

**ACOUSTIC THEORY
OF SPEECH PRODUCTION**

DESCRIPTION AND ANALYSIS OF CONTEMPORARY STANDARD RUSSIAN

Edited by

ROMAN JAKOBSON

Harvard University

AND

C. H. VAN SCHOONEVELD

Stanford University

II

1970
MOUTON
THE HAGUE · PARIS

ACOUSTIC THEORY OF SPEECH PRODUCTION

*With Calculations based on X-Ray Studies
of Russian Articulations*

by

GUNNAR FANT

Royal Institute of Technology Stockholm

Second printing

1970
MOUTON
THE HAGUE · PARIS

© Copyright 1960 in The Netherlands.
Mouton & Co. N.V., Publishers, The Hague.

*No part of this book may be translated or reproduced in any form by print, photoprint, microfilm,
or any other means, without written permission from the publishers.*

FIRST PRINTING: 1960

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

INTRODUCTION

This monograph is intended as a contribution to the understanding of the relations between physiological data concerning speech production and physical data from the description of the speech wave. One part of this task is general in character and involves a survey of the theoretical foundations of the description of speech signals and of the prediction of acoustic signals from articulatory data.

Part I is devoted to the general theory of speech production and calculation techniques with an emphasis on the acoustics of compound resonator systems which is treated on the basis of equivalent circuit theory. It serves as a theoretical foundation for the specific studies undertaken in *Part II* as well as a reference for the theory of simple resonator systems and the theory of sound sources in speech, dealt with in the appendices.

Part II presents the results of calculations based on the X-ray studies of Russian articulations. The production of all standard Russian vowels and consonants was X-rayed in 1951 by Dr. A. S. MacMillan and Dr. G. Kelemen at the Massachusetts Eye and Ear Infirmary, Harvard Medical School, under a plan of investigation drawn up by Prof. R. Jakobson, Harvard University, Prof. M. Halle, Massachusetts Institute of Technology (M.I.T.), and the author. These records and calculations have been prepared as a part of the project "Description and Analysis of Contemporary Standard Russian", directed by R. Jakobson in connection with the Department of Slavic Languages and Literatures at Harvard University, under the sponsorship of the Rockefeller Foundation.

The main object of these calculations was to attempt a reconstruction of the spectra of speech sounds from physiological data of their production. This involves a comparison of samples of connected speech with corresponding data derived from sustained forms of a subject's articulation during X-ray photography. Some critical comments on the traditional pattern of articulatory descriptions are added.

Part III provides a summary of the relations between articulation and speech wave, with applications to the theory of distinctive features. The first section of this part is intended to supplement the beginning of *Part I*, being introductory in character and more linguistically oriented.

The calculations extended over a long period of time. This was partially due to the complexity of the material under analysis. First, numerical calculations were carried out with paper and pencil, and rather rough approximations were needed in order to keep calculation times within practical limits. The next step was the use of high-speed electronic computing machines (e.g., the *BESK* machine in Stockholm), which made it possible to perform more detailed studies in a short time. Such calculations were performed on the vowel material in 1953–1954. The calculations on consonants were made in 1954–1957 with the aid of the transmission line analog *LEA* at the Speech Transmission Laboratory of the Division of Telegraphy and Telephony, Royal Institute of Technology (R.I.T.), Stockholm. The calculations of nasal sounds were delayed by the lack of physiological data. Independent investigations at the M.I.T. and at the R.I.T. have led to similar results.

It was felt at the completion of this work that despite their high technical quality, the X-ray studies employed and the vocal tract dimensions derived therefrom did not possess the degree of accuracy and completeness necessary for a complete exploitation of the potentialities of modern analog and numerical calculation techniques. High-speed X-ray film and an improved technique for measurements of the whole vocal tract are needed in order to broaden our knowledge of speech as articulatory events.

The acoustic theory of vowels is by now fairly well established, but there remains much work to be done on the description of the general properties of consonants. The calculations were carried out on the basis of simplified assumptions with regard both to physiological data and to theoretical foundations. It is believed, however, that the results permit a critical evaluation of the approximations utilized and that they will indicate the potentialities of the acoustic theory of speech production in its present form.

ACKNOWLEDGMENTS

The author greatly appreciates the stimulating discussions with Professor R. Jakobson and Professor M. Halle and the considerable amount of attention they devoted to the initial study of the X-ray material. The methods used for the conversion of articulatory data to acoustic data were influenced by very close contact with the speech research group under Professor K. N. Stevens at the Acoustics Laboratory of the Massachusetts Institute of Technology. The study of nasal sounds was made possible by the research on cavity dimensions performed by Dr. G. Bjuggren of Sabbatsbergs Sjukhus, Stockholm. The author is extremely grateful to Professor Torbern Laurent of the Royal Institute of Technology, Stockholm, for his never-failing encouragement and helpful advice. Valuable editorial aid was given by Dr. W. Jassem, the Technical Institute of the Polish Academy of Science, Poznań. The author is also indebted to lektor Eli Fischer-Jørgensen of the University

of Copenhagen, to Dr. J. L. Flanagan, Bell Telephone Laboratories, and to Professor M. Halle and Dr. H. M. Truby, R.I.T., for critical comments on the form and contents of the mimeographed pre-edition (Fant, 1958). My thanks are also due to Mrs. M. Richter, Mrs. S. Felicetti, and other members of the Speech Transmission Laboratory, R.I.T., for all their work in the preparation of this publication and the book.

The work was made possible by financial support from the Rockefeller Foundation, the Wallenberg Foundation, the State Council of Technical Research in Sweden, the Research Institute of the National Defence, Sweden, and USAF Contract, AF 61 (514)-1084.

PREFACE TO THE SECOND EDITION

The second edition of *Acoustic Theory of Speech Production* is identical in contents with the first edition. A few minor corrections have been made and a subject index and an author index has been added. The subject index should allow the reader to look up and locate specific technical terms and topics. It is felt that the main scope of the book is to provide the theoretical basis of the static aspects of the theory of speech production and specification whilst the study of speech dynamics belongs to a more recent period of developments. *Acoustic Theory of Speech Production* is also of interest to musical acoustics. The data given in the Appendices 2 and 3 and the general theory given in the main body of the book is of interest in analysis of wind instrument performance. A recent review of the field of acoustic analysis and synthesis of speech is given in the author's contribution to *Manual of Phonetics*, editor B. Malmberg, North-Holland Publ. Co., 1968.

The author is grateful to F. Fransson for help with corrections and to Si Felicetti for her careful editorial work.

Stockholm, August 1969

Gunnar Fant

TABLE OF CONTENTS

Introduction	5
Acknowledgments	6
PART I ACOUSTIC THEORY OF SPEECH	
<i>Chapter 1.1 GENERAL THEORY</i>	15
1.11 Source-Filter Description of Speech Production	15
1.12 Segmentation	21
1.13 The F-pattern	24
<i>Chapter 1.2 NETWORK THEORY OF VOCAL TRANSMISSION</i>	27
1.21 Network Representation of Acoustic Resonators and Horns	27
1.22 Methods of Numerical Calculations	36
1.23 Transform Equations for Speech Production	42
A. The Vocal Tract Transfer Function	42
B. Radiation, Source, and Other “Constant” Factors	44
C. The Complete Laplace Transform	45
D. The Inverse Transform	46
<i>Chapter 1.3 ANALYTICAL CONSTRAINTS ON THE COMPOSITION OF SPEECH SPECTRA</i>	48
1.31 Idealized Spectral Description of Voiced Sounds	48
1.32 The Relations Between Formant Frequencies and Spectrum Envelopes	54
1.33 Pole-Zero Decomposition of Consonants	60
<i>Chapter 1.4 THE F-PATTERNS OF COMPOUND TUBE RESONATORS AND HORNS</i>	63
1.41 The Twin-Tube Resonator. The Effect of Lip-Rounding	63
1.42 Horns as Single Resonators and Connecting Sections	67
1.43 Three-Parameter Models Approximating the Vocal Tract	71
A. Models Composed of Cylindrical Sections Only	72
B. Models With Horn-Shaped Tongue Section	79

PART II CALCULATIONS BASED ON X-RAY DATA

<i>Chapter 2.1</i> X-RAY PROCEDURE, SUBJECT, AND PHONETIC MATERIAL	93
<i>Chapter 2.2</i> METHODS AND APPROXIMATIONS	97
<i>Chapter 2.3</i> A STUDY OF VOWELS	107
2.31 Calculations of Formant Frequencies and Spectrum Envelopes	107
2.32 Articulatory and Acoustic Vowel Diagrams	110
2.33 The Relations Between Resonator Dimensions and Formant Frequencies	113
2.34 The Spatial Distribution of Sound Pressure. Formant Bandwidths	125
A. Calculations of Formant Levels on an Absolute Pressure Basis	125
B. Sound Pressure Distribution Within the Vocal Tract	128
C. The Dependencies of Formant Bandwidths on Various Resistive Elements Within the Vocal Tract	135
<i>Chapter 2.4</i> NASAL SOUNDS AND NASALIZATION	139
2.41 Physiological Data	139
2.42 Nasal Sounds Produced With Oral Closure	142
2.43 Nasalization	148
<i>Chapter 2.5</i> THE LIQUIDS	162
<i>Chapter 2.6</i> FRICATIVES, AFFRICATES, AND STOPS	169
2.61 Fricatives and Affricates	169
2.62 Stops	185
2.63 Idealized Models of Fricatives and Stops	191
2.64 Conclusions Regarding Source Characteristics of Fricatives and Stops	201

PART III SUMMARY

<i>Chapter 3.1</i> SEGMENTATION AND SPECIFICATION	207
<i>Chapter 3.2</i> THE RELATIONS BETWEEN THE F-PATTERN AND ARTICULATION	209
<i>Chapter 3.3</i> SOME ASPECTS OF THE THEORY OF DISTINCTIVE FEATURES	212
<i>Chapter 3.4</i> COMMENTS ON THE ACOUSTICAL NATURE OF DISTINCTIVE FEATURES	215

APPENDICES

<i>Appendix A.1 SPEECH WAVE ANALYSIS</i>	229
A.11 Intensity Measurements	229
A.12 Spectrum and Waveform Measurements	234
A.13 Spectrographic Illustrations of the Speech Material Utilized for the Control of the Consonant Calculations	242
<i>Appendix A.2 A STUDY OF SOURCE CHARACTERISTICS</i>	265
A.21 The Voice Source	265
A.22 Turbulent and Transient Sources	272
<i>Appendix A.3 ANALYTICAL STUDY OF SIMPLE RESONATOR MODELS WITH APPLICATIONS TO SPEECH PRODUCTION</i>	281
A.31 The Single Helmholtz Resonator	281
A. One Opening	281
B. Two Openings	283
A.32 The Double Helmholtz Resonator	284
A.33 The Single Tube as an Acoustic Resonator	290
A.34 Four-Tube Systems. Transform Equations for Arbitrary Source Locations	297
A.35 The Damping Effect of Series and Shunt Losses Within Twin-Tube Resonators	300
A.36 Summary of Twin-Tube Formulas for the Study of Resonator Damping. Applications to Vocal Tract Models	303
A. The Helmholtz Resonator	304
1. Radiation Resistance	304
2. Frictional Losses in the Resonator Neck Under Ideal Conditions	304
3. The Effect of a Superimposed Turbulent Air Stream. The Glottal Shunt	305
4. Heat Conduction and Cavity Wall Vibration Losses	306
B. Standing Waves in Tubes	307
1. Radiation Damping of an Open Tube	307
2. Radiation Damping of Standing Waves in a Tube With a Narrow Opening	308
3. Cavity Wall Damping of Standing Waves	308
C. Experiments and Calculations on the Performance of Single and Twin-Tube Hard-Walled Resonators	309
1. Single Tube, Open at one End	309
2. Twin-Tube Resonator Model	310
<i>Selected Bibliography</i>	313

PART I

ACOUSTIC THEORY OF SPEECH

1.1 GENERAL THEORY

1.11 *Source-Filter Description of Speech Production*

The speech wave is the response of the vocal tract filter systems to one or more sound sources. This simple rule, expressed in the terminology of acoustic and electrical engineering, implies that the speech wave may be uniquely specified in terms of *source* and *filter* characteristics. In spite of the technical phrasing it is apparent that this statement also covers essentials of the phonetician's concept of speech production.

The source-filter theory of speech production is exemplified by the block diagram of *Fig. 1.1-1* which contains a number of interconnected filter sections, each one representing a part of the vocal cavities. A vocal cord sound source is indicated in diagram a). The coupling of the nasal cavities to the rest of the vocal tract at the boundary between the pharynx cavity and the mouth cavity is indicated.

Other block diagrams of similar typography but with the separate filter blocks labeled *front cavity* and *back cavity* suggest themselves as a technical realization of phonetic terms. In such a description there easily arises a confusion as to the actual physiological boundaries intended. Thus, in addition to the possible reference to the mouth as front cavity and to the pharynx as back cavity there exists the possibility of dividing the vocal tract with reference to the sound source. An alternative division, in the case of vowels a subdivision, is in terms of the narrowest passage in the vocal tract. In fricatives and stops this region is close to the source but does not necessarily coincide with the source location. It should also be observed that the point of articulation of vowels, when defined by the highest point of the tongue, may come far from the place of maximum narrowing, the latter constituting the acoustically relevant basis for a back cavity versus front cavity division. This discrepancy is apparent for back vowels. In the case of very open vowels a division of the cavity system above the larynx into separate parts loses some of its significance. The more open a vowel is, the less will the separate parts of the vocal cavities act as independent resonators.

Diagrams of the type shown in *Fig. 1.1-1b* are used in quantitative treatments of

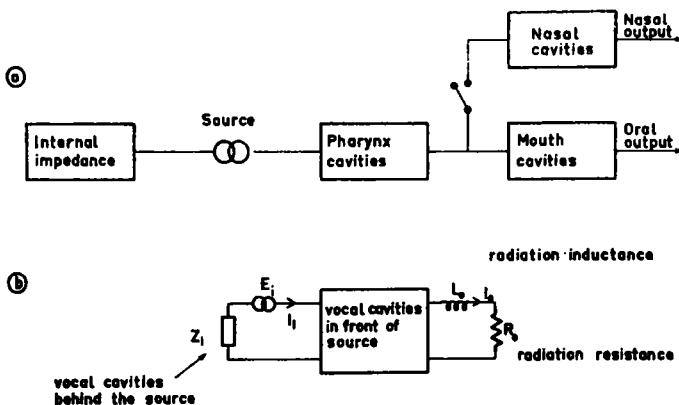


Fig. 1.1-1. Speech production as a filtering process.

- Schematic representation of the production of sounds employing a glottal source.
- Four-terminal network representation of the production of any non-nasal sound irrespective of the source location.

speech production within the frame of electrical circuit theory. The cavities in front of the source are represented by a four-terminal network, the output terminals of which are connected to the radiation impedance. The two-terminal network behind the source contains the source impedance¹ in series with the impedance of the back cavities.

The complete *filter function*, in technical terminology often referred to as the *transfer function*, is defined as the frequency dependent ratio of the pressure in the soundfield at a specified point in front of the speaker to the pressure of the sound source or to its volume velocity. In a more detailed analysis, see *Sections 1.23 and 1.31*, the frequency dependent properties of radiation from the mouth-opening of the speaker are considered as a separate factor within the filter function, in which case the intermediate reference for the output of the vocal cavities is the volume velocity through the lips which in *Fig. 1.1-1b* is denoted by the electrical symbol I_0 .

With the symbolic notations of S for source and T for the transfer function of the vocal tract filter, the product $P = S \cdot T$ represents the corresponding speech sound. In a more general and precise definition each of these categories is time and frequency dependent.²

¹ A current $i(t)$, amp, in the electrical circuit corresponds to a volume velocity $u(t)$, cm³/sec, in the corresponding acoustic system. Volume velocity $u(t) = v(t)A$ is the product of particle velocity $v(t)$, cm/sec, multiplied by the cross-sectional area A cm² of the system perpendicular to the direction of the airflow or oscillation. Voltage $e(t)$, volt, corresponds to pressure $p(t)$, dyne/cm². The ratio of the pressure to the volume velocity in terms of frequency transforms is the analogous acoustic impedance $Z = P(f)/U(f)$, dyne sec/cm³.

