac{S}{S_o} \rightarrow o$ the product C_2B_1 completely overrides the term D_1D_2 so that (XI.18) folds down to simply:

$$C_2B_1 = 0 \qquad \text{(XI.19)}$$

Now $B_1 = 0$ produces the formants contributed by the mouth tube, whereas

$$C_2 = 0 \tag{XI.20}$$

leads to nothing but (XI.15) and (XI.16), pertaining to a diabolo closed at both ends, with its corresponding triplet-tube depicted in figure XI.6.

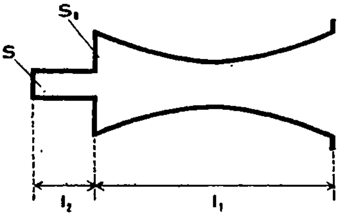


Figure XI.8
Front diabolo variant of the twin-tube resonator.

When, as is shown in figure XI.8, the mouth tube is replaced with a diabolo, application of (XI.18) leads to:

$$D_2D_1 + C_2B_1 = 0 \tag{XI.21}$$

In the special case of open saturation $\frac{S}{S_o} \to 0$, so that (XI.21) reduces to:

$$D_1D_2 = 0 \tag{XI.22}$$

leading to $D_2 = 0$ and $D_1 = 0$, which in turn, is nothing but (XI.9), pertaining to a diabolo closed at ONE end.

In practice one meets realisations of [u] with a second formant that is considerably lower than the value suggested by the half wave-length resonance of the back cavity. Obviously, the hour-glass constriction is at work here for which the exponential diabolo is a convenient model that lowers F_2. In this case the back diabolo variant of the twin-tube resonator, depicted in figure XI.7, may be applied and, more in particular, its variant of close saturation illustrated in figure XI.6. It is interesting to notice how, in the case of [u], the additional lowering of F_2 as a result of the diabolo-effect, enhances the contrast between [u] and [i].

In order to check the assumptions worked into the calculations of our models we have constructed several hard-walled plastic models, four of which have been depicted in figure XI.9.

The throat tubes of these four resonators have the same length l_2 and the same volume V. This may be easily checked by filling them with water and measuring their contents with the aid of a graduated glass. The mouth tubes of the resonators are identical. The classical Helmholtz formula, complete with end correction, for the FIRST formant of the three resonators, is as follows

$$F_1 = \frac{c}{2\pi}\sqrt{\frac{S_1}{(l_1 + 0.85d_1)V}} \qquad \text{(XI.23)}$$

When we consider the second formant of the resonators No 2 and No 3 as the half wave-length resonance of the saturated throat tube we have

$$F_2 = \frac{c}{2l_2} \qquad \text{(XI.24)}$$

Formula (XI.24) does not hold good for resonator No 1, containing a conical diabolo.[3] Neither does it hold good for resonator No 4.

The values for F_1 and F_2, obtained with the aid of formulas (XI.23) and (XI.24) and the numerical data presented in figure XI.9 have been collected in Table XI.1.

There is an easy method of producing the separate formants of necked resonators, namely, by blowing over the edge of the neck, in that way giving birth to a periodical series of eddies, in the same manner as the edge-tone of an organ pipe, see figure XI.10.

[3] For reasons of mechanical simplicity this diabolo has been made conical instead of exponential.

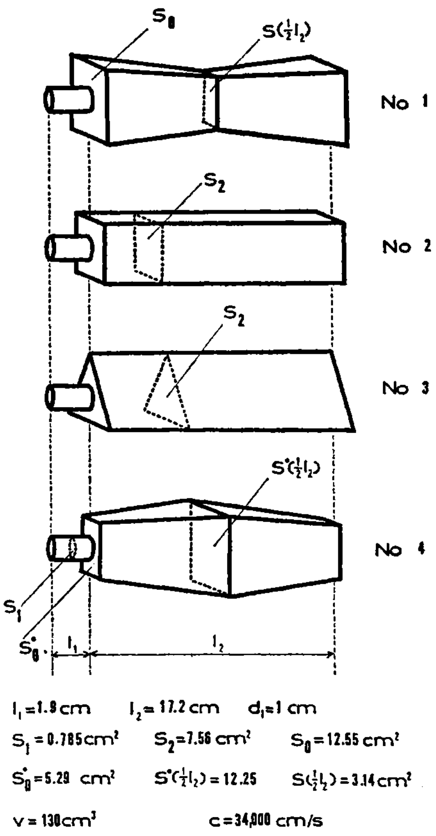


Figure XI.9

Comparison of the formants of four particular resonators.

By adjusting the velocity of blowing one is able to excite the desired formant: the higher the velocity, the higher the formant. The periodicity of a formant may be conveniently and accurately determined by feeding the tone via a microphone into an electronic counter.

Strictly speaking, the edge tone-technique of exciting the formants is in itself a formant definition. The discrepancies between the calculated and measured frequencies must partly be judged in this light.

Table XI.1 shows that resonators No 2 and No 3 are practically

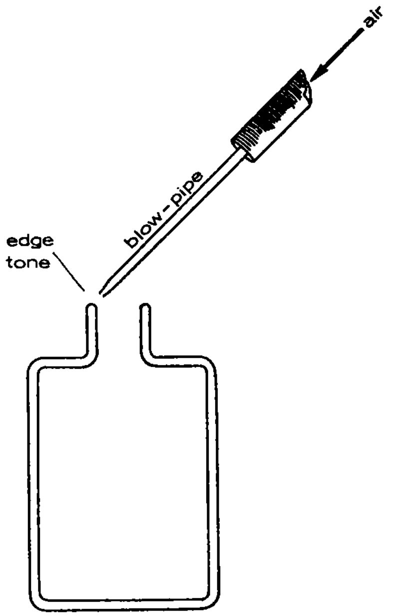
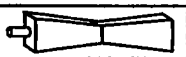
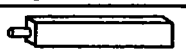
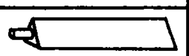
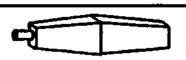


Figure XI.10
Formant excitation by means of edge tones.