² The frequency dependency is conveniently stated by means of Laplace transforms with $s = \sigma + j\omega$ as the complex frequency variable: $P(s) = S(s)T(s)$. (1.1-1)

This reads that the Laplace transform $P(s)$ of the pressure in the soundfield in front of the speaker is

There is some degree of correspondence between the phonetic term *phonation* and the technical term *source* and similarly between *articulation* and *filter*. This analogy implies, of course, that phonation is held apart from articulation in the sense of the generation of sound versus the specific shaping of its phonetic quality. The vocal tract filter system is dependent on the position of the articulators and a direct translation is possible, at least when dealing with idealized vocal tract models, as will be shown in *Chapter 1.4*.

When discussing the production of speech it should be noted that the source S of the formula $P = S \cdot T$ is an acoustic disturbance superimposed upon the flow of respiratory air and is caused by an object giving rise to friction or to a transient release or checking of the air stream and, in the case of voiced sounds, by a quasi-periodic modulation of the airflow due to the opening and closing movement of the vocal cords. Prior to the acoustical stage of phonation where the source S belongs, there is thus an aerodynamic stage comprising expiration (sometimes inspiration) of air, the main parameters of which are the volume of air exhaled (or inhaled) per unit time and the subglottal air pressure. The latter should not be confused with the acoustic sound pressure of a source.

The basic property of a vocal cord sound source is its periodicity expressed by the duration T_0 of a complete voice period or by the inverse value of the *voice fundamental frequency*

$$F_0 = 1/T_0. \quad (1.1-2)$$

Fundamental pitch and fundamental frequency are not synonymous but these terms can often be used interchangeably due to the close one-to-one correspondence. In more strict terminology *pitch* is a tonal *sensation* and *frequency* a property of the sound *stimulus*. The duration of a pitch cycle always varies somewhat from one period to the next. Such variations are in part systematic determining the intonation pattern, in part accidental or rather unintentional, but nevertheless of importance for the naturalness of human speech. Only speaking machines are capable of producing a perfectly monotonic pitch.

A voice source is further characterized by its *spectrum envelope* $S(f)$ which is a specification of the amplitudes of the source harmonics as a function of their frequency. This source spectrum envelope reflects personal characteristics of the speaker but varies also with voice register, fundamental pitch, and voice intensity. In quasi-

the product of the source transform $S(s)$ and the transfer function, i.e., the filter transform $T(s)$. An extensive use will be made of Laplace transforms within this work since they provide an efficient tool for making concise statements concerning waveform and spectrum aspects of speech production. Laplace transforms possess a greater generality than Fourier transforms, but the latter are in some instances a useful supplement for the acoustic specifications. As Fourier series have been extensively made use of in phonetic theory since Helmholtz, they will be utilized in Fig. 1.1-2 providing a simplified description of the production of voiced sounds. The time dependency may be included by substituting $P(s)$ for $P(s,t)$, etc.

scientific terminology this description is misleadingly referred to as the number of overtones present in the voice.

The classification of speech sounds in terms of source characteristics has its root in classical phonetics. The terms *harmonic/inharmonic*, referring to the spectrum, and their respective waveform synonyms *periodic/non-periodic* apply to the existence versus non-existence of a vocal sound source. Apical or uvular trills are also produced with a periodic variation of the filter function, in the instance of voiced variants, the main source being supplied from the vocal cords. In unvoiced variants of these sounds the source at the place of articulation is the primary source, but the rate of vibration of this source will not exceed 30 c/s and is thus generally much lower than the lowest possible fundamental frequency of samples of voiced speech. For purposes of phonetical classification the term *harmonic* should be restricted to sounds produced from a vocal cord sound source, thus excluding the unvoiced trills.

The terms *harmonic* or *periodic* are not adequate, from a strictly physical standpoint. Because of the variations always present it would be more appropriate to speak of voiced sounds as *quasi-periodic*. The term *voice* will be adopted here with reference both to a source category and a feature of the speech wave. The following source possibilities exist:

- A. No source (silence);
- B. Voice source only;
- C. Mixed voice source and noise sources;
- D. Noise source (one or several).

Practical use has been made of the two binary categories *voice/voiceless*, and *noise/noiseless* in the design of electrical speech synthesizers, e.g., the *Voder* and the *Vocoder* (Dudley, 1939; Dudley et al., 1939) from which all modern instruments for the production of synthetic speech derive; see Dudley (1956).

The term *noise source* refers to the primary acoustic disturbance within the vocal tract responsible for the generation of whispered sounds, aspirated sounds, fricated sounds, and exploded sounds. The source is *continuous* if the sound is *sustainable*, and *interrupted* if the shortness of the duration and the particular speed of onset and decay are crucial (Jakobson, Fant, Halle, 1952). Most noise sources are *turbulent*, which is a technical expression for the physical conditions under which the sound has been generated. The only non-turbulent noise sources are the sudden release of an over-pressure or the sudden checking of an air stream. The sources in this category will be called *transient* in agreement with standard electronical practice. A transient source is not identical with an interrupted turbulent source, but the former precedes the latter in the production of an unvoiced exploded stop sound, the turbulent noise source originating from random disturbances of the airflow at constricted passages and obstacles in the vocal tract, the transient source being contained in the impact on the vocal cavities of the pressure release at the instant of explosion.

Two types of *turbulent noise* sounds should be considered. One is the *fricative noise* produced under conditions of a relatively narrow constriction, in which case it is

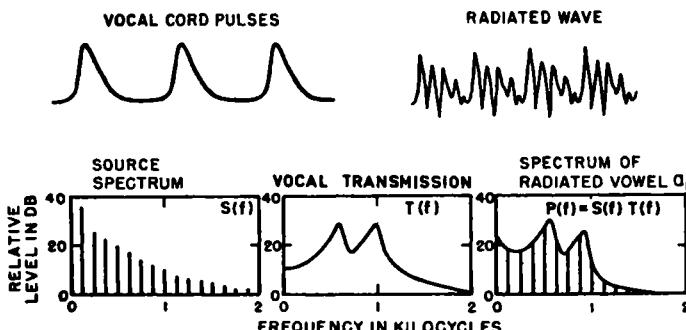


Fig. 1.1-2. Simplified source-filter decomposition of the spectrum of a two-formant voiced sound. The waveform of the periodic airflow through the glottis is converted into a harmonic spectrum $S(f)$ which multiplied by the filter characteristics $T(f)$ of vocal transmission provides the spectrum $P(f)$ of the radiated vowel which in turn may be specified by its waveform, as seen in the upper left of the figure.

essentially the cavities and parts of the vocal tract in front of and at the place of constriction that participate in the shaping of the sound. This fricative sound interval is shorter in a stop sound than in a corresponding fricative. The other type of sound is what could be called *open aspiration* and is often phonetically transcribed with an adscript ^h. It is produced with greater articulatory opening than members of the class of fricative sounds. The larger opening and the occurrence of more than one source within the vocal tract, e.g., an additional glottal noise source, contribute toward emphasizing the formants that depend on the entire vocal tract and not merely those of the front parts. In the following the term *aspiration* will be utilized in the specific *h-sound* sense and not to specify the entire sound produced by the outlet of an expiratory pulse.³ *Frication* and *aspiration* as here defined may occur simultaneously or in succession, or only one of the two noise categories may be chiefly present. In the noise interval of a stop sound, assuming an increasing degree of opening, the aspiration must come after the frication if both are present. The properties of voice and noise sources will be dealt with in more detail in Appendix A.1.

The analytical decomposition of speech production into the source and filter components will next be demonstrated by a simple example pertaining to a voiced sound. The glottis represents a high impedance termination of the vocal tract, and it is thus possible and convenient to define the voice source by the pulsating airflow through the glottis. This is a sawtooth-shaped periodic *time function* of the type indicated in the upper left of Fig. 1.1-2, which can, by means of the Fourier transform, be expressed as a harmonic spectrum, as indicated in the lower left part of the figure. The frequency selective transfer characteristics are introduced by the process of multiplying the amplitude of each harmonic $|S(f)|$ of the *source spectrum* by the value of the appropriate gain factor $|T(f)|$ of the *filter function* at the frequency f :

$$|P(f)| = |S(f)| \cdot |T(f)|. \quad (1.1-3)$$

³ Compare a similar treatment by Schatz (1954) and Fischer-Jørgensen (1956).

The phase of each harmonic is the sum of the phase of the corresponding source harmonic and the phase of the filter function, which relation can be written

$$\varphi_P(f) = \varphi_S(f) + \varphi_T(f). \quad (1.1-4)$$

This is a process of synthesis which can be materialized and controlled in any detail in a speaking machine. The spectrum envelope amplitude data $|P(f)|$ are a better reference for phonetic studies than the waveform of the sound, as studied from an oscillogram of the sound pressure $p(t)$. The technical object of spectrographic speech analysis is to derive the spectrum envelope function $P(f)$ from the speech wave $p(t)$, as picked up by a microphone. These methods are reviewed in *Appendix A.1*. The phase function $\phi_P(f)$ is generally neglected since it does not add any substantial information.

The spectral peaks of the sound spectrum $|P(f)|$ are called *formants*. Referring to *Fig. 1.1-2*, it may be seen that one such resonance has its counterpart in a frequency region of relatively effective transmission through the vocal tract. This selective property of $|T(f)|$ is independent of the source. The frequency location of a maximum in $|T(f)|$, i.e., the *resonance frequency*, is very close to the corresponding maximum in spectrum $P(f)$ of the complete sound. Conceptually these should be held apart but in most instances resonance frequency and formant frequency may be used synonymously. Thus, for technical applications dealing with voiced sounds it is profitable to define formant frequency as a property of $T(f)$.

The basic principle of the theory of voiced sounds is that, to a first order of approximation, the filter function is independent of the source. The formant peak will thus only accidentally coincide with the frequency of a harmonic. The formant frequencies can change only as a result of an articulatory change affecting the dimensions of the various parts of the vocal tract cavity system and thus the filter function. Conversely, but with the limitations implied by the concept of compensatory forms of articulation, the formant frequencies provide information about the position of the speaker's articulatory organs. If these formant frequencies are held constant and the fundamental frequency is raised one octave, the result is ideally that twice as many pulses per second are emitted from the voice organs. The distance between adjacent harmonics in the spectrum will be doubled, and the number of harmonics up to a certain fixed frequency limit will thus be halved. If a specific formant, for instance the first, comes close to the 6th harmonic at the lower pitch, it will be the 3rd harmonic that comes closest to the same formant in the case of the higher pitch. The concepts of formant frequency and harmonic number should not be confused.

Providing there is a definite tongue constriction within the vocal tract separating a back cavity from a front cavity, there exist primary but not sufficient conditions for describing the vocal tract network as a double resonator. Extensive use has been made of this model in the literature and it is generally proposed that the first formant is associated with the resonance of the back cavity and the second formant with the resonance of the front cavity. One of the purposes of the present publication is to

show the considerable limitations of this rule and to provide a sounder foundation for the physiological interpretation of the vocal resonances. As shown in *Chapter 1.4*, all parts of the vocal tract contribute to the determination of all formants but with varying degrees depending on the actual configuration. Compound tube or horn models should replace the Helmholtz resonator models of the vocal tract.

The intensity variations of a single harmonic or of a group of harmonics at a certain place within the frequency range depend both on the source and on the filter. As will be described in *Chapter 1.3*, the intensity level of a group of spectral components will decrease when the formants at lower frequencies are shifted to still lower frequency positions, or when any adjacent formant is shifted further away from the particular frequency region under observation.

The influence on the sound spectrum of the type of voice and the relative voice effort due to the source spectrum variations should also be considered. A reduction of voice effort, with a fixed location of all formants, leads to a decrease of the level of harmonics which is more prominent in the higher frequencies than in the lower part of the spectrum. This is due to a more steeply falling slope of the source spectrum envelope normally accompanying the lowering of the voice level.

Formants are labeled F_1 , F_2 , F_3 , etc., in the order they occur in the frequency scale. The notation F_1 , F_2 , F_3 , etc., refers to the frequencies of the formants or to the frequencies of the corresponding vocal tract resonances. In the simplified presentation of *Fig. 1.1-2* only the first two formants were included. F_3 and F_4 etc., are, however, always present, though with varying intensities. These higher formants are primarily of importance in front vowels.

Distances between formants in the frequency scale average 1000 c/s for males. This statistical average is physiologically correlated with the total length of the vocal tract. Because of shorter cavity lengths females thus have larger average formant spacings and higher average formant frequencies than males. Similar relations hold for children compared with adults; see the more detailed presentation of *Appendix A.1*.

Two speakers uttering the "same" vowel thus generally have somewhat different formant frequencies depending on their particular vocal tract dimensions. The spread of formant data may be specifically large if all possible contextual variants of a phoneme as well as all possible speaker categories are taken into account. However, in a particular context it is always to be expected that any speaker following the code of his language will produce phonemically different sounds by means of consistent distinctions in the formant pattern. This is the basis of the theory of distinctive features; Jakobson et al. (1952).

1.12 Segmentation

A basic problem in speech analysis is the degree of divisibility of the speech wave. The most common approach has been to start with the linguistic criteria in terms of a phonemic transcription and to impose this as a basis for division. By a systematic

comparison of the sound patterns of different contexts it is possible to make general statements as to what sound features are typical for a particular phoneme. Such studies are necessary, but in order to avoid ambiguities in the labeling of the successive observable sound units such an investigation should be preceded by an initial process of segmentation and description of the speech wave on the basis of its physical structure, and in terms of phonetic rather than phonemic units. There is need for extensive investigations of this type in order to establish an objective basis for dealing with phonetic problems. The book *Visible Speech*, by Potter, Kopp, and Green (1947), is a useful reference, but a more detailed and systematic mapping work is yet to come. This will be a laborious task but not an impractical one since only a limited number of pattern classes need to be recognized.

Detailed studies of this type lead to a description of the speech wave as a succession of sound units with fairly distinctly defined boundaries. These boundaries can conveniently be established in terms of the two basic categories of speech production, the source and the filter (Fant, 1952). Thus the boundary between a nasal continuant and a following oral continuant is determined by a change in the vocal tract filter system at the break of the oral closure. The transition from a voiced palatal continuant to a voiced vowel represents a more gradual filter change, as seen by the formant transitions. The boundary between a sound interval of aspiration and a following vowel produced with the same position of the articulators is mainly due to the change of source. The sequence of a dental fricative continuant to a vowel is characterized by a fairly abrupt change of the source and also of the filter system since the filtering is highly dependent on the position of the source. The spectrographic picture shows the high frequency area of random striations typical of the turbulent sound followed by the periodic fine-structured formant bars predominantly in the lower part of the spectrogram typical of the vowel. The step from silence to a following sound interval is a matter of source change only, provided the articulators do not vary significantly during the onset of the source, as they may do in stop sounds.

A purely acoustic segmentation of the speech wave thus results in a number of minimal sound units of the dimension of a speech sound, or smaller, that may be labeled according to their production. The number of such successive acoustic units of a speech utterance is generally greater than the number of signs in a phonetic or phonemic transcription. If this transcription is to be mapped onto the spectrographic record, the investigator will be forced to follow some convention for assigning each acoustic interval to a specific graphic sign. There is no harm in such a procedure provided all investigators have the appropriate means for observing the acoustic boundaries and follow the same conventions. Data on consonant duration in the literature should be regarded with some scepticism in this respect, especially if the data have been obtained from oscillographic or kymographic records only.

It is well known that a listener's identification of a particular phoneme is, in many instances, dependent on cues from successive acoustic sound intervals, that is, not only from the single interval or the several intervals included in the traditional

phonetic transcription but also from those conventionally assigned to following or preceding phonemes. The most common example of this is the significance of the formant transitions in the first part of a vowel as cues to the identity of the preceding consonant. Such transitional cues are of greatest importance for the distinction between various nasal consonants and also within the class of voiced stops. High energy fricatives, on the other hand, are less dependent on such cues.

The word [mama], for example, containing four speech sounds, may be dissected into four corresponding acoustic units. The same would be the case with the word [nana]. If a tape-recording of these words is divided into these appropriate segments and resliced with an exchange of all [m]-sound intervals for [n]-intervals there does not result a corresponding phonemic shift, as observed from listening tests. An interchange of the "vowels", on the other hand, does cause a shift in identification; see e.g., Malécot (1956).

As another example, the two words [pæt] and [kæt] may be chosen. The convention would be to set the boundaries of the vowel at those instances in time where the voicing starts and ends. If these test words are cut at the onset of voicing, and the initial [p] is exchanged for the initial [k], or vice versa, there will be a marked tendency of listeners to identify each of the resliced words according to the particular initial consonant segment it contains.⁴ This can be explained from an examination of the articulatory process. At the onset of voicing the articulators have moved so far towards the typical vowel position that the main part of the transition is completed. The first, and very important, part of the transition has taken place during the aspiration interval (Fischer-Jørgensen, 1954, 1956; Fant, 1957). Another reason for the commutability of the consonant noises in the above example is their relatively high energy.

Segmentation—with reference to stop consonants into explosion, frication, and aspiration—as mentioned in the last paragraph of *Section 1.11*, cannot always be derived from the spectrographic data. It is often hard to separate explosion from frication, since the former is very short in duration. The duration of the explosion is that of the transient response of the vocal tract. It is thus of a duration of the order of the inverse of the bandwidth of the particular formant to be studied and therefore insignificantly small in the high frequency fricative area of a dental stop, for example. The frication is more apparent in palatal and dental stops than in labials, and the boundary towards the aspiration is seldom sharp. Aspiration as distinct from frication may be recognized by a reduction of the intensity in the high frequency region above 4000 c/s and the appearance of formants that have a continuity with those of the following vowel. The first formant is generally very weak in aspirated sound intervals except in combinations with some back vowels.

An acoustic segmentation of the word [bæt] cannot rely on the onset of voice only, since the voice source may be active during the whole initial [b]. The introductory *voice bar* at the bottom of the spectrogram, if present, signals the presence

⁴ Experiments performed by H. M. Truby.

of vocal cord vibrations during the loading up of the oral cavities to the state of over-pressure before release. In Russian this voice bar is a necessary attribute of the voiced stops. In Swedish and in English the onset of the voice source may come only at the explosion or very soon after. In the absence of a voice bar, that is, when the /p/-/b/ distinction is primarily of the fortis/lenis type, the duration of the time interval from the explosion to the first vocal cord vibration of the following vowel is longer for the [p] than for the [b]. The latter sound lacks the aspiration and has generally a very weak fricative interval. It is thus apparent that the first part of the formant transitions is carried by a voice source in [b], and by the aspirative noise source in [p]. From an articulatory point of view the vowel in [pæt] starts with the aspirative interval of the [p].

The similarity in the production of stops and nasal consonants is apparent when, for instance, an initial [b] with voiced occlusion is compared with an initial [m]. The sound transmission to the outer air from the vocal tract during the sound interval of complete oral closure, is through the vibrating walls of the cavities in the case of the stop sound, and through the nose in the interval of nasal murmur, the latter path of transmission evidently being the more efficient one. The release of the lip barrier is, however, associated with a release of an overpressure that is stronger in the case of the stop sound and contributes to the greater intensity contrast between the closed and open sound intervals of the stop, as compared with the nasal.

1.13 The F-pattern

A maximally complete source-filter decomposition of the speech wave data implies a detailed mapping of the speaker's vocal tract during the speech process so that its transmission properties, i.e., filter function may be specified. The source characteristics of a short sample of the speech wave are then derived by subtraction, on a decibel scale, of the filter function from the spectrum of the sample in conformity with Eq. 1.1-1. Such an analysis has been applied to fricatives and stops; see *Chapter 2.6*.

For the more general purpose of making the data from acoustic speech analysis maximally useful for phonetic purposes it is possible to perform a partial and approximate reconstruction of the source and of the filter function without the vocal tract mapping. To some extent this simply involves a physiological interpretation of spectrograms by the investigator based on a general knowledge of speech production.

The fine structure of the sound or smaller part of the sound to be studied reveals the type of source, whether it is a single transient or continuant, containing noise or no noise, and voice or no voice, and the particular combination of those features.

The fine structure may be removed from the intensity-versus-frequency curve of a sample by drawing a *spectrum envelope* smoothly combining the successive harmonics. This envelope should reveal the *formant structure*, or, in other words, the topology of formant peaks. If possible, it should also retain major dips in the spectrum corresponding to anti-resonance effects. The spectrum envelope may, for practical

purposes, be *smoothed*, so that adjacent formant peaks that need not be considered in isolation may be treated as part of a single formant, or, rather, as a formant group. This would be the case for the main formant of a dental stop or fricative at frequencies above 4000 c/s, the detailed formant structure of which appears to be of less importance. Similarly, for certain perceptual studies, it may be desirable to refer to the second and higher formants of a front vowel as a single formant group.

Those formants that are substantially derived from the oral part of the vocal tract, i.e., the pharynx and mouth, are of specific interest in speech analysis. There are three main reasons:

- 1) The predictability of the filter function and thus of the major spectral shape of a vowel spectrum from formant frequencies only (Fant, 1956). This is an instance of source-filter decomposition without articulatory mapping.
- 2) The possibility of inferring the articulation of any sound given the evidence of the spectrum envelope and in particular the frequencies of those formants that have a continuity with the formants of an adjacent vowel.
- 3) The importance of transitional cues for the perception of speech. These cues are contained in the formant frequency variations in intervals of the speech wave adjacent to a consonant.

There is thus a motivation to attempt a generalization of the *hub* or *locus* concepts by the introduction of the following term:

F-pattern. The F-pattern, at any instant of time, is defined as the resonance frequencies of the oral part of the vocal tract or those resonance frequencies that show a continuity with the oral resonances of an adjacent sound. The F-pattern coincides closely with the observable formant peak frequencies of sounds produced from a glottal, preferably voiced, source. For this reason it is advisable to choose a method of formant frequency determination that provides an identity of resonance frequency and of formant frequency. The F-pattern thus comprises F_1 , F_2 , F_3 , F_4 , etc. Each of these frequencies may be referred to by the term *locus* or *position* pertaining to the frequency scale. The F_2 -locus is identical with Potter, Kopp, and Green's (1947) term *hub* defined as the *visible or hidden position of the second formant within a spectrogram*.⁵ The materialization of the F-pattern presupposes the existence of a source, not necessarily a glottal source. Because of the continuities and limited range of articulatory movements, the F-pattern is continuous throughout a speech utterance, and it may therefore be of some interest to follow it continuously within a speech record, even at the silent intervals, by a process of interpolation.

The transitional characteristics of consonant-vowel and vowel-consonant combi-

⁵ The restriction, *when the sound is made alone*, was added to this definition referring to the possibility of sustaining the particular articulation. It is suggested here that F-locus be reserved for F-patterns of specific sounds, primarily for the limiting positions of the articulators, e.g., at the state of maximum closure for consonants, and that the term F-position be utilized for any arbitrary interval of connected speech. Stevens and House (1956) utilize the term locus according to this articulatory principle. The Haskins group (Delattre et al., 1955) associates the term locus mainly with empirically determined rules for preferred formant transitions of their synthesis patterns.

nations may thus be specified by the F-pattern at successive points in the time scale, the F-pattern being a fairly accurate and concise acoustic correlate to articulation in the more precise sense of vocal tract configuration. A formant transition may be rising or falling with regard to the direction of its frequency variation. An alternative terminology is to call a transition of a formant $n = 1, 2, 3, \dots$ etc., positive or negative according as the F_n -locus of a consonant is higher or lower than the frequency F_n of a following or a preceding sound (Halle et al., 1957).

It is well known—see, e.g., Delattre (1951); Stevens and House (1956); Fant (1957)—that a low F_1 signals articulatory closure, a very low F_2 signals a retracted articulation combined with lip closure, a very high F_2 signals a palatal position of the tongue and a high F_3 indicates a prepalatal or dental articulation. Dentals have a medium high F_2 -position, and the F-pattern of labials varies with tongue position. The differential effect of a lip closure is a lowering of all loci in varying amounts. A very low F_3 signals retroflex modification. Relations between the F-pattern and articulation will be discussed in more detail in *Section 1.43* and throughout *Part II*.

When extracting data from spectrographic studies of speech for the purpose of phonetic descriptions the general rule is thus: *the vowel spectrum is sufficiently specified by the F-pattern*, whereas *the consonant spectrum should in addition be specified by its spectrum envelope*. As previously stated, the main shape of a vowel spectrum envelope can be computed from a knowledge of the formant frequencies, i.e., from the F-pattern. This is true of most sounds produced with a glottal source. Some restrictions occur for lateral sounds and nasalized vowels. The visible formants in the spectrum of the nasal murmur are only partially equivalent to the F-formants, i.e., those of the oral pathways. In a fricative consonant it is true that some of the subpeaks within a formant region or some isolated relatively weak formants reflect a part of the F-pattern, but these are not sufficient for a reconstruction of the entire consonant spectrum.

1.2 NETWORK THEORY OF VOCAL TRANSMISSION

1.21 Network Representation of Acoustic Resonators and Horns

The mathematical treatment of the speech production process involves the following successive operations. The first one is the mapping of the vocal cavities in terms of an area function describing the cross-sectional area perpendicular to the air stream from the glottis to the radiating surface at the lips. Secondly, this area function has to be approximated by a sufficiently small number of successive parts, each of a constant cross-sectional area. The transmission properties of this system are next calculated and added to the assumed frequency characteristics of the source. The last step is to perform a maximally concise presentation of the results by converting the calculated frequency characteristics into a set of poles and zeros. When dealing with voiced sounds the formant frequencies are of primary interest. In addition, or as a further attempt to remove the physical redundancy, the main shape of the calculated spectrum can be discussed, as in the previous chapter.

The network representation of the vocal tract is based on the analogous acoustic impedance which is the ratio of pressure to volume velocity. It should not be confused with the specific impedance which is the ratio of pressure to particle velocity or with the concept of mechanical impedance which is the ratio of force to particle velocity. The advantage of adopting the analogous impedance lies in the continuity of both volume velocity and pressure along the resonator network.

The theory of simple Helmholtz resonators is, as will be discussed in detail in *Section 2.33*, only partially applicable to the study of vocal transmission because of the relatively large dimensions involved. In a more general treatment the acoustic resonator system should be divided into small sections by means of cuts perpendicular to the direction of the wave propagation. One such section of a resonator system, specified by the cross-sectional area A and length l , can be represented by a series *inductance* $L = \rho l/A$ followed by a shunt *capacitance* $C = lA/\rho c^2$. In addition, a *resistance* R in series with L and a *conductance element* G paralleling C should be included. This equivalent circuit may be used as a fair approximation for sections

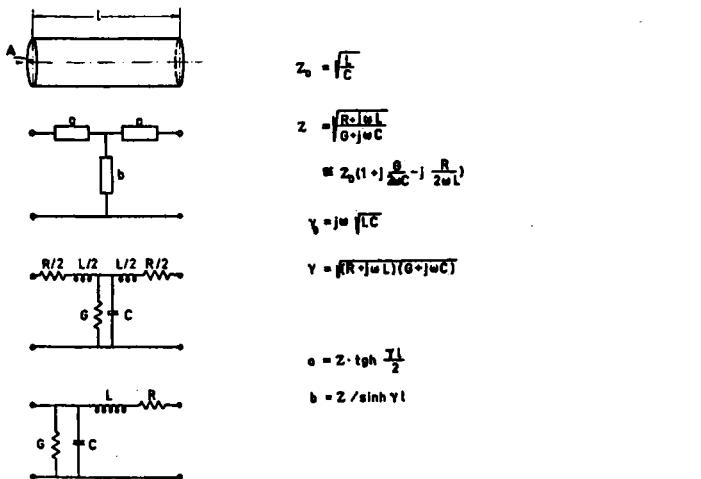


Fig. 1.2-1. The T-network equivalent of a single-tube section and lumped element approximations valid for large frequencies or short lengths.

of finite length up to a frequency where the wavelength is larger than $8l$. In acoustic systems there is also a critical upper frequency where wave propagation in radial direction sets in. The corresponding wavelength is of the order of twice the largest cross-dimension. In general, assuming relatively small losses through the resistances R and $1/G$, the resonance frequencies of an arbitrarily complex resonator system are determined by the L and C elements alone, and the R and G elements or rather the R/L and G/C ratios will determine the resonance bandwidths.

When the length of the section is not very small, the accuracy of representation is improved by placing half of the inductance symmetrically around the capacitance. A homogeneous section of constant cross-sectional area A and arbitrary length l can be specified in terms of an equivalent circuit or the T-circuit indicated at the top of Fig. 1.2-1 containing the elements

$$\begin{aligned} a &= Z \operatorname{tgh} \frac{\Gamma}{2}, \text{ and} \\ b &= Z / \sinh \Gamma. \end{aligned} \quad (1.21-1)$$

The characteristic impedance Z and the transfer constant Γ are related to the distributed elements L , C , R , G per unit length by the classical expressions:

$$\begin{cases} Z = \sqrt{\frac{R + j\omega L}{G + j\omega C}}, \text{ and} \\ \Gamma = l\sqrt{(R + j\omega L)(G + j\omega C)} = l(a + j\beta) = l\gamma, \end{cases} \quad (1.21-2)$$

where γ is the complex propagation constant containing the attenuation constant a and the phase constant β .

Assuming small losses

$$\begin{cases} Z = Z_0[1 - j\alpha_R/\beta + j\alpha_G/\beta], \text{ and} \\ \gamma = \alpha_R + \alpha_G + j\beta, \end{cases} \quad (1.21-3)$$

where

$$\begin{cases} Z_0 = \sqrt{L/C}, \\ \beta = \omega\sqrt{LC}, \\ \alpha_R = R/2Z_0, \text{ and} \\ \alpha_G = GZ_0/2. \end{cases} \quad (1.21-4)$$

Since the inductance L and capacitance C per unit length of the acoustical transmission line are

$$\begin{aligned} L &= \rho/A, \text{ and} \\ C &= A/\rho c^2, \end{aligned} \quad (1.21-5)$$

Eq. 1.21-4 can be rewritten:

$$\begin{cases} Z_0 = \rho c/A, \\ \beta = \omega/c, \\ \alpha_R = \frac{1}{c} \cdot \frac{R}{2L} = \frac{RA}{2\rho c}, \text{ and} \\ \alpha_G = \frac{1}{c} \cdot \frac{G}{2C} = \frac{G\rho c}{2A}. \end{cases} \quad (1.21-6)$$

If losses are neglected, the uniform resonator section is sufficiently specified by its length and cross-sectional area. This relation holds approximately for any shape of the cross-section and thus not only for cylindrical resonator sections.

Numerical calculations of vocal transmission performed without the aid of analog or digital computers may be rather complicated. It therefore pays to start out with an optimally simple cavity approximation that is sufficiently representative for the scope of the investigation.

One possible means of simplification is to utilize the theory of horns for the representation of a larger part of the vocal tract where the area changes continuously. The acoustic theory of horns, as presented by Morse (1948), provides the physical foundation for such calculations. However, in addition to the wave equation an equivalent network representation is necessary for the incorporation of a horn as a part of a compound cavity system. Laurent (1940, 1956) has developed equivalent circuits for continuously inhomogeneous transmission lines of a type that exactly corresponds to the class of horns treated by Morse.

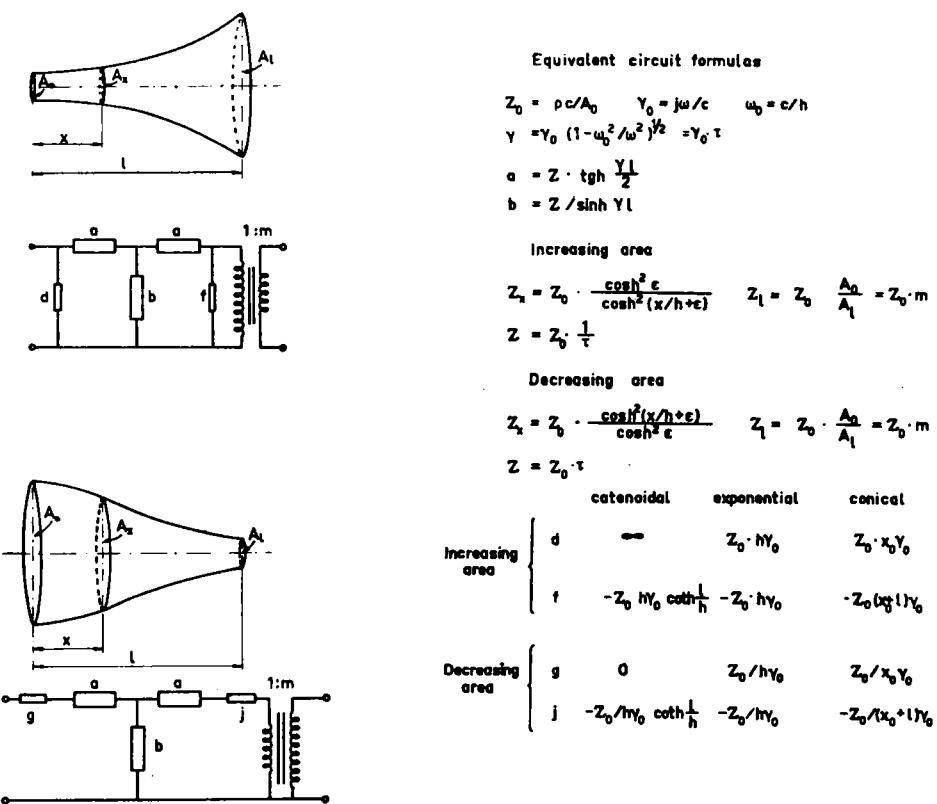


Fig. 1.2-2. Horn resonators and their equivalent networks, derived by Laurent (1940, 1956). They can be used for a complete calculation of conical, exponential, catenoidal, and other types of horns as a part of a one-dimensional wave propagation system.