TABLE XI.1

Resonator No.	F_1 (Hz)		F_2(Hz)	
	calculated	measured	calculated	measured
1 	254	250	×	830
2 	254	250	988	1065
3 	254	247	988	1068
4 	254	247	×	1205

identical in acoustic respect. Obviously, the SHAPE of the cross-area, be it square or triangular, is of little importance.

The resonators No 1, No 2, and No 3 have practically the same F_1 because they have the same volume and the same mouth tube. The second formant of No 1, however, is considerably lower than those of No 2 and No 3. This reduction in frequency is the result of a corresponding reduction of the cross-area in the middle of the throat tube by a factor 4. Though resonator No 1 contains a conical diabolo, it displays the same trend as our model with the exponential diabolo: a reduction of F_2.

As mentioned earlier in this chapter, the formant-generating expression (XI.15) is sensitive to the sign of the flare m. This behaviour should be borne out by our plastic resonators. Resonator No 4 in figure XI.9, though it is a conical diabolo, is meant to illustrate the case of an exponential twin-horn with a positive value of m. As shown in Table XI.1, resonator No 4 is characterized by $F_2 = 1205$ Hz, zo indeed above the $\frac{1}{2}\lambda$ resonance. This rise is, in this case, the result of an augmentation of the cross-area in the middle of the throat tube by a factor 2.

Applications of the saturated twin-tube with front diabolo will be discussed in Chapter XII.

XII

THE TWIN-TUBE RESONATOR WITH FRONT DIABOLO AS A MODEL FOR THE FORMANTS OF MONOLITHIC DIPHTHONGS AND VOWELS IN RUNNING SPEECH

There is a traditional tendency to use the term: diphthong for any sequence of two vowels. In this chapter we shall not deal with sequences of two vowels in which BOTH vowels have a phonemic function. This is, for instance, the case in DUTCH for:

taai, [t] [a] [i], meaning "tough"
taak, [t] [a] [k], meaning "task"

Here final [i] is opposed to the consonant [k] and has, as it were, the function of a consonant. The elements [a] and [i] are easily recognizable without the intervention of instrumental tricks like segmentation techniques, perhaps because they function as separate phonemes in the (Dutch) language.

We shall restrict ourselves here to discussing three Dutch diphthongs of a vastly different nature, namely the diphthongs in normal orthography written as *ou* (or also *au*), *ui* and *ei* (or also *ij*). They appear, for instance, in the following key-words:

bout [b] [ou] [t], meaning "bolt"
buit [b] [ui] [t], meaning "loot"
meid [m] [ei] [t], meaning "maid"
bijt [b] [ei] [t], meaning "bite".

The phonetic symbols [ou], [ui] and [ei] do not in the least pretend to indicate the nature of the two vowel elements: they are merely (ill-)inspired by orthography. Translated into phonetic reality

[ou] means going from something near [ɑ] to something near [u]
[ui] means going from something near [ʌ] to something near [ü]
[ei] means going from something near [ɛ] to something near [i].

As a matter of fact this is not a very satisfying description but until now no one has done very much better.

TABLE XII.1

Dutch word	Phonetic transcription	English translation
bout	b ou t	bolt
buit	b ui t	loot
bijt	b ei t	bite
boet	b u t	barn
boot	b o t	boat
bot	b ɔ t	blunt
baat	b a t	profit
beet	b e t	bit
bed	b ɛ t	bed
biet	b i t	beet
bad	b ɑ t	bath
hout	h ou t	wood
hoed	h u t	hat
haat	h a t	hatred
had	h ɑ t	had
huid	h ui t	skin
heet	h e t	hot
kous	k ou s	stocking
kaas	k a s	cheese
kas	k ɑ s	cash
kees	k e s	Dutch boy's name
kies	k i s	molar
kuis	k ui s	chaste
koos	k o s	chose
kus	k ʌ s	kiss
	etc., etc.	

Distributional properties of the monolithic diphthongs in Dutch. In order to avoid possible confusion inspired by orthography, we have emphasized the diphthongs.

We shall, in this chapter, call the diphthongs [ou], [ui] and [ei] MONOLITHIC diphthongs because, whatever their articulatory structure may be, they function as single phonemes in Dutch. Consequently, the complete monolithic diphthong is opposed to other vowels or other monolithic diphthongs, as is shown in Table XII.1.

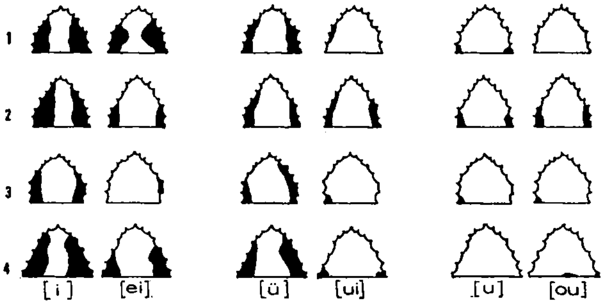


Figure XII.1

Palatograms of four different speakers pronouncing the three monolithic diphthongs of Dutch. They are depicted together with the palatograms of the corresponding final elements pronounced as separate vowels.
All palatograms have been taken from the collection present at the Institute of Phonetic Sciences of the University of Amsterdam. This collection was set up by the then director dr. Louise Kaiser.

Apart from distributional reasons there are phonetic grounds for putting the monolithic diphthongs in a special class in Dutch.