The general expression for the horn area $A(x)$ as a function of the axial coordinate x is

$$A(x) = \frac{A_0}{\cosh^2 \epsilon} \cosh^2 \left(\frac{x}{h} + \epsilon \right), \quad (1.21-7)$$

which implies an increasing area. Conversely, the decreasing area function has the form

$$A(x) = \frac{A_0 \cosh^2 \epsilon}{\cosh^2 \left(\frac{x}{h} + \epsilon \right)} \quad (1.21-8)$$

When $\epsilon = 0$ the horn is catenoidal; when $\epsilon = \infty$ the horn is exponential; and when $\epsilon = \frac{x_0}{h} - \frac{j\pi}{2}$ with h approaching infinity, the horn is conical. The constants can, of course, be given intermediate values corresponding to horns of intermediate shape.

In terms of these constants, as chosen by Morse, the elements of the equivalent circuits, Fig. 1.2-2 as derived by Laurent, can be specified as follows:

<i>Resonator length</i>	<i>l</i>
<i>Nominal characteristic impedance</i>	$Z_0 = \rho c / A_0$
<i>Nominal propagation constant</i>	$\gamma_0 = \alpha + j\omega/c$
<i>Horn cutoff frequency</i>	$\omega_0 = c/h$
<i>Propagation constant</i>	$\gamma = \gamma_0(1 - \omega_0^2/\omega^2)^{1/2} = \gamma_0 \cdot \tau$
<i>Transfer constant</i>	$\Gamma = l \cdot \gamma$
<i>Characteristic impedance</i>	$\begin{cases} Z = Z_0/\tau & (\text{increasing area}) \\ Z = Z_0 \cdot \tau & (\text{decreasing area}) \end{cases}$
<i>Series element of T-network</i>	$a = Z \operatorname{tgh} \frac{\gamma l}{2}$
<i>Parallel element</i>	$b = Z / \sinh \gamma l$
<i>Transformer impedance ratio</i>	$m^2 = A_0 / A_l$

The additional series or parallel elements of the equivalent network are

$$d = Z_0 h \gamma_0 \coth \frac{l}{h},$$

increasing area

$$f = -Z_0 h \gamma_0 \coth \left(\frac{l}{h} + \epsilon \right),$$

$$g = \frac{Z_0}{h \gamma_0} \coth \frac{l}{h}, \text{ and}$$

decreasing area

$$j = -\frac{Z_0}{h \gamma_0} \coth \left(\frac{l}{h} + \epsilon \right).$$

For the special case of catenoidal, exponential, and conical horns:

	Increasing area		Decreasing area	
	<i>d</i>	<i>f</i>	<i>g</i>	<i>j</i>
<i>Catenoidal horn</i>	∞	$-Z_0 h \gamma_0 \coth \frac{l}{h}$	0	$\frac{-Z_0}{h \gamma_0} \coth \frac{l}{h}$
<i>Exponential horn</i>	$Z_0 h \gamma_0$	$-Z_0 h \gamma_0$	$Z_0 / h \gamma_0$	$-Z_0 / h \gamma_0$
<i>Conical horn</i>	$Z_0 x_0 \gamma_0$	$-Z_0 (x_0 + l) \gamma_0$	$Z_0 / x_0 \gamma_0$	$-Z_0 / (x_0 + l) \gamma_0$

The catenoidal horn is a useful model since at least one of the two additional impedance elements of the equivalent network vanishes. The rate of flare is zero at $x = 0$ which makes continuous interconnections possible. Thus the whole mouth cavity including a palatal constriction and front and back parts as in the production of the vowels [i] and [e] may be represented by a single network, since the parts in front of and behind the narrowest point can then be regarded as perfectly matched. The same is true of a mouth cavity of concave area function, i.e., a part of the vocal tract containing an area maximum. In back vowels the whole pharynx can be represented by a convex area function.

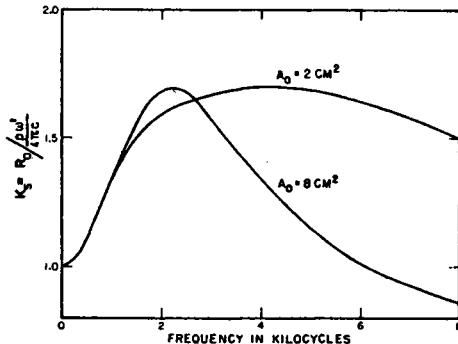


Fig. 1.2-3. The radiation resistance factor $K_S = R_0 / \frac{\rho \omega^2}{4\pi c}$ for the frequency range of primary interest for the calculation of vocal tract transmission. At very low frequencies the baffle effect is negligible and the radiation resistance is one-half the value for the infinite baffle case. At higher frequencies the combination of a spherical baffle of 9 cm radius and a radiating area of the values indicated in the figure account for an increase of radiation resistance towards, but not fully approaching, the infinite baffle value.

The theoretical treatment above takes care of the complete one-dimensional wave propagation. It should be stressed that the reflections at the mouth-opening of a speaker are very great and that the simplified theory of horn loudspeakers, based on the outgoing wave only, are not applicable. At first glance, the transformer in the equivalent circuit would indicate that the vocal tract power output should be proportional to the radiating area at the lips. This relation holds, however, only at frequencies well above the cutoff frequency of the horn and in general only for the spectrum level between formants; see further the discussion in *Section 1.42*. At frequencies which are high in comparison with the cutoff frequency, the equivalent circuit equals that of a tube of the same length as the horn and with a constant cross-sectional area equal to that of the throat of the horn. This line equivalent is terminated by the ideal transformer. At lower frequencies the effect of the transformer is counteracted by the impedance of the circuit elements d and f .

Ideal hard-walled tubes are not loss free. There are frictional losses in a boundary layer at the surface of the tube which can be taken into account by the series resistance

$$R = \frac{S}{2A^2} (2\mu\omega\rho)^{\frac{1}{2}}, \quad (1.21-9)$$

where $\mu = 1.84 \cdot 10^{-4} \text{ g/cm sec}$ is the coefficient of viscosity. In the case of tubes of circular cross-sectional area, the ratio of area A to circumference S equals half the radius r and the resistance is

$$R = \frac{1}{Ar} (2\mu\omega\rho)^{\frac{1}{2}}. \quad (1.21-10)$$

The viscous boundary layer also has the effect of reducing the apparent area for calculating the inductance L per unit length. Thus

$$L = \frac{\rho}{A} \left[1 + \frac{S}{A} (\mu/2\omega\rho)^{\frac{1}{2}} \right]. \quad (1.21-11)$$

The correction term is negligible for vocal tract calculations.*

In several acoustic textbooks, e.g., Mason (1948); Trendelenburg (1950), the heat conduction losses at the cavity walls are taken into account as an apparent increase in the coefficient of viscosity. This is an incorrect representation of the equivalent circuit. The frictional losses are proportional to the square of the current, and the heat conduction losses are proportional to the voltage squared. The latter should thus be represented by a conductance element G as follows:

$$G = S(\kappa-1) \frac{\beta}{\varrho c} (K_h/2\omega C_p \rho)^{\frac{1}{2}}, \quad (1.21-12)$$

where

K_h = coefficient of heat conduction of air,

C_p = specific heat at constant pressure of air,

C_v = specific heat at constant volume of air,

$\kappa = C_p/C_v$, and

$\beta = \dot{\omega}/c$.

Under normal atmospheric conditions (1 atmos, 20°C) the following numerical expressions are obtained

$$\begin{aligned} R &= \frac{1.66 \cdot 10^{-3} f^{\frac{1}{2}}}{A(2A/S)}, \\ G &= 2.2 \cdot 10^{-7} \cdot S \cdot f^{\frac{1}{2}} \\ a_R &= \frac{2.01 \cdot 10^{-5} f^{\frac{1}{2}}}{2A/S}, \\ a_G &= \frac{0.91 \cdot 10^{-5} f^{\frac{1}{2}}}{2A/S}. \end{aligned} \quad (1.21-13)$$

It can thus be seen that heat conduction cannot be neglected since its contribution to the attenuation constant is of the order of one-half that from the frictional losses. The effect of heat conduction in the vocal cavities is probably small compared with the damping effect of energy losses through the vibrations of the cavity walls. These can be included in the theory by means of a shunt impedance, which may conveniently be thought of as composed of a large inductance, parallel with a high resistance. The inverse of this resistance is the loss conductance G_s . According to van den Berg

* A similar correction for the capacitance with respect to heat conduction losses is given by Benade (1968): On the propagation of sound waves in a cylindrical conduit, *J. Acoust. Soc. Am.*, 44, 616-623 (1968).

(1953), the pharynx walls possess a positive reactance at frequencies of interest and contribute significantly to the damping of the first formant; see *Sections A.36-A.4*, where also similar calculations by Stevens (1958) are reviewed. The effect of the reactive part of vocal tract cavity wall impedances on the tuning of the resonance frequencies is small, except during conditions of complete or almost complete closure at a tongue constriction or at the lips (see *Section 2.34-B*).

Irrespective of the physical mechanism responsible for energy loss at the cavity walls, it can be seen that the attenuation constant for an arbitrarily shaped section of fixed cross-sectional area, is proportional to the circumference. This is true of both a_R and a_G . The nasal cavities have an especially large shape factor $S/2(A\pi)^{\frac{1}{2}}$, owing to their relatively complex configuration. The division into two separate passages contributes to make the total surface large compared with the volume. A narrow tongue constriction also deviates substantially from the circular cross-sectional shape. Assuming an elliptical shape and a height of one-ninth of the width, the shape factor comes close to 2. By definition the shape factor of a circular section equals 1.

Calculations have shown that the particle velocities in narrow constrictions can be large enough to cause a considerable non-linear increase in the frictional resistance at low frequencies. Calculations have been performed by Ingård (1953) who expresses the total resistance of a narrow aperture that terminates abruptly into the adjacent cavities by the formula

$$R = \frac{1.66 \cdot 10^{-3} f^{\frac{1}{2}}}{Ar} [1 + 2r + 2r \cdot 0.7(v/100)^{1.7}], \quad (1.21-14)$$

where r is the radius of the aperture. The second term within the brackets is the linear end correction to the aperture resistance which contains the sum of the inner and outer end corrections, and the third term represents the non-linear increase in this end correction. The latter term is independent of the neck length as long as the particle displacement v/ω is small compared with this length.

The applicability of this formula to the vocal tract is not very useful since there are seldom sharp discontinuities in area except at the teeth or at the upper edges of the vocal cords.

In the production of any continuant voiced or unvoiced speech sound there is a flow of air superimposed on the formants or other spectral components constituting the signal structure. As is shown in *Section A.21*, the flow of air through a narrow constriction under turbulent conditions gives rise to a flow dependent resistance of the order of

$$R_{diff} = 2R_{flow} = \rho v / A, \quad (1.21-15)$$

where A is the cross-sectional area and v the particle velocity.

The radiation impedance terminates the vocal tract filter. It contains a resistive part in which the radiated energy is consumed in series with a reactance representing the effective mass of vibrating air at the lips.

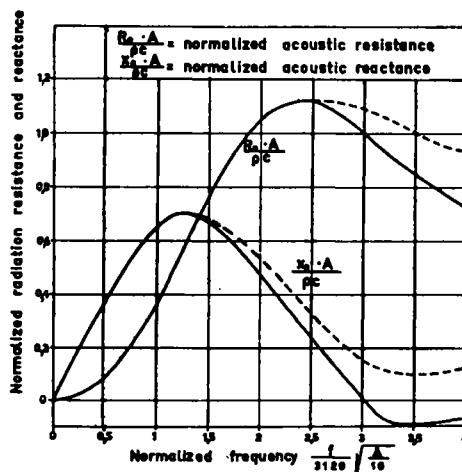


Fig. 1.2-4. Broken line curve — normalized radiation resistance and inductance for a small circular baffle on the surface of a sphere of radius 9 cm. Solid lines pertain to the approximation utilized in the design of LEA.

The formulas, given by Morse (1948) for a radiating circular baffle on the surface of a sphere of radius $a = (A_0/\pi)^{1/4} = 9 \text{ cm}$ have been adopted for the calculations. At frequencies where $\omega a/c < 2$, i.e., $f < 6000(10/A_0)^{1/4}$, which is the frequency range of interest for the calculation of vocal transmission, the radiation resistance can be expressed as

$$R_0 = \frac{\rho\omega^2}{4\pi c} K_S(\omega), \quad (1.21-16)$$

where A_0 is the radiating area. The function $K_S(\omega)$ is a frequency dependent factor related to the baffle effect of the head, as shown in Fig. 1.2-3. At very low frequencies $K_S = 1$, as would be expected for a simple point source. At higher frequencies of the order of 2000 c/s, K_S approaches a maximum value of 1.7 and then decreases. Earlier calculations on vocal transmission performed by Chiba and Kajiyama (1941); Dunn (1950); and van den Berg (1953) were based on the application of the infinite baffle theory, which corresponds to a value of $K_S = 2$ in Eq. 1.21-16.

The radiation resistance is apparently almost independent of the radiating area up to a frequency of 3000 c/s. At higher frequencies a large area provides a smaller radiation resistance than a small radiating area. The full extent of the radiation resistance versus frequency curve is shown by the broken line in Fig. 1.2-4. The ω^2 dependency extends up to about $\omega(A_0/\pi)^{1/4}/c = 1.85$, i.e., a frequency of 6500 c/s at the comparatively large mouth-opening $A_0 = 8 \text{ cm}^2$. At higher frequencies the radiation resistance approximately matches the characteristic impedance $\rho c/A_0$ at the terminating section of the vocal tract.

The solid line curves in Fig. 1.2-4 pertain to the resistance and reactance of the

radiation impedance filter section utilized in the Swedish electrical line analog *LEA*. The accuracy is good within the frequency range of interest, as can be seen.

The low frequency approximation to the radiation reactance can be expressed as the end correction length

$$l_0 = 0.8(A_0/\pi)^{\frac{1}{2}}. \quad (1.21-17)$$

Some radiation takes place at any abrupt transition from a smaller area A_0 to a larger cross-sectional area A in a compound cavity system. The internal end correction length corresponding to this additional reactance in the system is according to Ingård (1953),

$$l_i = 0.48(A)^{\frac{1}{2}}[1 - 1.25(A_0/A)^{\frac{1}{2}}], \quad (1.21-18)$$

provided the aperture area A_0 is smaller than 0.16 A .

This additional end correction can be of some importance as a contribution to the inductance of the opening between the teeth. It also provides a small contribution to the glottis inductance and to the outlet of the larynx tube.

The accuracy requirements for the estimation of the radiation inductance cannot be set very high since there is some uncertainty as to the true location of the radiating surface at the lips. This surface is not confined to a single plane and probably follows the curvature of the lips and teeth to some extent. It cannot be dependent on the degree of labialization only, and it is probably frequency dependent and conditioned by the shape of the front cavity. It has been assumed in the present calculations that the radiation takes place at a plane located not further than 0.5 cm in front of the front teeth. This rule was adopted for all unrounded vowels and consonants.

1.22 *Methods of Numerical Calculations*

Once the vocal tract has been specified in terms of a lattice network as in *Fig. 1.2-5*, the calculations can proceed according to standard methods in circuit theory.

For voiced sounds the vocal transmission is defined by a current transfer ratio U_0/U_q , where U_0 is the output volume velocity at the lips and U_q is the volume velocity supplied by the voice source. When dealing with sound sources higher up in the vocal tract, the transmission is defined by the ratio U_0/E_v , where E_v is the pressure of the source. The source E_v is regarded as a series element inserted in the series branch connecting sections v and $v-1$ of the network.¹ The radiation impedance is denoted by d_0 . Assuming a lattice of symmetrical T-structure as in *Fig. 1.4-5* with a series element a_n and shunt elements b_n and $a_n + b_n = d_n$, the following formulas can be derived:

¹ This is a point of major systematic interest. A source voltage E paralleling the line is inconceivable since it would short-circuit the system and thus upset the impedance structure. A parallel constant current source is a third alternative. It would preserve the impedance structure but cause a different set of zeros than the series source and may be outruled on this basis. The series source provides calculated data in conformity with the structure of measured speech spectra.

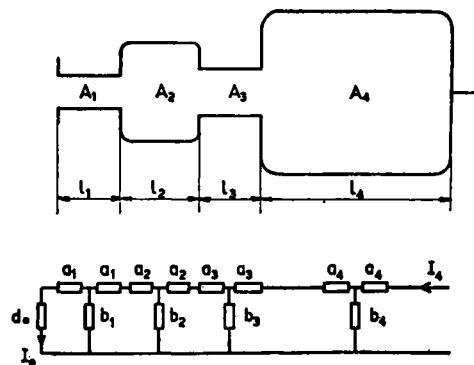


Fig. 1.2-5. Generalized network of a 4-section resonator. Each of the four sections of the cavity system is represented by a complete T-network equivalent.

$$U_0/U_q = b_q \Delta_{0q}/\Delta, \text{ and} \quad (1.22-1)$$

$$U_0/E_\nu = \Delta_{0\nu}/\Delta, \quad (1.22-2)$$

where Δ is the principal determinant and Δ_{0q} and $\Delta_{0\nu}$ are co-factors.

The poles of the vocal tract transfer function enter via the principal determinant, and possible zeros originate from the co-factor. The factor $b_q \Delta_{0q}$ is equal to the product of all shunt elements and can be brought into the principal determinant. No zeros occur in a transfer ratio, i.e., as a result of a voice source. The zeros of $\Delta_{0\nu}$ coincide with the condition of infinite impedance seen from the source towards the glottis end.

This procedure of calculation will next be exemplified by analysis of the 4-section system of Fig. 1.2-5. Besides the glottis source U_4 a source E_3 is inserted in the connection between sections 2 and 3.

loop nr

$$1 \quad (d_0 + d_1)U_0 - b_1 U_1 = 0 \quad (1.22-3a)$$

$$2 \quad -b_1 U_0 + (d_1 + d_2)U_1 - b_2 U_2 = 0 \quad (1.22-3b)$$

$$3 \quad -b_2 U_1 + (d_2 + d_3)U_2 - b_3 U_3 = E_3 \quad (1.22-3c)$$

$$4 \quad -b_3 U_2 + (d_3 + d_4)U_3 = b_4 U_4 \quad (1.22-3d)$$

$$\Delta = \begin{vmatrix} d_0 + d_1 & -b_1 & — & — \\ -b_1 & d_1 + d_2 & -b_2 & — \\ — & -b_2 & d_2 + d_3 & -b_3 \\ — & — & -b_3 & d_3 + d_4 \end{vmatrix} \quad (1.22-4)$$

$$\Delta_{0q} = \Delta_{04} = - \begin{vmatrix} -b_1 & — & — \\ d_1 + d_2 & -b_2 & — \\ -b_2 & d_2 + d_3 & -b_3 \end{vmatrix} = b_1 b_2 b_3, \quad (1.22-5)$$

$$\frac{U_4}{U_0} = \begin{Bmatrix} (d_0+d_1)/b_1 & -b_1/b_2 \\ -1 & (d_1+d_2)b_2 \\ \text{---} & -1 \\ \text{---} & \text{---} \end{Bmatrix} \quad (1.22-6)$$

$$\Delta_{03} = \begin{Bmatrix} -b_1 & \text{---} & \text{---} \\ d_1+d_2 & -b_2 & \text{---} \\ \text{---} & -b_3 & d_3+d_4 \end{Bmatrix} = b_1 b_2 (d_3 + d_4). \quad (1.22-7)$$

The zeros occur when $d_3 + d_4 = 0$.

A 4-section resonator was used by Dunn (1950) for calculation of the spectrum envelope of vowels neglecting losses and with the further approximation of sections 1 and 3 as being pure inductances. Calculations on 4-section resonators taking into account losses have earlier been performed by the author (1950a) and by van den Berg (1953). The lumped element approximation of sections 1 and 3 used in all these studies gives rise to considerable errors at frequencies above the 3rd formant. The complete analysis of the 4-tube resonator in terms of hyperbolic functions is given in *Section A.34*.

If losses are disregarded in the calculations, it will be necessary to add representative values of formant bandwidths to the calculated data. The determinantal expressions *Eq. 1.22-4,5,6,7* are rather simple to handle in the loss free case, even if a large number of sections are used. The incorporation of every added section involves two additional multiplications and one addition. If the formant frequencies are the main objectives of an investigation, it is equally simple or perhaps even simpler to evaluate the impedance, as seen from the glottis termination of the network. The formants occur at the frequencies where the reactance is infinite. Disregarding losses,

$$Z_{tq}(\omega) = jZ_q \operatorname{tg} \left[\frac{\omega l_q}{c} + \operatorname{artg} \left[\frac{A_q}{A_{q-1}} \operatorname{tg} \left(\frac{\omega l_{q-1}}{c} + \dots \right. \right. \right. \\ \left. \left. \left. + \frac{A_3}{A_2} \operatorname{tg} \left\{ \frac{\omega l_2}{c} + \operatorname{artg} \left[\frac{A_2}{A_1} \operatorname{tg} \left(\frac{\omega l_1}{c} + \operatorname{artg} \frac{Z_0}{Z_1} \right) \right] \right\} \dots \right) \right] \right] \quad (1.22-8)$$

With the aid of a nomogram of the function $\operatorname{artg}(mtg\varphi)$, each recurrent step of the calculation is reduced to two operations. This is a method that has been utilized for the calculation of the resonant modes of compound electrical cavity resonators.

It is equally feasible to calculate the input impedance at the radiating end. The formants then occur at frequencies where this impedance, Z_{t0} , is zero. It is also possible to break up the lattice network at any nodal point and determine the formant frequencies from the condition that the sum of the reactances to the left and the right of the point of observation shall equal zero.

In general, when the resistive elements of the network are taken into account, the current transfer ratio will be complex:

$$\begin{aligned} U_q/U_0 &= 1/H(\omega) = N(\omega) = N_b(\omega) + jN_a(\omega), \text{ and} \\ \varphi(\omega) &= \operatorname{arg} N_a(\omega)/N_b(\omega). \end{aligned} \quad (1.22-9)$$

If a closed form expression for U_q/U_0 can be retained, the poles of U_0/U_q are found by substituting $j\omega$ for the complex frequency variable $s = \sigma + j\omega$ and equating $N_b(s)$ to zero. This method becomes, however, very impractical for network structures exceeding two or three loops. When numerical calculations are carried out, the roots have to be evaluated by approximate methods. One obvious method is to calculate the absolute value, $|U_0/U_q|$, at sufficiently close frequency intervals so that the peaks of the resonance curves and their bandwidths (3 dB below peaks) can be determined graphically. These measures of resonance frequencies F_n and resonance bandwidths B_n may be adopted as an approximation to the angular frequencies ω_n and the damping constants σ_n of the poles according to the relations

$$\left\{ \begin{array}{l} \omega_n = 2\pi F_n, \text{ and} \\ \sigma_n = -\pi B_n. \end{array} \right. \quad (1.22-10)$$

This procedure is the same as measuring formant frequency and bandwidth from amplitude-frequency spectra of speech sounds. However, even if small losses are assumed, there will be many errors involved in the evaluation of two poles that come very close and thus determine a double peak. The two peaks may fuse into one maximum, but two poles cannot coincide except as a limiting condition when the cross-sectional area of a constriction is reduced to zero.

One effective method of sharpening the selectivity in the analysis of pole frequencies suggested by Huggins (1952), is to determine the frequencies where the second derivative of the phase function, i.e., $\frac{d^2\varphi(\omega)}{d\omega^2}$ goes through zero.

A single resonance

$$H_1(j\omega) = \frac{\sigma_1^2 + \omega_1^2}{[j(\omega - \omega_1) + \sigma_1][j(\omega + \omega_1) + \sigma_1]} \quad (1.22-11)$$

contributing to $H(j\omega)$ has the phase function

$$\varphi_1(\omega) = \operatorname{arg} \frac{\omega - \omega_1}{\sigma_1} + \operatorname{arg} \frac{\omega + \omega_1}{\sigma_1}, \quad (1.22-12)$$

$$\frac{d\varphi_1(\omega)}{d\omega} = \frac{\sigma_1}{\sigma_1^2 + (\omega - \omega_1)^2} + \frac{\sigma_1}{\sigma_1^2 + (\omega + \omega_1)^2}, \quad (1.22-13)$$

$$\frac{d^2\varphi_1(\omega)}{d\omega^2} = 2\sigma_1 \left[\frac{\omega - \omega_1}{[\sigma_1^2 + (\omega - \omega_1)^2]^2} + \frac{\omega + \omega_1}{[\sigma_1^2 + (\omega + \omega_1)^2]^2} \right], \quad (1.22-14)$$

$$\frac{d^2\varphi_1(\omega)}{d\omega^2} = 0 \text{ when } \omega = \omega_1 \sqrt{\frac{2\omega_0}{\omega_1} - \frac{\omega_0^2}{\omega_1^2}}, \quad (1.22-15)$$

or $\omega_{1e} = \omega_1 \sqrt{2 \sqrt{1 + \frac{1}{4Q_1^2}} - 1 - \frac{1}{4Q_1^2}}, \quad (1.22-16)$

where $\omega_0^2 = \omega_1^2 + \sigma_1^2$ and $Q_1 = \omega_1/2\sigma_1$.

The value ω_{1e} is a close estimate of the pole frequency. The decrement is determined from

$$\left[\frac{d\varphi_1(\omega)}{d\omega} \right]_{\omega=\omega_{1e}} \simeq \frac{1}{\sigma_1}, \quad (1.22-17)$$

provided $Q_1 > 1$. The above analysis is applicable to a first formant of very low frequency and large bandwidth, but it also represents the case of two higher formants of the same bandwidth and a small difference in pole frequency. In the former case the first formant pole is close to its negative conjugate.

A less effective method for detecting the pole frequencies in these situations is to determine the frequencies where the phase of the transfer function equals $(2n-1)\pi/2$.

The condition for two close resonances merging into a single maximum is that their frequency difference shall equal the bandwidth of each when considered in isolation. With application to a first formant of very low frequency, this condition implies that the resonance frequency F_1 be one-half the bandwidth. Under these circumstances, the first formant has a function identical with that of a matched half section low-pass prototype filter. If the Q of the first formant is higher than 1, the peak frequency ω_{pn} comes close to the angular frequency ω_n of the pole:

$$\omega_{pn} = \omega_1(1 - 1/4Q^2)^{\frac{1}{2}}, \quad (1.22-18a)$$

or

$$F_{pn} = F_1(1 - 1/4Q^2)^{\frac{1}{2}}, \quad (1.22-18b)$$

which means that a first formant of 100 c/s bandwidth and a pole frequency of 250 c/s has its peak at 245 c/s, and the frequency of 90 degrees phase shift is to be found at 255 c/s. The frequency of zero second derivative of the phase function, according to Eq. 1.22-14, is 248.8 c/s, which is rather close to the ideal value.

The calculations above exemplify the theorem that the phase information of minimum phase networks is entirely predictable from the amplitude information and vice versa. The highly selective phase analysis, or similar operations on the basis of the second derivative of the amplitude function, has only a limited use for the analysis of actual speech, since it might be hard to distinguish minor irregularities and parasitic or spurious formants from the basic formants of oral origin, i.e., those of the F-pattern. This objection is, however, serious only in automatic formant extractors that label the formants as F_1 , F_2 , F_3 , etc., according to their order of occurrence in the spectrum. The phase detection method can also be adopted for

the extraction of pole data from numerical calculations of $U_0/U_q = H(j\omega)$ if the derivatives are exchanged for difference values.

One further method for evaluating the poles from the numerical data that have been utilized for the present calculations is to start out from the loss free case, $N_a(\omega) = 0$, and determine the frequencies ω_{n1} where $N_b(\omega) = 0$. The damping constant σ_n and a correction term $\Delta\omega_n$ to the angular frequency ω_n of the pole s_n are then obtained by linear approximation of the complex function $N(j\omega)$ in the vicinity of its zero.

$$s_n = j\omega_{n1} - N(\omega_{n1})/N'(\omega_{n1}), \quad (1.22-19)$$

$$N'(\omega_{n1}) = \left[\frac{dN(s)}{ds} \right]_{s=j\omega_{n1}} = \frac{1}{j} \left[\frac{dN_b(\omega)}{d\omega} + j \frac{-N_a(\omega)}{d\omega} \right]_{\omega=\omega_{n1}}, \quad (1.22-20)$$

$$s_n = \sigma_n + j(\omega_{n1} + \Delta\omega_n),$$

$$\sigma_n = \frac{N_a N'_b}{N_a'^2 + N_b'^2}, \quad (1.22-21)$$

$$\Delta\omega_n = -\sigma_n \frac{N'_a}{N'_b}.$$

In the case of high Q formants, the correction term $\Delta\omega_n$ is negligible.

The essential scope of the calculations, or at least the first step, is to determine the pole frequencies from $N_b(\omega) = 0$. When dealing with a complicated network, it is desirable to reduce the numerical labor to a minimum. The following procedure, exemplified by the calculation of $\omega_1 = 2\pi F_1$, has proved to be useful. First make a guess at the most probable resonance frequencies. These are denoted ω_{e1} . Calculate $N_b(\omega_{e1})$. Next make use of the general relations between formant frequencies and spectrum envelope to predict the probable error involved in ω_{e1} . Expand $N_b(\omega)$ in its zeros

$$N_b(\omega) = (1 - \omega^2/\omega_{e1}^2)(1 - \omega^2/\omega_{e2}^2)(1 - \omega^2/\omega_{e3}^2)(1 - \omega^2/\omega_{e4}^2)k_{r4}, \quad (1.22-22)$$

where k_{r4} is the residue factor for resonances above the fourth as defined and evaluated in *Section 1.31*. Differentiate:

$$N'_b(\omega) = \frac{2}{\omega_{e1}} \cdot \frac{N_b(\omega)}{(1 - \omega^2/\omega_{e1}^2)}, \quad (1.22-23)$$

$$(\omega = \omega_{e1}),$$

and make use of a linear extrapolation.

The estimated deviation of ω_{e1} from ω_1 is

$$\Delta\omega_1 = -N_b(\omega_{e1})/N'_b(\omega_{e1}), \quad (1.22-24)$$

where $N_b(\omega_{e1})$ is obtained from the determinant calculations on the network and

$N_b(\omega_{e1})$ from the zero expansion above. This method was utilized in the first set of calculations carried out on *BARK*² in Stockholm. The machine was programmed to perform the error estimation and thus automatically supply itself with new and more probable values at the end of each recurrent cycle. The convergency was very good. A number of 2 to 5 iterative cycles were needed for the evaluation of each resonance frequency with an accuracy of better than 1 per cent.

This method, supplemented by interpolation, is recommended for those who undertake calculations on the basis of a detailed network structure without the aid of high-speed computing devices. It is actually not more complicated to perform a calculation of formant frequencies from a 10-section vocal tract than to calculate the ten first harmonics of a waveform. It is questionable whether phoneticians nowadays care to utilize numerical calculations. Those who want a mathematical check on their physiological X-ray investigations but lack an electrical analog machine can, however, make use of the technique developed above.