In the first place there is the fact, that the elements of a monolithic diphthong can only be brought to attention by segmentation techniques. The elements are only heard as such when they are separately exposed in ABSENCE of the rapid transition from the first element to the second. When the transition is separately exposed the Dutch listener states he 'hears' the 'complete' diphthong.

In figure XII.2 the oscillogram of Dutch *ei* has been depicted. It shows what one hears when certain parts of the sound curve are picked out and presented separately to a listener. The [ɛ] like element has a high F_1 clearly visible in the picture as a damped oscillation whereas the [i] like element displays a much lower F_1. In the normal perception of a word containing [ei] the rapid intervocalic transition heard as [ei] completely overrides the other cues coming from [ɛ] and [i]. Figures XII.3 and XII.4 show the remaining two monolithic diphthongs of the same speaker.

As the oscillograms show, the intervocalic transition from the first element to the second one takes place in the very short time-interval of some six vocal periods. Now the, generally, slow tongue movements do not, in connection with the twin-tube model,suggest such rapid changes of the formants. Therefore, we must look for an accelerating mechanism that translates relatively slow movements of the articulators into rapid movements of the formants in the frequency domain.

It is a known fact, see figure XII.1, that, for the monolithic diphthongs of Dutch, the palatograms of the final elements do NOT coincide with the palatograms of these elements when they are being realized as separate vowels. In the cases of [ei] and [ui] we see very clearly, that the blade of the tongue has less contact with the palate than in [i] and [ü], especially in front. The case of [ou] is not very illuminating because neither in [ou] nor in [u] the contact between the tongue and the palate is of any importance. We may safely assume, that the monolithic diphthongs are realized via an articulatory mechanism that is different from that of the normal vowels. More in particular we may say that the tip of the tongue is lowered in the diphthongs.

Experiments with two-formant vowel generaters show that the monolithic diphthongs of Dutch can be acceptably produced by keeping F_2 constant and varying only F_1, it being understood that each diphthong has its own fixed value of F_2.

Summarizing a model generating the vocal gestures of the monolithic diphthongs of Dutch should have the following features

a. a constant F_2
b. a suitable palatogram
c. an accelerating mechanism.

These requirements are met by the saturated twin-tube with front diabolo as discussed in Chapter XI and depicted in figure XI.8.

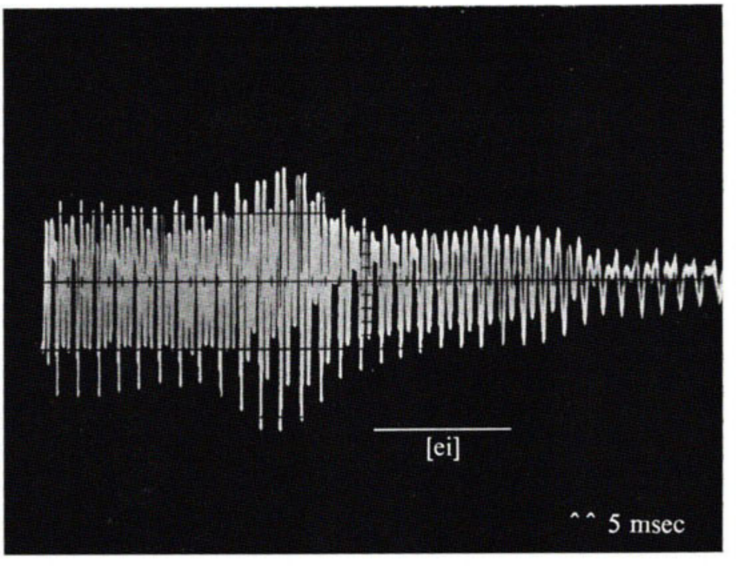


Figure XII.2
Oscillogram of the monolithic diphthong [ei] in Dutch.

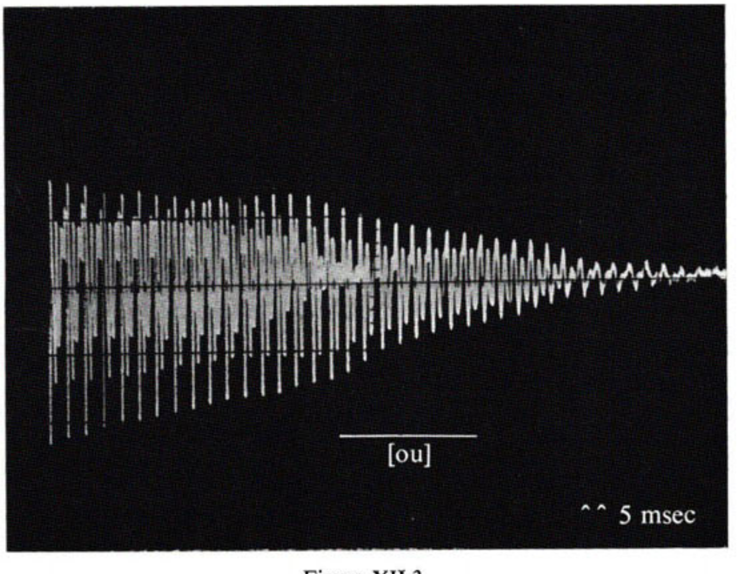


Figure XII.3
Oscillogram of the monolithic diphthong [ou] in Dutch.

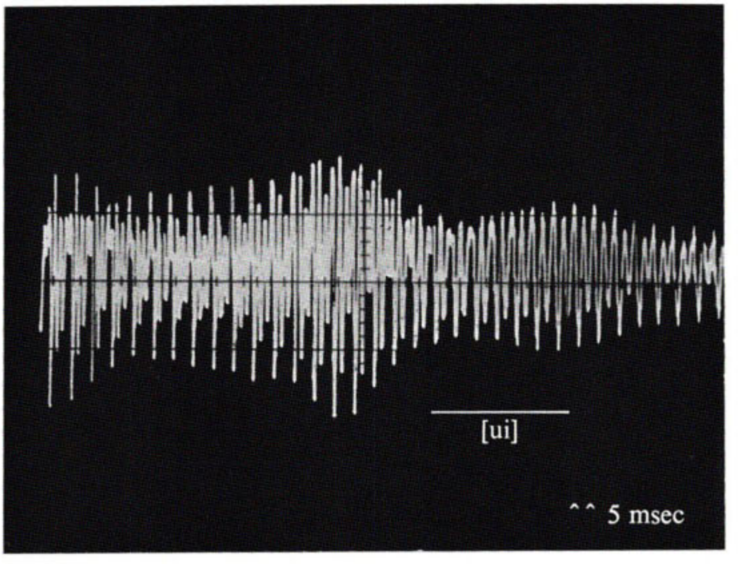


Figure XII.4
Oscillogram of the monolithic diphthong [ui] in Dutch.

The constant F_2 is provided by the $\frac{1}{4}$ wave-length mode of the throat tube having the frequency $\frac{c}{4l_2}$ (unless that frequency is higher than that of the $\frac{3}{4}$ wave-length mode of the mouth tube, but the latter frequency is practically constant too.

The widening shape of the front diabolo is in harmony with the requirement of less tongue-palate contact in front.

Unfortunately, we have not been able to predict the avalanche effect in F_1 by calculation but, as shown in Chapter XI, we have verified this effect experimentally by realizing the model in hardware.

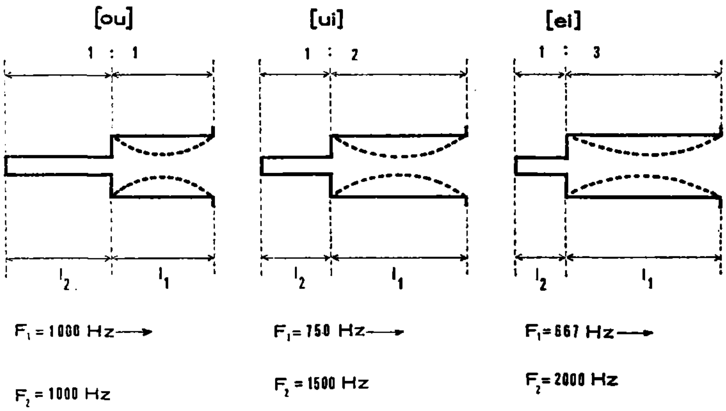


Figure XII.5

The saturated twin-tube with front diabolo as a model for generating the formants of the monolithic diphthongs of (accepted) Dutch.

Note: The arrows behind the first formants indicate that F_1 makes a downward shift when the front-tube is squeezed into the shape of a diabolo.

Figure XII.5 illustrates the three values of F_2 we have selected for our model that generates the vocal gestures of the three monolithic diphthongs of Dutch.

The selected values of F_2 at the same time pin down the starting points of F_1. The dotted diabolo's indicate that the F_1-shifts are

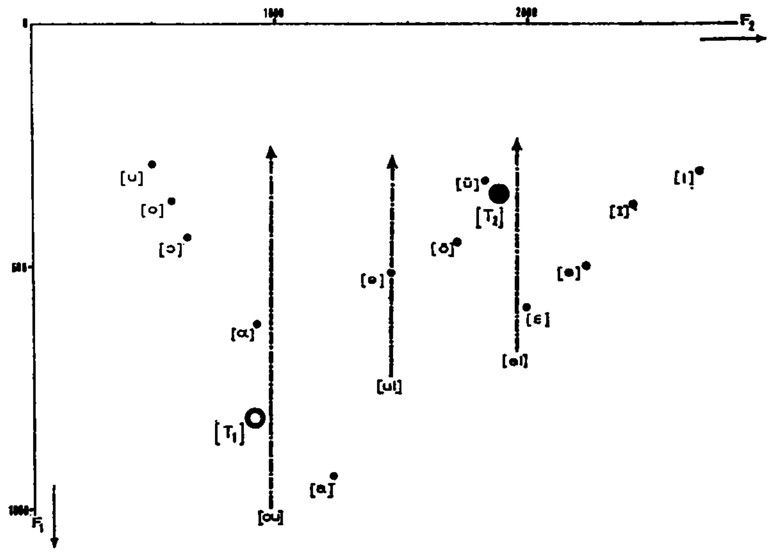


Figure XII.6

The three normalized vocal gestures of the saturated twin-tube model with front diabolo, corresponding with the monolithic diphthongs [ou], [ui] and [ei] of Dutch. The gestures have been placed against the background of the vowel system of Dutch, including the two (tentative) vowel targets [T_1] and [T_2] of running speech.

being brought about by 'diabolic' constriction of the mouth tube, symbolized by the change of the flare *m* of one half of the symmetrical diabolo.

Anybody studying the oscillograms of the monolithic diphthongs will discover that it is not necessary to run through the complete gestures indicated in figure XII.6. As predicted by my co-operator Mr. J. G. Blom[1] a limited gesture in the correct direction often suffices. Moreover, he will discover that the length of the final element depends on the phonetic environment in which the diphthong appears and may in point of fact be very short.

[1] Private communication.

In order to possible avoid confusion we here want to emphasize that we restrict ourselves in this monograph to the monolithic diphthongs of accepted Dutch. For instance, we here leave alone the other monolithic diphthongs in Dutch, or in other languages, where the intervocalic transition, be it accelerated or not, does NOT carry the phonemic cue. We hope to discuss the diphthong problems in more detail in a separate monograph.