1.23 Transform Equations for Speech Production

A. THE VOCAL TRACT TRANSFER FUNCTION

The vocal tract transfer function for non-nasalized sounds relating volume velocity through the lips to volume velocity at the glottis can be put in the form

$$U_0/U_q = H(s) = \frac{K_G}{\prod_1^{\infty} (1-s/\hat{s}_n) (1-s/\hat{s}_n^*)} = \frac{K_G K_{rg}(s)}{\prod_{n=1}^g (1-s/\hat{s}_n) (1-s/\hat{s}_n^*)}, \quad (1.23-1)$$

where \hat{s} , \hat{s}_n^* are the conjugate poles. The constant K_G is related to the losses through vibration of the cavity walls. These losses are small and K_G will thus be given the value unity. At zero frequency $H(s)$ approaches unity, which corresponds to a continuity of volume velocity. In Eq. 1.23-1 the infinite product series representing ideal one-dimensional wave propagation is substituted for an approximation covering the first g poles and a correction factor³ $K_{rg}(s)$ which is the residue from the higher poles within the frequency region up to and including pole No. g .

If subglottal or nasal coupling is to be considered, or when the source is located higher up in the vocal tract, there enters an additional zero function $H_z(s)$ which includes a scale factor K_z :

$$H(s) = H_p(s)H_z(s), \quad (1.23-2)$$

where the pole function $H_p(s)$ is identical to $H(s)$ of Eq. 1.23-1.

$$H_z(s) = K_z \cdot s \prod_1^{\infty} (1-s/\bar{s}_n) (1-s/\bar{s}_n^*). \quad (1.23-3)$$

² Binary Arithmetic Relay Calculator of the Swedish Board of Computing Machinery in Stockholm.

³ A derivation of the factor $K_{rg}(s)$ is given by Fant (1959). It is of considerable importance for speech synthesis; see Section 1.31.

In nasalized vowels there is some wave propagation through the nasal tract, more precisely the fraction $L_M/(L_N+L_M)$ if resistive elements are disregarded. L_M is the effective inductance of the mouth outlet as seen from the uvula and L_N is the inductance of the nasal passages as seen from the same point.

$$L_M = \rho \int_{uvula}^{lips} \frac{dx}{A(x)}. \quad (1.23-4)$$

If finite values of the resistance elements R_M and R_N in series with L_M and L_N are taken into account, the scale factor for the oral and the nasal outputs are

$$K_{zM} = \frac{L_N}{L_N + L_M} \cdot \frac{s - s_N}{s - s_{NM}}, \text{ and}$$

$$K_{zN} = \frac{L_M}{L_N + L_M} \cdot \frac{s - s_M}{s - s_{NM}} = 1 - K_{zM}, \quad (1.23-5)$$

where

$$s_N = -R_N/L_N,$$

$$s_M = -R_M/L_M, \text{ and}$$

$$s_{NM} = -(R_N + R_M)/(L_N + L_M). \quad (1.23-6)$$

If $R_N/L_N > R_M/L_M$ there will be a small boost in the mouth output at very low frequencies. Since $K_{zM} + K_{zN} = 1$, the effect is totally canceled if the nasal and the oral outputs are added in equal proportion. The finite vocal tract wall impedance has a shunting effect analogous to that of the nasal cavities. The transform equations derived for the nasal passages are applicable to any network shunting the vocal cavities.

A similar reasoning can be utilized for the evaluation of the constant K_z associated with the cavities behind the source. At frequencies well below the conjugate poles and zeros of the complete system, the current volume velocity delivered by the source $E_v(s)$ is given by

$$\lim_{s \rightarrow 0} U_v(s) = E_v(s) s C_b,$$

where

$$C_b = \frac{1}{\rho c^2} \int_{source}^{termination} A(x) dx. \quad (1.23-7)$$

C_b is the capacitance of the total volume contained behind the source. In unvoiced sounds the vocal cords do not effectively terminate the vocal tract. The trachea and the lungs enter the picture also, though chiefly at low frequencies. The current delivered from the source is also the output current. Thus $K_z = C_b$ and

$$H_z(s) = C_b \cdot s \prod_1^{\infty} (1 - s/\bar{s}_n) (1 - s/\bar{s}_n^*). \quad (1.23-8)$$

The function $H_z(s)$ has a single zero at the origin. When the source is located within a narrow constriction, the effect at higher frequencies of this zero and the first pair of conjugate poles from $H_p(s)$ is equal to that of a single low frequency pole on the real axis representing an internal impedance largely composed of the inductance and resistance within the source constriction. If all the conjugate zeros of $H_z(s)$ are effectively neutralized by poles of $H_p(s)$, the back cavities can be disregarded completely and the pressure can be substituted for a volume velocity source

$$U_v = \frac{E_v(s)}{R_v + sL_v}, \quad (1.23-9)$$

where $R_v + sL_v = Z_v(s)$ is the transform of the constriction impedance. This impedance then terminates the effective cavity network. The transfer from voltage source to current source, speaking in terms of electrical quantities, is useful when dealing with sounds produced at the glottis. It is also useful for calculating the front cavity response of stop sounds and fricatives under the conditions of small front-to-back coupling.

B. RADIATION, SOURCE, AND OTHER "CONSTANT" FACTORS

The transfer from volume velocity through the lips $U_0(\omega)$ to pressure $P_1(\omega)$ in the soundfield at a distance l is approximately determined by the following power equation:

$$U_0^2 \cdot R_0 = \frac{P_1^2}{\rho c} 4\pi l^2; \quad (1.23-10)$$

Assuming a spherical baffle:

$$R_0 = \frac{\rho \omega^2}{4\pi c} K_S(\omega). \quad (1.23-11)$$

Thus

$$\frac{P_1}{U_0} = \frac{\rho \omega}{4\pi l} \sqrt{K_S(\omega)}. \quad (1.23-12)$$

However, the radiation is uniform in all directions only at low frequencies. The combined effect of the directivity of radiation and the increase of R_0 in excess of ω^2 can be calculated from the data given by Morse (1948) pertaining to a small circular baffle on the surface of a sphere

$$\frac{P_1}{U_0} = \frac{\rho \omega}{4\pi l} K_T(\omega), \quad (1.23-13)$$

where $K_T(\omega)$ is the total correction needed.

A frequency curve of $20 \log_{10} K_T(\omega)$ for a sphere of radius 9 cm is shown in Fig. 1.2-6. The maximum value of this correction is 7 dB at high frequencies and it reaches 5 dB at 2000 c/s. This curve is, however, only one of several constant frequency functions that should be taken into account together with the voice source slope and the radiation transfer curve. Among other *constant* frequency characteristics contrib-

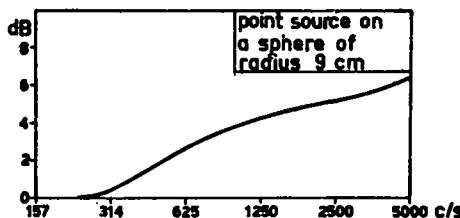


Fig. 1.2-6. Frequency characteristics of sound energy radiated in normal incidence from a small area on the surface of a sphere of radius 9 cm comprising the combined effect of directivity of radiation and increase of radiation resistance in excess of ω^2 .

uting to the attenuation of the highest part of the spectrum above 5000 c/s, there is the transmission through sinus Morgagni tuned by the false vocal cords (van den Berg, 1955b), and the shunting effect of the sinus piriformis on both sides of the larynx tube and the influence of other cross modes representing particle motion perpendicular to the air stream. The departure of radiation resistance from the ω -proportionality at very high frequencies in a direction towards the limiting $\rho c/A_0$ value contributes also, but to a very small extent. The actual baffle effect of the head plus body is also a minor feature of constant character that is not very well known.

The voice source characteristics and other constant functions of low-pass character could be approximated by four poles on the negative real axis in the form

$$U_q = \frac{U_{q0}}{\prod_{r=1}^4 (1 - s/s_r)} \quad (1.23-14)$$

The first two of these poles pertaining to the voice source are both of the order of $s_{r1} \approx s_{r2} = 2\pi \cdot 100$ c/s, but the variability with regard to speaker and stress is apparent. The third and fourth poles are of the order of $s_{r3} = 2\pi \cdot 2000$ and $s_{r4} = 2\pi \cdot 4000$. These are needed for synthesis in a transmission line analog but have no function in a formant circuit synthesizer since such a system preserves the natural spectrum level of voiced sounds up to 3500 c/s only.

C. THE COMPLETE LAPLACE TRANSFORM

The complete transform of ideal voiced non-nasal sounds measured at a distance of l cm from the lips is obtained from Eq. 1.23-1,13,14.

$$P_d(s) = \frac{\rho s}{4\pi l} K_T(s) \frac{U_{q0}}{\prod_{r=1}^4 (1 - s/s_r)} \cdot \frac{K_{rg}(s)}{\prod_{n=1}^g (1 - s/\hat{s}_n) (1 - s/\hat{s}_n^*)} \quad (1.23-15)$$

comprising the effects of radiation and radiation directivity, the voice source characteristics, and other low-pass filter effects. Each of the first g formants is represented

by a pair of conjugate poles and the influence of higher poles is contained within the factor $K_{rg}(s)$. Eq. 1.23-15 is the formal operator transform underlying the specific filter function to be developed in Eq. 1.3-11 and illustrated by Fig. 1.3-5.

The oral output of nasalized vowels can be represented by the following transform in which the nasality effect has been separated:

$$P_l(s) = \frac{\rho s}{4\pi l} K_T(s) \frac{U_{q0}}{\prod_{r=1}^g (1-s/s_r)} \cdot \frac{s-s_N}{s-s_{NM}} \cdot \frac{L_N}{L_N+L_M} \cdot \frac{\prod_{b=1}^f [(1-s/\bar{s}_b)(1-s/\bar{s}_b^*)]}{\prod_{b=1}^f [(1-s/\hat{s}_b)(1-s/\hat{s}_b^*)]} \cdot \frac{K_{rg}(s)}{\prod_{n=1}^g (1-s/s_n)(1-s/s_n^*)}. \quad (1.23-16)$$

A non-nasal sound produced from a source $E(s)$ with finite coupling between the parts in front of and behind the source has the transform

$$P_l(s) = \frac{\rho s}{4\pi l} K_T(s) E(s) s C_b \frac{\prod_{n=1}^h (1-s/\bar{s}_n)(1-s/\bar{s}_n^*)}{\prod_{n=1}^g (1-s/\hat{s}_n)(1-s/\hat{s}_n^*)} \cdot \frac{K_{rg}(s)}{K_{rh}(s)}. \quad (1.23-17)$$

Of course, $K_{rg}(s)$ and $K_{rh}(s)$ can be left out if an infinite number of conjugate poles and zeros are included.

In the case of small coupling, those of the g poles that are predominantly dependent on the front resonator can be specified as a separate factor. It should be noted that the lowest conjugate pole (corresponding to the first formant) may become non-oscillatory if the source is situated in a very narrow constriction. In that case s_n and s_n^* are negative, real, and different.

D. THE INVERSE TRANSFORM

The inverse transform of $P(s)$ represents the pressure versus time speech wave. The inverse transform of Eq. 1.23-15 extended to periodic stationary conditions⁴ is

$$p(t) = \sum_{m=0}^v \left\{ \sum_{r=1}^4 A_{re} s_r(t-mT_0) + (-1)^n \sum_{n=1}^g A_{ne} \sigma_n(t-mT_0) \cdot \cos [\omega_n(t-mT_0) + \varphi_n] \right\}. \quad (1.23-18)$$

This is the superposition of a finite number of damped oscillations and non-oscillatory exponentials excited during each fundamental period from the first $m = 0$ to $m = v$,

⁴ A more extensive treatment is given by Fant (1959).

with the last pulse omitted during the utterance. The factor $(-1)^n$ implies that alternating phases should be used in an electrical analog synthesizer of the parallel type.⁵ The constants A_r , A_n , and φ_n are, of course, completely determined by the particular set of poles and other constants of the frequency transform $P(s)$.⁶ The initial amplitudes are closely related to the formant amplitudes, which is shown in Fant, 1959.⁷

In general, the length of the voice fundamental period $T_0 = 1/F_0$, as well as the initial amplitudes A_r and A_n , will vary from one period to the next. Besides these time dependent changes related to source variations, there can occur gradual variations of the pole frequencies s_r and $s_n = \sigma_n \pm j\omega_n$. Thus the extent of a formant transition within a voice fundamental period could be studied. The generalized time dependent definition of formant frequency and bandwidth is thus:

$$\begin{aligned} F_n(t) &= \omega_n(t)/2\pi, \text{ and} \\ B_n(t) &= -\sigma_n(t)/\pi. \end{aligned} \tag{1.23-19}$$

In broad-band spectrograms from low-pitched male voices this effect may be seen, e.g., in the transition from a labial voiced stop to a vowel. In narrow-band spectrograms such rapid frequency variations cannot be detected owing to the greater averaging time of the filter.

Formant bandwidth variations or rather σ_n -variations within a fundamental period can be observed in oscillographic records. Appreciable time-dependent variations of B_1 can be expected owing to the flow-dependent resistances at the glottis and at very narrow supraglottal passages, but also from the superpositional effect from several glottis flow discontinuities within a fundamental period. It is possible that the peak factor within a voice fundamental period could be of a more direct auditory significance than a bandwidth or decay constant.

⁵ Also shown by Weibel (1955).

⁶ See also Flanagan (1957b). These dependencies are those discussed in Section 2.33.

⁷ This publication also contains direct and inverse transforms of stationary periodic sounds.

1.3 ANALYTICAL CONSTRAINTS ON THE COMPOSITION OF SPEECH SPECTRA

1.31 *Idealized Spectral Description of Voiced Sounds*

The relative importance of the separate formants of voiced sounds decreases with increasing order above F_2 . F_1 and F_2 are the main determinants of vowel quality.¹ F_3 and F_4 contribute significantly to the phonetic quality of front vowels, but in back vowels they are of minor importance only. F_3 and F_4 , as well as F_0 , also provide certain information on personal voice characteristics.

From the representation of speech production in terms of Laplace transforms, outlined in *Section 1.23*, it is clear that an ideal non-nasalized vowel is completely and uniquely specified by the source characteristics plus the data on formant frequencies and bandwidths. Formant levels can be calculated from these data and therefore constitute redundant information (Fant, 1956).

The number of variables may thus be reduced from the 20-40 harmonics of a harmonic specification to four formants plus source characteristics required by the theory of Laplace transforms. The bandwidth data are to a certain extent predictable from the formant frequency data, and the source characteristics are more or less constant features for a speaker, primarily dependent on the amount of vocal stress. The only remaining independent variables in the production of voiced sounds are thus the formant frequencies, i.e., the F-pattern; see also *Section 1.13*. The great importance of formant frequencies for speech specification agrees of course with practical experience from analysis and synthesis, but it is of some interest that it can be mathematically proved.

The dependent relationships between formant frequencies and the shape of the spectrum envelope have both theoretical and practical implications in speech analysis and speech synthesis. Under practical conditions no vowel is ideal and there exist minor humps, dips, and even extra formants due to the actual composition of the voice source spectrum and to the coupling to the subglottal system. Nasality and also

¹ See *Section A.12* for definitions, measurement techniques, and ranges of variation.

intermodulation from the recording and analyzing equipment have similar effects. The complete but more redundant harmonic specification should be used if it is considered desirable to include all observable details in the description. On the other hand, it can be convenient to make use of the conditioned relationships between formant frequencies, i.e., the F-pattern, and spectrum shape to decide what is a formant and what is a mere distortion. The dependency of voiced sounds on source and filter disregarding phase information can be put in the form

$$|P(f)| = |U(f)| |H(f)| |R(f)|, \quad (1.3-1)$$

where brackets indicate absolute values and (f) function of frequency. $|U(f)|$ symbolizes the amplitude-versus-frequency characteristics of the source,² $|H(f)|$ the frequency selective gain function of vocal transmission, and $|R(f)|$ the frequency characteristics of radiation, i.e., the conversion from volume velocity through the lips to pressure in the soundfield. The product $|H(f)R(f)|$ constitutes the complete filter function, $T(f)$, compare *Eq. 1.1-3*.

As discussed in connection with *Eq. 1.23-14* and *1.23-15* it is, however, profitable to rearrange *Eq. 1.3-1* so as to bring together all functions that do not vary significantly in voiced sounds. The combined source and radiation characteristics constitute a spectrum that falls off at the approximate rate of *6 dB/octave*. $|R(f)|$ is approximately proportional to frequency, f , and it will be assumed that $|U(f)|$ is approximately proportional to $1/f^2$ above a cutoff frequency of *100 c/s*. This relation can thus be written

$$|U(f)| |R(f)| = P_k \frac{(f/100)}{1 + (f/100)^2}, \quad (1.3-2)$$

where P_k is a constant determining the particular sound pressure level.