On several occasions, we drew attention to the fact that in free running speech one uses a vowel system that is vastly different from the triangular set of formant positions characteristic of isolated key words.[2] The behaviour of the talkers in running speech can best be described by saying that, see figure XII.7, they do not aim at 12 vowel targets (the 12 vowels of Dutch) but instead of 12 at 2 only. The centres of these two targets have been transferred from figure XII.7 to figure XII.6 and introduced there as [T_1] and [T_2]. Though their positions are admittedly only tentative and need further confirmation by big scale measurements, they give rise to the following hypothesis:

In free running speech one is inclined, one talker more so than the other, to realize the vowels via a reduced system of articulation that may be described by the model of the saturated twin-tube with front diabolo.

The basic shape of all vowels is the open vowel with narrow throat tube.

One contents oneself with producing only two different contrasting lengths of the throat tube.

One contents oneself with producing only two different contrasting ('diabolic') constrictions of the mouth tube.

The shortest throat tube always appears together with the narrowest mouth tube, whereas, on the other hand, the longest throat tube always appears together with the widest mouth tube.

[2] See, for instance, H. Mol, *Fundamentals of Phonetics I, The Organ of Hearing* (Mouton & Co, The Hague 1963), fig. 21.

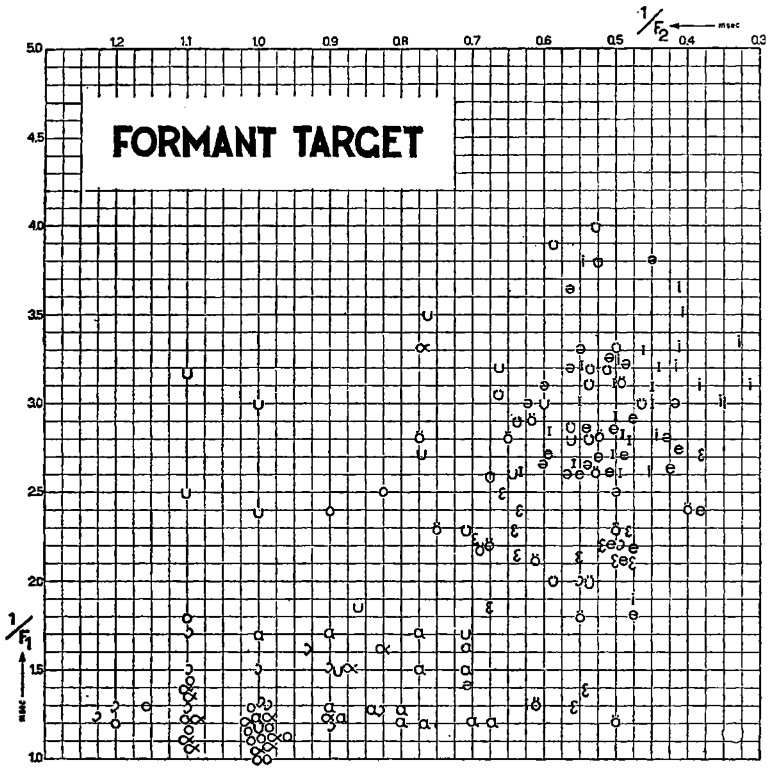


Figure XII.7

(Taken from Fundamentals of Phonetics I: The Organ of Hearing by H. Mol, Mouton & Co, The Hague 1963). Period of second formant versus frequency of first formant for 12 Dutch vowels appearing in a simple, freely pronounced sentence. The data of this pilot investigation refer to 15 talkers taken at random from a group of 100 talkers. The numbers on the axes represent periods, in harmony with the fact that the formants were indeed measured as periods and not as frequencies.

XIII

APPENDIX

1. NASALITY

As early as 1951 Pierre Delattre states[1]: "Starting from an oral vowel, if we lower the velum while holding all other speech organs immobile, the frequency of formant 3 rizes considerably while the frequencies of formants 1 and 2 remain fairly stable."

In the same paper he indicates how the rise of formant 3 "does not seem to be related to the nasal quality of the nasal vowels".

So, obviously, the cue for nasality must be carried by other things than the frequency locations of the formants. Especially in languages where this cue is thought to have a phonemic function it will be worthwhile to look for it in the sound waves leaving the vocal tract of the speaker. Now as the twin-tube method only predicts the formant positions, it is at first sight unfit for application to the nasal vowels. However, the experiment described by Delattre is somewhat misleading. This is apparent from what he remarks about the French nasal vowels when he describes what happens when one denasalizes these vowels "by raising the velum and holding all other organs immobile": "The results of such denasalizing does not give the French oral vowels [ɛ], [œ], [ɔ], [ɑ] but some strange vowels that do not exist in French (nor probably in any language), for the organic positions of the four French nasals (and their formants 1 and 2) are not the same of any French orals. This can

[1] Pierre Delattre, "The physiological Interpretation of Sound Spectrograms", *PMLA*, Volume LXVI No 5 (Sept, 1951), pp. 864-875.

be shown by synthetic speech as well as by human speech".

In other words, the French nasals already clearly distinguish themselves from the orals by their formant positions, a contrast that can indeed be described by the twin-tube model. On top of that comes the cue of nasality, for which the twin-tube model in its present form cannot present a description. One might raise the interesting question whether the direct acoustic cue of lowering the velum as such, really functions phonemically.

2. THEVENIN'S THEOREM AS A VEHICLE FOR DEFINING THE FORMANTS IN A LOSS-FREE ACOUSTICAL NETWORK

According to Thévenin's theorem the current U through the impedance Z belonging to an electrical network, see figure XIII.1, may be written as:

$$U = \frac{e}{Z + Z_o} \qquad \text{(XIII.1)}$$

where Z_o is the impedance of the network looking into the terminals 1, 2 and the electromotoric force e is the open voltage across the terminals 1, 2, that is the voltage when Z is removed.

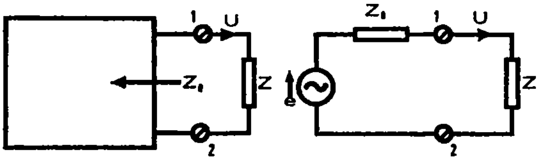


Figure XIII.1
Thévenin's theorem.

As a rule e, Z and Z_o are functions of frequency. Now, when there are no losses in the network, it is possible that for some special frequencies:

$$Z + Z_o = 0 \qquad \text{(XIII.2)}$$

so that:

$$|U| = \infty \tag{XIII.3}$$

(provided nothing dramatic happens to e for those frequencies). The frequencies, given by (XIII.2), for which (XIII.3) is true may be defined as resonance frequencies or, in this case, as formant frequencies.

In the analogous acoustical case, we may apply (XIII.2) for defining and calculating the formants of the twin-tube resonator, see figure XIII.2.

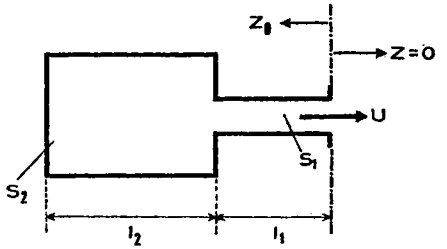


Figure XIII.2
Thévenin – method of defining the formants of the twin-tube resonator.

As the volume velocity U in the mouth opening comes to mind first as an interesting victim of resonance phenomena, we have

$$Z = 0 \tag{XIII.4}$$

so that (XIII.2) boils down to

$$Z_o = 0 \tag{XIII.5}$$

as formant generating equation.

Let us first derive some auxiliary formula's. From (III.25) we may directly conclude, that the acoustical impedance of a tube, characterized by its general circuit parameters A, B, C and D, and terminated by an impedance Z_2, is given by

$$Z_1 = \frac{AZ_2 + B}{CZ_2 + D} \tag{XIII. 6}$$

Starting from (XIII.6) and at the same time applying (VI.1) we can easily derive the following formulas for a tube with constant cross-area S and the length l:

$$\text{closed tube: } \frac{\varrho_o c}{jS}\cot\frac{\omega}{c} \quad (Z_2 = -j\infty) \qquad \text{open tube: } j\frac{\varrho_o c}{S}\tan\frac{\omega l}{c} \quad (Z_2 = 0) \tag{XIII.7}$$

We are now able to calculate Z_o via (XIII.6). We have to consider that A, B, C and D pertain to the mouth tube and that Z_2 refers to the throat tube closed at the throat end. Moreover, it is sufficient to introduce only A and B because $Z_o = 0$ when

$$AZ_2 + B = 0 \tag{XIII.8}$$

After some elaboration of (XIII.8) we finally arrive at:

$$\tan\frac{\omega l_1}{c}\tan\frac{\omega l_2}{c} = \frac{S_1}{S_2} \tag{XIII.9}$$

This is nothing but the well-known formant formula of the twin-tube.

As illustrated in figure XIII.3 there is another way of finding the formants of the twin-tube.

Instead of the volume velocity in the mouth opening we may study the resonances of the volume velocity at the point of articulation, that is at the place where the mouth tube and the throat tube meet.

In this case application of (XIII.2) leads to simply:

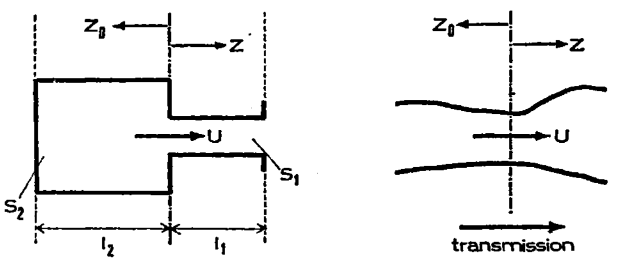


Figure XIII.3

Thévenin-method of defining the formants of the twin-tube resonator.

$$\frac{\varrho_o c}{jS_2}\cot\frac{\omega l_2}{c} + j\frac{\varrho_o c}{S_1}\tan\frac{\omega l_1}{c} = 0 \qquad \text{(XIII.10)}$$

or, again:

$$\tan\frac{\omega l_1}{c}\tan\frac{\omega l_2}{c} = \frac{S_1}{S_2} \qquad \text{(XIII.9)}$$

When calculating the formants of the twin-tube Fant[3] also divides the resonator at the point of articulation. Principally, it should be possible to cut the resonator at an arbitrary spot but we shall not deal with this somewhat tricky problem in this monograph.

3. NUMERICAL SOLUTION OF THE FORMANT FORMULA

For those exploring the behaviour of the twin-tube model numerical solution of the formant formula in the following form

$$\cos\frac{\omega l}{c} = \frac{1-k}{1+k}\cos\beta\frac{\omega l}{c} \qquad \text{(XIII.11)}$$

is of great help. At the Institute of Phonetic Sciences of the University of Amsterdam this solution is performed on the computer IBM 1130 available there. The program relevant to this purpose was made by mr J.G. Blom, engineer in chief at this Institute, who called it TWNTE.

Application description

Program TWNTE is a users oriented computer program to solve the formant formula of the twin-tube model.

The program is self-instructive. No programming knowledge is required.

Within reasonable limits the user can make a free choice of all parameters including the velocity of sound.

For each set of parameters the program computes the first 4

[3] Fant, *l.c.* p. 63.

formants, with an accuracy of 1 c/s, using the method of Bolzano iteration.