The vocal tract transfer function $|H(f)|$ will next be broken down into two basic factors according to the presentation in *Eq. 1.23-1*. One is to be uniquely specified by the first four formants and the other shall take care of all the constant frequency characteristics. Thus

$$|H(f)| = k_{r4}(f) |H_1(f)| |H_2(f)| |H_3(f)| |H_4(f)|, \quad (1.3-3)$$

where $|H_1(f)|$ is the contribution from the first formant, $|H_2(f)|$ from the second formant, and so on. The factor $k_{r4}(f)$ contains the remaining frequency characteristics and can physically be explained as the contribution from formants of higher number than the fourth. It is, of course, also possible to base the specification on three variable formants only and to include the effects of the remaining higher formants in a residual factor denoted $k_{r3}(f)$.

The magnitude of $k_{r3}(f)$ or $k_{r4}(f)$ is by no means negligible, as can be seen from *Fig. 1.3-1* which shows the transfer function of a tube of the effective length 17.5 cm closed at the driving point end and open at the radiating end. This is an idealized

² Instead of the symbol $S(f)$ for source adopted in the generalized *Eq. 1.1-3* it has been preferred to adopt the symbol $U(f)$ which signifies a volume velocity source spectrum.

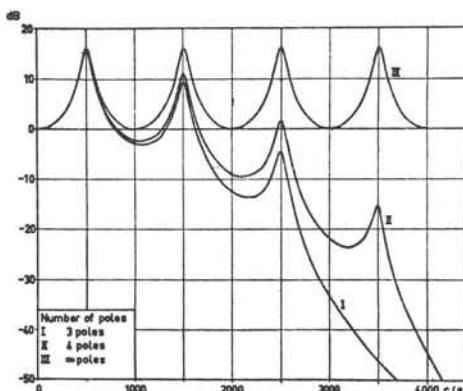


Fig. 1.3-1. Frequency characteristics of the ratio of output volume velocity to input volume velocity of an idealized resonator with formant frequencies at odd multiples of 500 c/s and formant bandwidths of 100 c/s. Curve III is the complete solution according to the theory of a transmission line short-circuited at the far end and open-circuited at the input, i.e., a constant current source is anticipated. Curve II is a 4-formant approximation, i.e., a 4-pole system function, and curve I is an approximation with 3 poles only.

resonator simulating the articulation of a neutral open vowel. The formants occur at odd integers of 500 c/s, i.e., 500, 1500, 2500, 3500, ... c/s, and the spectrum level at the valleys between formants is everywhere the same. The losses are assumed to be constant, i.e., all formants have the same bandwidth, viz., $B = 100$ c/s. The peak levels of the formants are the same. When this transfer function is simulated by a few resonant circuits in cascade placed at the appropriate center frequencies, it is found that the spectrum level falls off fast at high frequencies but that the match is good at low frequencies. The difference between curve III, i.e., $|H(f)|$ and curve II, i.e., $|H_1(f)H_2(f)H_3(f)H_4(f)|$, is the factor $k_{r4}(f)$ which can be analytically calculated³ by means of a series expansion and summation of the contributions from the 5th and higher poles up to infinity:

$$20 \log_{10} k_{r4} = 0.54 x^2 + 0.00143 x^4 \quad (\text{dB}). \quad (1.3-4a)$$

Similarly:

$$20 \log_{10} k_{r3} = 0.72 x^2 + 0.0033 x^4 \quad (\text{dB}), \quad (1.3-4b)$$

and

$$20 \log_{10} k_{r2} = 1.06 x^2 + 0.0102 x^4 \quad (\text{dB}), \quad (1.3-4c)$$

where $x = f/f_1$ and $f_1 = c/4l_{tot}$. Here l_{tot} denotes the total length of the vocal tract.

The constant factors are brought together in Fig. 1.3-2 where curve I constitutes $|U(f)R(f)|$ according to Eq. 1.3-2, and curve II is identical with k_{r4} . The curve III = I+II is thus the sum of all the constant frequency dependent factors that must be incorporated in a speech synthesizer containing four formant circuits in cascade. The practical need of this correction in the synthesis of speech has been demonstrated with the Swedish speaking machine OVE (Fant, 1953a, 1957).

³ See Fant (1959) for the detailed derivation.

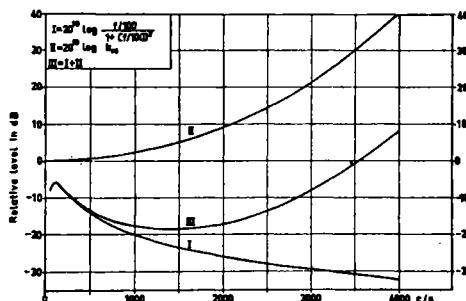


Fig. 1.3-2. Summation of the constant factors contributing to the spectrum envelope of vowels, including the idealized frequency characteristics of the voice source, of radiation, and of the poles higher than number 4.

- I. is the sum of a $+6 \text{ dB/octave}$ rise for the radiation and an idealized voice source spectrum envelope falling off at a rate of -12 dB/octave at frequencies above 100 c/s .
- II. is the contribution from the poles of higher number than 4 to the spectrum level at frequencies below 4000 c/s .
- III. = I + II is the total frequency correction to be applied to the sum of the resonance curves for the first four formants when calculating the vowel spectrum envelope from the formants or when driving a 4-pole resonance analog of the vocal tract by a source consisting of a series of unit impulses.

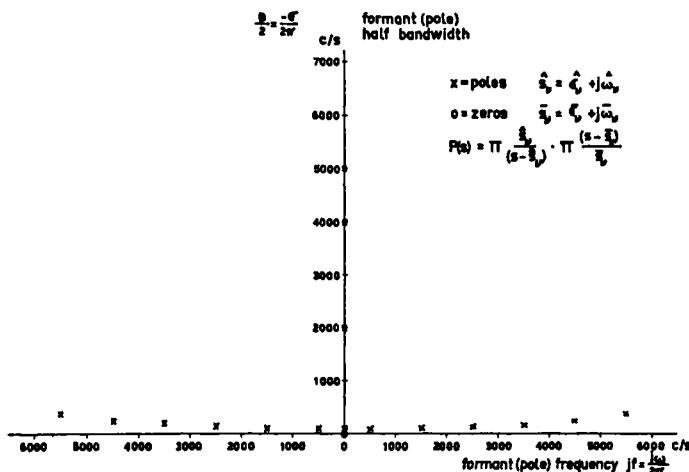


Fig. 1.3-3. Complex frequency representation of the constituents of a vowel. The conjugate poles pertain to the idealized neutral vowel. The zero at zero frequency is an approximation of the radiation characteristics. The poles on the negative real bandwidth axis approximate the frequency characteristics of the voice source utilized in an earlier work.

The mathematical aspects of a formant as a determinant of spectral qualities are contained in the complex frequency concept pole. As implied by the treatment in *Section 1.23*, the frequency characteristics of the complete sound can be derived from a number of points, poles and zeros, in the complex frequency plane; see *Fig. 1.3-3*.

A few minor changes have been made in the usual form of representation. Thus all scale values are divided by 2π and the coordinate system has been rotated 90 degrees in order to make the frequency axis horizontal. A formant enters via a pair of conjugate poles of frequency, plus and minus the formant frequency F_n . The ordinate of the diagram, $\sigma_n/2\pi$, is accordingly one-half of the formant bandwidths, i.e., $B_n/2$. The frequency characteristics of the voice source enter via poles on the *bandwidth* axis. The particular source function exemplified in *Eq. 1.3-2* corresponds to two poles at 100 c/s. In *Fig. 1.3-3* the source poles are placed at 100 c/s, 2000 c/s, 4000 c/s, and 5000 c/s. These approximate the source frequency characteristics utilized by Stevens et al. (1953).⁴ The effect of radiation enters via a zero at the origin. The formant poles are placed at frequencies which are odd integers of 500 c/s and are assumed to have bandwidths increasing with frequency.

Source, vocal transmission, and radiation have accordingly been represented by points in the complex frequency plane. Any one of these functions $U(f)$, $H(f)$, $R(f)$, or their product; see *Eq. 1.3-1*, can easily be derived from the following graphical procedure:

Draw vector lines from all the poles to a point $(f,0)$ on the frequency axis and also a separate set of vectors from the poles to the origin. Calculate the product of the lengths of the vectors to the origin and divide it by the product of the vectors to the point on the frequency axis. If there are no zeros in the complex frequency plane this ratio is the desired value of the system function. If zeros occur, divide the pole function by a zero function that is obtained by a completely analogous operation. The zero at the origin is treated differently. Only the vector from this zero to the point $(f,0)$ on the frequency axis is included.

Ideal vowel spectra have no other zeros than the one at the origin, introduced by the radiation characteristics. As previously mentioned, this is a +6 dB/octave contribution that is canceled at frequencies above 100 c/s by the lowest source pole on the bandwidth axis. The remaining source poles each provide a -6 dB/octave contribution to the overall slope at sufficiently high frequencies.

Poles and zeros of finite frequency value, i.e., the points that do not lie on the bandwidth axis, must occur in conjugate pairs. Every pole or zero of positive frequency position is thus paired with a point at negative frequencies. The zeros enter the system function when there are shunting cavity systems in the speech production or when the cavities behind the source must be taken into consideration, as will be described later.

A pair of conjugate poles s_n , $s_n^* = F_n \pm j\omega_n$ define the contribution $H_n(f)$ from

⁴ This source function is not superior to the 100 c/s double pole, as judged from synthesis with OVE.

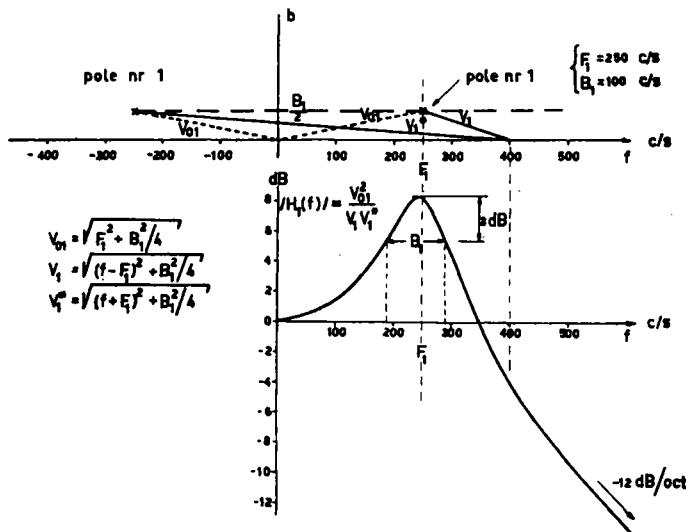


Fig. 1.3-4. The derivation of the resonance curve of a formant from its frequency and bandwidth, i.e., from the two quantities that determine the formant pole. The particular values utilized in the example, i.e., a formant frequency of 250 c/s and a formant bandwidth of 100 c/s could pertain to the first formant of a high front vowel. The procedure is begun by plotting the two conjugate poles, i.e., the points $B_1/2, \pm F_1$. The resonance curve is then

$$20 \log_{10} \frac{V_{01}^2}{V_1 V_1^*},$$

where V_{01} is the length of the vector from the origin to the poles and V_1 and V_1^* are the lengths of the vectors drawn from the frequency f on the f -axis to the positive pole and to the negative pole respectively.

the formant number n . The absolute value of the contribution is according to the vector representation

$$|H_n(f)| = \frac{s_n s_n^*}{|(s-s_n)| |(s-s_n^*)|} \Big|_{(s=j\omega)}, \quad (1.3-5a)$$

$$\text{or } H_n(f) = \frac{F_n^2 + (B_n/2)^2}{\sqrt{(f-F_n)^2 + (B_n/2)^2} \sqrt{(f+F_n)^2 + (B_n/2)^2}}, \quad (1.3-5b)$$

where, as before, the formant is characterized by its frequency $F_n = \omega_n/2\pi$ and its 3-dB bandwidth $B_n = F_n/2\pi$.

The frequency curve $H_n(f)$ is identical with the voltage transfer function of a series RLC -circuit, defined by the ratio of condenser voltage to the impressed series voltage within the loop.

1.32 The Relations Between Formant Frequencies and Spectrum Envelopes

The shape of $|H_1(f)|$, pertaining to a first formant at 250 c/s of bandwidth 100 c/s, is shown in Fig. 1.3-4. At low frequencies $|H_n(f)|$ always approaches the value unity, i.e., 0 dB. The value of $|H_n(f)|$ at the formant frequency is

$$Q = \frac{F_n}{B_n}. \quad (1.3-6)$$

Providing the Q of formants is large Eq. 1.3-5 may be put in the simplified form

$$|H_n(f)| = \frac{1}{\sqrt{(1-x^2)^2 + x^2/Q^2}}, \quad (1.3-7)$$

where $x = f/F_n$ is the relative frequency⁵ under observation. Because of large Q values, the part of the resonance curve above $x = \sqrt{2}$ can be written

$$|H_n(f)| \simeq \frac{1}{x^2 - 1}, \quad (1.3-8)$$

$$x > \sqrt{2},$$

and at still higher frequencies

$$|H_n(f)| \simeq \frac{1}{x^2}, \quad (1.3-9)$$

$$x > 2$$

or

$$|H_n(f)| \simeq \frac{F_n^2}{f^2},$$

$$f > 2 F_n.$$

At a frequency of $f = \sqrt{2} F_n$ the resonance curve $H_n(f)$ has fallen to the value unity, i.e., 0 dB, and at higher frequencies the curve falls off at a rate of 12 dB/octave. Since both f and F_n enter, the effect of shifting the formant frequency F_n down one octave will be to decrease the spectrum level by the constant amount of 12 dB at all frequencies above $x = 2$, and increasingly more in the frequency region just above the original position of the formant peak. This is a fundamental rule governing the interrelations between formant frequencies and the shape of a spectrum envelope.

A decomposition of an ideal vowel into its variable components $|H_1(f)|$, $|H_2(f)|$, $|H_3(f)|$, and $|H_4(f)|$, each of the form of Eq. 1.3-7 and the remaining constant factor $k_{r4}(f) |U(f)| |R(f)|$ the previously defined by Eq. 1.3-2 and 1.3-4a has been carried out in Fig. 1.3-5, pertaining to an F-pattern of odd integers of $F_1 = 500$ c/s and further illustrating the effect of shifting F_1 from 500 c/s to 250 c/s retaining the rest of the F-pattern. The factorialized presentation of Eq. 1.3-1 and 1.3-3 has been replaced by a summation of the logarithms of the separate factors, that is, by a summation of spectral levels in dB.

⁵ More precisely $x = f/\sqrt{F_n^2 + (B_n/2)^2}$ and $Q = (\sqrt{F_n^2 + B_n^2/4})/B_n$; compare Section 1.22.