Diagnostic messages are provided.

The program is written in Fortran IV (subset) for IBM 1130 but can easily be modified for other computers.

Minimum machine configuration:

Central processing unit
Console typewriter
Console printer

The following items are given below:

TYPICAL OUTPUT
SAMPLE OF DIAGNOSTIC MESSAGES
PROGRAM LISTING.

```
PROGRAM TWNTE,
TWNTE COMPUTES F1,F2,F3,F4 OF THE TWINTUBE MODEL OF THE VOCAL TRACT.
DEFINITION OF PARAMETERS.

                          .   ..............
                          .   .
      ....................
      .                   .
      .     01            .            02
      .                   .
      ....................
                          .   .
                          .   ..............
                          X D X
      X     0.5L          X      0.5L       X


K=02/01 (RATIO OF CROSSSECTIONS)
BETA=D/0.5L (RELATIVE ECCENTRICITY)
L=LENGTH IN CM
V=35000. CM/SEC (VELOCITY OF SOUND)


TO CHANGE V SET DATA ENTRY SWITCH 14.
USE KEYBOARD TO ENTER PARAMETER VALUES.
USE A DECIMAL POINT AND PRESS END OF FIELD KEY.
IF YOU STRIKE A WRONG KEY PRESS ERASE FIELD KEY AND REENTER PARAMETERS.
FIRST PRESS PROGRAM START KEY.


        K                BETA               L
   (        )         (        )         (        )        V= 35000.CM/S
      1.                  0.            17.5

F1=   500.C/S   F2=  1500.C/S   F3=  2500.C/S   F4=  3500.C/S


PRESS PROGRAM START KEY TO ENTER NEW PARAMETERS.
SET DATA ENTRY SWITCH 15 AND PRESS PROGRAM START KEY TO STOP.

        K                BETA               L
   (        )         (        )         (        )        V= 35000.CM/S
      8.                  0.            17.5

F1=   784.C/S   F2=  1217.C/S   F3=  2784.C/S   F4=  3217.C/S

        K                BETA               L
   (        )         (        )         (        )        V= 35000.CM/S
     .125                 0.            17.5

F1=   217.C/S   F2=  1784.C/S   F3=  2217.C/S   F4=  3784.C/S

        K                BETA               L
   (        )         (        )         (        )        V= 35000.CM/S
     .125                 0.            13.

F1=   292.C/S   F2=  2401.C/S   F3=  2984.C/S   F4=  5093,C/S

        V
   (        )
     70000.
RESET DATA ENTRY SWITCH 14 AND PRESS PROGRAM START KEY.
```

```
     K                  BETA               L
  (       )          (        )         (       )         V= 70000.CM/S
    1.                  0.               17.5

F1=  1000.C/S   F2=  3000.C/S   F3=  5000.C/S   F4=  7000.C/S

     K                  BETA               L
  (       )          (        )         (       )         V= 70000.CM/S
    8.                  0.               17.5

F1=  1568.C/S   F2=  2433.C/S   F3=  5567.C/S   F4=  6433.C/S

     V
  (         )
   35000.
RESET DATA ENTRY SWITCH 14 AND PRESS PROGRAM START KEY.

     K                  BETA               L
  (       )          (        )         (       )         V= 35000.CM/S
   .125                .78               22.

F1=   246.C/S   F2=   971.C/S   F3=  1823.C/S   F4=  2697.C/S

END OF TWNTE
```

```
     K                  BETA               L
  (       )          (        )         (       )         V= 35000.CM/S
 7.5                   1.              17.5
** PARAMETER OUT OF FIELD,
** ENTER PARAMETERS AND PRESS END OF FIELD KEY.
     K                  BETA               L
  (       )          (        )         (       )         V= 35000.CM/S
     12.               1.              17.5
** K OUTSIDE RANGE 0.1,10,
** ENTER PARAMETERS AND PRESS END OF FIELD KEY.
     K                  BETA               L
  (       )          (        )         (       )         V= 35000.CM/S
    8.                 2.                17.5
** BETA OUTSIDE RANGE -1,+1,
** ENTER PARAMETERS AND PRESS END OF FIELD KEY.
     K                  BETA               L
  (       )          (        )         (       )         V= 35000.CM/S
    8.                .5                 28.
** L OUTSIDE RANGE 10,25,
** ENTER PARAMETERS AND PRESS END OF FIELD KEY.
     K                  BETA               L
  (       )          (        )         (       )         V= 35000.CM/S
```

```
// JOB
// FOR
*LIST AML
*ONE WORD INTEGERS
*IOCS(TYPEWRITER,KEYBOARD,DISK)
*LIST SOURCE PROGRAM
C
C     PROGRAM TWNTE
C
C     THIS BASIC FORTRAN IV WRITTEN PROGRAM COMPUTES
C     THE F1,F2,F3 AND F4 OF THE TWINTUBE-MODEL.
C     TWNTE IS PRIMARY WRITTEN FOR AN IBM 1130 SYSTEM,
C     BUT CAN EASELY BE MODIFIED FOR EVERY COMPUTER
C     WITH TYPEWRITER IN AND OUTPUT.
C
C     AUTHOR  J.G.BLOM.
C
C     TWNTE
      REALLNGTH
      DIMENSIONFOR(4) ,AC(6)
    5 FORMAT('PROGRAM TWNTE,'/'TWNTE COMPUTES F1,F2,F3,F4 OF THE ',
     1'TWINTUBE MODEL OF THE VOCAL TRACT.'/
     1'DEFINITION OF PARAMETERS.'/)
   10 FORMAT(22X,'.',3X,'..............')
   15 FORMAT(22X,'.',3X,'.')
   20 FORMAT(6X,'.....................')
   25 FORMAT(6X,'.',15X,'.')
   30 FORMAT(6X,'.',5X,'O1',8X,'.',10X,'O2')
   35 FORMAT(22X,'X D X')
   40 FORMAT(6X,'X',5X,'0.5L',6X,'X',6X,'0.5L',6X,'X'//)
   45 FORMAT('K=O2/O1 (RATIO OF CROSSSECTIONS)'/
     1'BETA=D/0.5L (RELATIVE ECCENTRICITY)'/
     1'L=LENGTH IN CM'/
     1'V=35000. CM/SEC (VELOCITY OF SOUND)')
   50 FORMAT(//'TO CHANGE V SET DATA ENTRY SWITCH 14.'/'USE KEYBOARD TO
     1ENTER PARAMETER VALUES.'/'USE A DECIMAL POINT AND PRESS END OF FIE
     1LD KEY.'/'IF YOU STRIKE A WRONG KEY PRESS ERASE FIELD KEY AND REEN
     1TER PARAMETERS.'/'FIRST PRESS PROGRAM START KEY.'//)
   55 FORMAT(8X,'K',15X,'BETA',16X,'L'/
     13X,'(         )',9X,'(         )',9X,'(         )',9X,'V=',F7.0,'CM/S')
   60 FORMAT(/'PRESS PROGRAM START KEY TO ENTER NEW PARAMETERS.'/
     1'SET DATA ENTRY SWITCH 15 AND PRESS PROGRAM START KEY TO STOP.'/)
   65 FORMAT(8X,'V'/3X,'(          )')
   70 FORMAT('** PARAMETER OUT OF FIELD.')
   75 FORMAT('** ENTER PARAMETERS AND PRESS END OF FIELD KEY.')
   80 FORMAT('** L OUTSIDE RANGE 10,25,')
   85 FORMAT('** BETA OUTSIDE RANGE -1,+1,')
   90 FORMAT('** K OUTSIDE RANGE 0.1,10,')
   95 FORMAT('** V OUTSIDE RANGE 10000,100000,')
  100 FORMAT('RESET DATA ENTRY SWITCH 14 AND PRESS PROGRAM START KEY.'/)
  105 FORMAT('END OF TWNTE'//)
  110 FORMAT(3(F4.0,2F7.0))
  115 FORMAT(/4('F',I1,'=',F7.0,'C/S',3X)/)
  120 FORMAT(F4.0,F9.0,F4.0)
      INT = 0
      C = 35000.
C     C    =VELOCITY OF SOUND
      WRITE(1,5)
      WRITE(1,10)
      WRITE(1,15)
      WRITE(1,20)
```

```
      WRITE(1,25)
      WRITE(1,30)
      WRITE(1,25)
      WRITE(1,20)
      WRITE(1,15)
      WRITE(1,10)
      WRITE(1,35)
      WRITE(1,40)
      WRITE(1,45)
      WRITE(1,50)
  125 PAUSE
      CALL DATSW(15,NSTOP)
      GOTO(260,130),NSTOP
  130 CALL DATSW(14,NV)
      GOTO(135,165),NV
  135 WRITE(1,65)
      B = 0.
      BB = 0.
      READ(6,120)B,C,BB
      B = B + BB
      IF (B) 140,145,140
  140 WRITE(1,70)
      WRITE(1,75)
      GOTO 135
  145 IF (C - 10000.) 155,150,150
  150 IF (C - 100000.) 160,160,155
  155 WRITE(1,95)
      WRITE(1,75)
      GOTO 135
  160 WRITE(1,100)
      PAUSE
  165 WRITE(1,55)C
      DO 170 I=1,6
  170 AC(I) = 0.
      READ(6,110)AC(1),AK,AC(2),AC(3),BETA,AC(4),AC(5),LNGTH,AC(6)
C     AK   =RATIO OF CROSSSECTIONS
C     BETA =RELATIVE ECCENTRICITY
C     LNGTH=LENGTH OF TWINTUBE
      CALL DATSW(15,NSTOP)
      GOTO(260,175),NSTOP
  175 DO 180 I=2,6
  180 AC(1) = AC(1) + AC(I)
      IF (AC(1)) 185,190,185
  185 WRITE(1,70)
      WRITE(1,75)
      GOTO 130
  190 IF (LNGTH - 10.) 200,195,195
  195 IF (LNGTH - 25.) 205,205,200
  200 WRITE(1,80)
      WRITE(1,75)
      GOTO 130
  205 IF (ABS(BETA) - 1.) 215,215,210
  210 WRITE(1,85)
      WRITE(1,75)
      GOTO 130
  215 IF (AK - 10.) 220,220,225
  220 IF (AK - .1) 225,230,230
  225 WRITE(1,90)
      WRITE(1,75)
```

```
      GOTO 130
  230 P = 6.283825 * LNGTH / C
      P = 6.283825 * LNGTH / C
      Q = P * BETA
      R = (1. - AK) / (1. + AK)
      F = 0.
      DO 250 K=1,4
      N = 0
      DF = 32.
      T1 = ( - 1.) * * K
      GOTO 240
  235 DF = DF / 2.
      IF (DF - .25) 250,250,240
C     STOP ITERATION IF STEP SMALLER THAN .5 C/SEC.
  240 N = N + 1
      T2 = ( - 1.) * * N
  245 F = F - T2 * DF
      IF (T1 * T2 * (COS(P * F) - R * COS(Q * F))) 235,250,245
  250 FOR(K) = F + .5
      WRITE(1,115)(I,FOR(I),I=1,4)
      INT = INT + 1
      IF (INT - 1) 255,255,125
  255 WRITE(1,60)
      GOTO 125
  260 WRITE(1,105)
      CALL EXIT
      END
```

INDEX