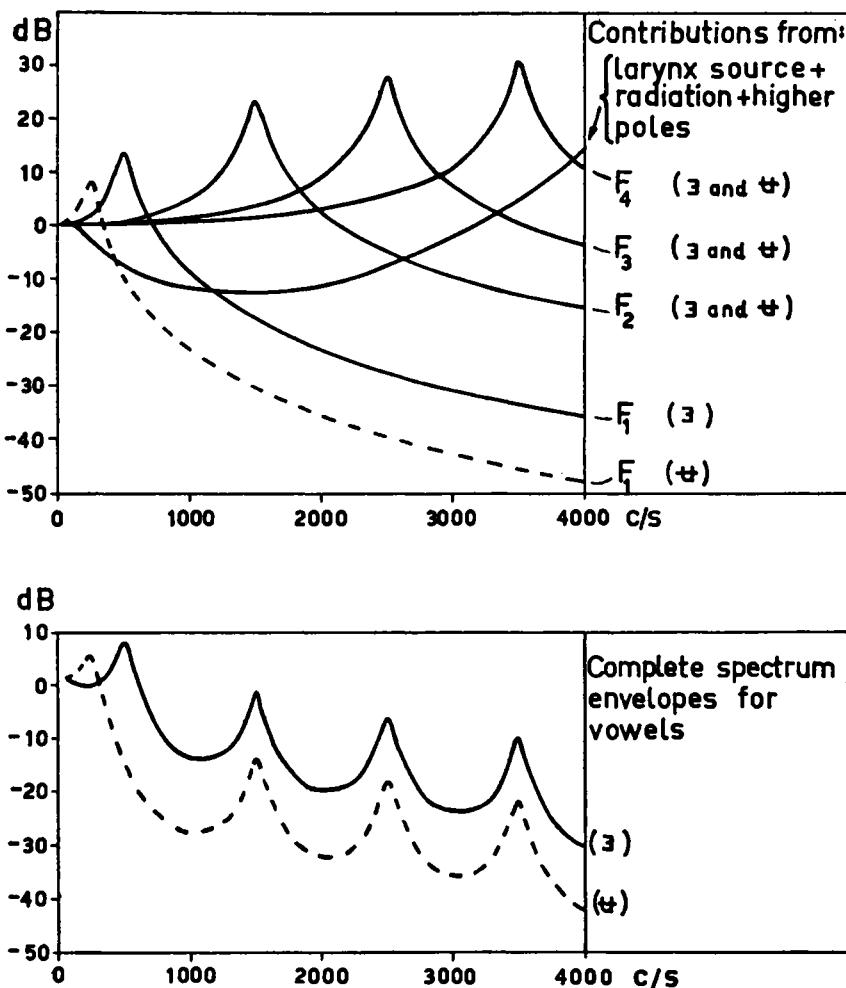


Fig. 1.3-5. The effect of a shift in frequency of the first formant on the level of the spectrum envelope demonstrated by the decomposition of two vowel spectra in elementary resonance curves. [ə] denotes the open tube idealized neutral vowel, with formant frequencies at 500, 1500, 2500, ... c/s. A shift in the frequency of the first formant only from 500 to 250 c/s results in a new sound, approximately [ɛ], in which the part of the spectrum envelope above 500 c/s is practically identical with that of the original vowel except for a loss of level of 12 dB.

$$\begin{aligned}
 20 \log_{10} |P(f)| = & 20 \log_{10} k_{r4}(f) |U(f)R(f)| + 20 \log_{10} |H_1(f)| + \\
 & + 20 \log_{10} |H_2(f)| + 20 \log_{10} |H_3(f)| + \\
 & + 20 \log_{10} |H_4(f)|, \tag{1.3-10}
 \end{aligned}$$

or
$$20 \log_{10}|P(f)| = 20 \log_{10}k_r(f)P_k \frac{f/100}{1+(f/100)^2} - \sum_{n=1}^* 10 \log_{10} \left[\left(1 - \frac{f^2}{F_n^2}\right)^2 + \frac{f^2}{F_n^2 Q_n^2} \right] (dB). \quad (1.3-11)$$

All formant bandwidths are 100 c/s. The lower part of Fig. 1.3-5 shows the calculated spectrum envelope of the ideal neutral vowel denoted [ɛ] and defined by $F_1 = 500$ c/s, $F_2 = 1500$ c/s, $F_3 = 2500$ c/s, $F_4 = 3500$ c/s and also the resulting spectrum envelope when F_1 is lowered to 250 c/s. The phonetic quality of a sound of the latter F-pattern resembles [u] as in Norwegian *hus*. Observe the 12-dB lower spectrum level at frequencies well above F_1 in [u] as compared to [ɛ]. The total intensity integrated over the whole spectrum is also lower when F_1 is shifted to the lower frequency position.

The technique employed in Fig. 1.3-5 has been further elaborated for the calculation of complete spectra of voiced sounds on the basis of formant frequencies. Instead of starting from measured formant data, it was considered desirable to perform a mathematical synthesis on the basis of formant frequencies that were systematically varied in steps of 250 c/s and 125 c/s in F_1 , 250 c/s and 500 c/s in F_2 , and 500 c/s in F_3 . F_4 was held constant at 3500 c/s. The phonetic symbols attached to the calculated spectrum envelopes of Fig. 1.3-6 are only an indication of the application of IPA symbols, and they are thus not intended as an attempt at a physical specification of the standard IPA symbols. The shape of each of these spectrum curves is, however, fairly representative of a natural sound of the same formant frequencies. It has accordingly been found that the formant level variations in natural speech as a function of F-pattern variations, are well predictable from the theory (Fant, 1956).

There are two different types of questions to be posed when interrelating formant frequencies and the relative levels within a spectrum envelope. The first, and mathematically more simple one, was that introduced by Fig. 1.3-5. What are the changes in envelope level at any particular frequency due to a shift in frequency of one or more of the formants? The answer is obtained simply by a reference to the elementary resonance, or rather filter curve, of a formant, Eq. 1.3-7, or to its approximations, Eq. 1.3-8 and 9.

The second, and more intricate, question pertains to the level changes at the peak of the variable formant, in which case the location of this formant relative to all other formants must be considered. This evaluation is facilitated by plotting all formants as points within the complex frequency plane and drawing vectors from each of these to the frequency under observation, as earlier discussed and in conformity with the basic form of the elementary formant filter function, Eq. 1.3-5. The contribution of the non-variable spectral constituents comprising the vocal cord source spectrum and the effect of radiation and of higher poles, as summarized in Fig. 1.3-5, must also be taken into account.

On the basis of these techniques the following general statements can be made.

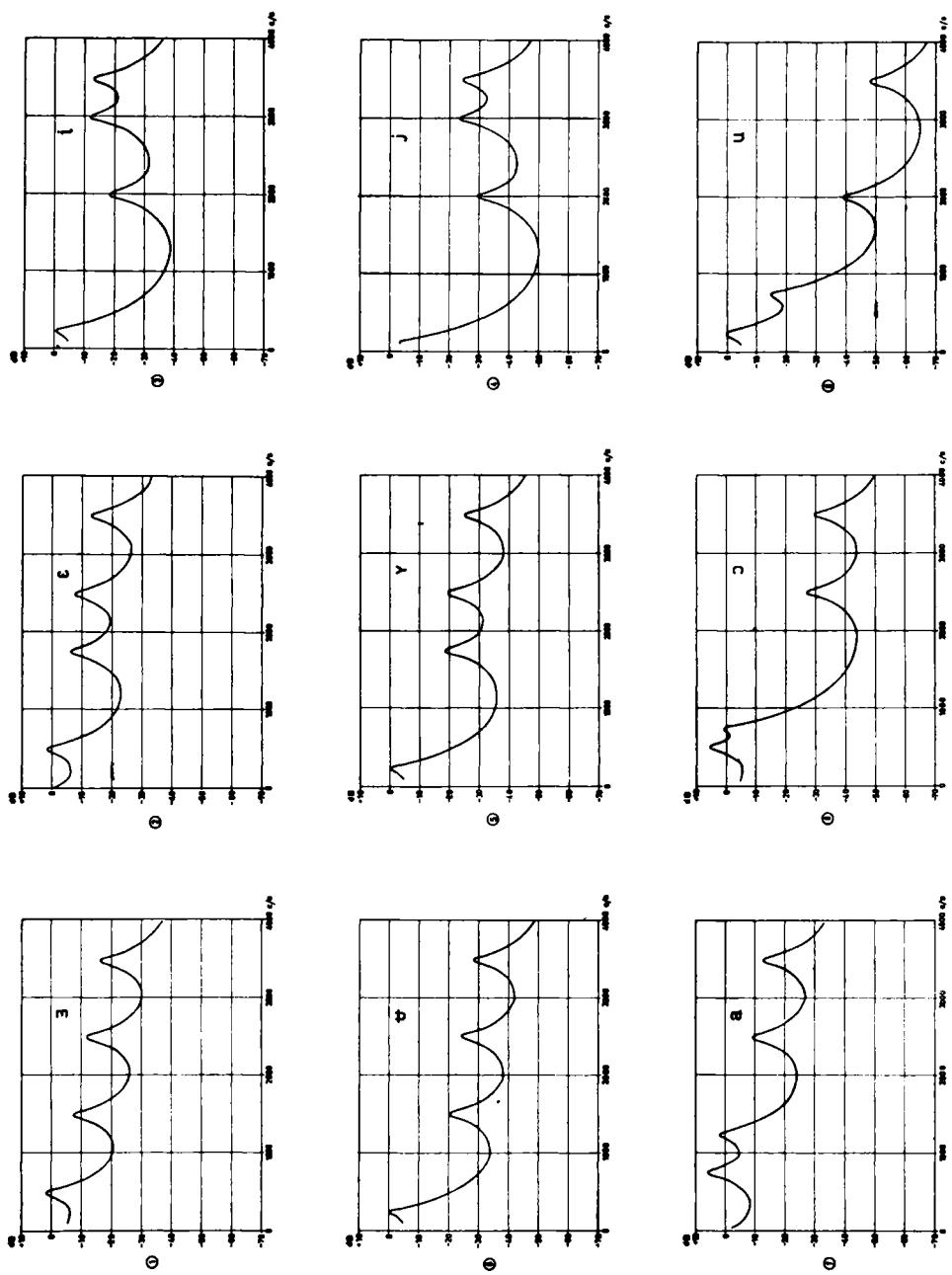


Fig. 1.3-6. Vowel spectrum envelopes derived from a mathematical synthesis in terms of elementary resonance curves, one for each formant as in Fig. 1.3-5. A voice source of -12 dB/octave slope and formant bandwidths of 100 c/s were assumed. These calculated spectra have been transcribed by the closest IPA symbols but should not be interpreted as phonetic quality norms. They demonstrate that spectrum shape and intensity levels are predictable from formant frequencies.

- 1) A shift in the frequency F_n of a formant brings about an intensity level change of the sound which is mainly confined to frequencies above F_n and amounts to $+12 \text{ dB}$ for an increase of one octave in F_n (see *Eq. 1.3-8* and *Fig. 1.3-5* and *1.3-6*). This may be called the low-pass filter rule; any formant acts as a low-pass filter of cutoff frequency $\sqrt{2} F_n$, where F_n is the formant frequency. The differential effect of an increase in F_1 only may be observed by comparing the spectrum envelopes of [a], [u], [e], [y], and [i], [j].
- 2) If the distance between two relatively close lying formants is reduced by a factor 2 the peaks are raised by 6 dB and the valley between the formants is raised by 12 dB . If the two formants come too close they will merge to a single observable maximum as discussed in *Section 1.42*. The effect of a simultaneous increase in F_1 and decrease in F_2 may be seen by comparing the curve labeled [a] with that of either [e] or [a].
- 3) The intensity variations of the first formant as a result of a shift of F_1 are not very great, provided F_1 does not come very close to F_2 . The 6 dB/octave -increase in L_1 expected from the increase in Q_1 due to an F_1 -rise at constant B_1 , is almost completely counteracted by the combined source and radiation characteristics providing a -6 dB/octave slope. The residue from the higher poles makes the latter slope less negative at frequencies above 500 c/s and thus conditions an L_1 -increase with rising F_1 .
- 4) The total overall intensity values of the synthetic sounds, illustrated in *Fig. 1.3-6*, are primarily conditioned by F_1 , and the first formant contributes always more to the total intensity than any other formants. In terms of loudness sensation, on the other hand, the second formant becomes more important, but does not generally contribute more than the first except at very low reception levels. An ordering in terms of the total intensity, integrated over the entire spectrum envelope, provides the following series: [a] [ɔ] [ɛ] [ə] [i] [u] [y] [j]. That order is fairly representative of natural speech.

Articulatory narrowing causes a shift down of F_1 and accordingly a drop of intensity, especially for the higher formants. The consonantal sign [j] has been assigned to curve No. 3 of *Fig. 1.3-6* since it differs from [i] basically by virtue of a lower F_1 . Noise elements are not necessary for the identification of the phoneme [j]. The syllable [ji] can be synthesized with the previously mentioned speaking machine *OVE* on the basis of formant frequency variations, retaining only those intensity variations that are conditioned by the analytic relations dealt with above. These are also sufficient for preserving syllabic divisions within sentences like *How are you, I love you* (Fant, 1953a, 1957).

The general relations between formant frequencies and formant levels discussed above are useful for discussing basic spectral attributes and distinctive features of vocalic sounds, i.e., *compactness* and its opposite *diffuseness*, and *gravity* and its opposite *acuteness*, as defined by Jakobson et al. (1952), and later by Jakobson and

Halle (1956). Such a discussion was undertaken in the article mentioned earlier (Fant, 1956).

A statement as to compactness or gravity has relational significance only, i.e., it may be used when two sounds are being compared. Ideally, a feature should be expressible by a simple quantitatively well-defined spectral attribute, in other words, a parameter expressing the common denominator of all minimal pairs to be considered. This is desirable but not always sufficient, e.g., for the purpose of a maximally precise phonetic description.

Vowel compactness defined as the degree of energy concentration in the neighborhood of 1000 c/s is not so easy to express by a formula in spite of the apparent simplicity. One formulation is to make use of the vowel [ɑ] as a polar point representing maximum compactness. Similarly the vowel [i] could be adopted as the maximally acute vowel. This choice of physical references leads to an unavoidable interdependency between the parameters utilized for the two basic features, that is, a variation along the axis of one parameter implies some degree of variation along the axis of the other parameter. The maximally acute sound becomes the maximally diffuse, i.e., minimally compact sound. This is, however, of no concern from the linguistic point of view, since once the acute phonemes (front vowels) have been separated from the grave phonemes (back vowels), there is no point in a further use of the grave/acute opposition. Compactness (degree of opening) on the articulatory level is utilized both within front and back vowels for a further subdivision.

From an engineering point of view it seems preferable to extend the orthogonality principle to the physical criteria, i.e., to search for independent spectral attributes and parameters (see also the discussion in *Part III*). One example of an orthogonal set of parameters is increasing F_1 for increasing compactness and decreasing F_2 for increasing gravity. The translation of formant frequency shifts to changes of spectral form may be undertaken by means of rule 1) above for both features, as far as essentials are concerned. It follows that compactness defined by F_1 alone is mainly manifested by an increasing intensity level in all parts of the spectrum. In addition, it follows from rule 2) that the increase in the level of the valley between F_1 and F_2 is especially apparent providing F_2 is reasonably close to F_1 . The essentials of the gravity feature when F_2 is the criterion may also be derived from rule 1) above. A shift up of F_2 increases the intensity level of all formants positioned higher up on the frequency scale. A shift down of F_2 has the reverse effect. The increase in level of the first formant due to the approach of F_2 is noticeable only when F_2 comes fairly close to F_1 , as found from rule 2). Gravity/acuteness is thus manifested by a shift of the main spectral energy in a *lower/higher* direction on the frequency scale. It should be observed that this shift is not only due to a shift of the energy of the second formant but depends also on the intensity level changes in the other formants (Fant, 1956).

An alternative set of orthogonal parameters is $F_1 + F_2$ and $F_2 - F_1$ which results from a 45-degrees' rotation of the F_1 versus F_2 plane. The parameter $F_2 - F_1$, i.e., the frequency spacing between the second and the first formants of a voiced sound

represents the degree of spread of the spectral energy. The parameter $\frac{1}{2}(F_1 + F_2)$ is an approximative measure of the *center of gravity* of the spectrum. The limiting condition of a small *formant spread* $F_2 - F_1$ is, according to rule 2), a one-formant sound and the limiting conditions of a small $\frac{1}{2}(F_1 + F_2)$ is, according to rule 1), a high degree of gravity, or a very low center of gravity to use the specific terminology. As discussed in *Part III*, this set of parameters has some advantages over that of F_1 and F_2 but should not be confused with the compactness and gravity features. The articulatory correlate to a low/high $F_2 - F_1$ is tongue retraction versus tongue advancement, the palatal position possessing the maximum $F_2 - F_1$. The main articulatory correlate to a low $F_1 + F_2$ is lip-rounding, but tongue-backing contributes also.

1.33 Pole-Zero Decomposition of Consonants

As implied by *Eq. 1.23-16* and *1.23-17*, the effect of coupling on the nasal cavities or the effect of a finite coupling on the subglottal tract, i.e., the trachea and the lungs, can be taken into account by an additional factor of poles and zeros. The oral system function $H(f)$ of *Eq. 1.3-1* is thus multiplied by a system function $N(f)$ containing the distortion due to the shunting cavity system. In the case of nasal coupling the conjugate poles of $N(f)$ are the nasal formants. Each such additional formant is paired with an anti-resonance, i.e., a conjugate zero in $N(f)$. The contribution to the frequency characteristics of such a pair is the product of a resonance $|N_{pn}(f)|$ and an anti-resonance curve $|N_{zn}(f)|$. Assuming small losses as in *Eq. 1.3-7*,

$$|N_n(f)| = |N_{pn}(f)| |N_{zn}(f)| = \left[\frac{(1-x_{zn}^2)^2 + x_{zn}^2/Q_{zn}^2}{(1-x_{pn}^2)^2 + x_{pn}^2/Q_{pn}^2} \right]^{\frac{1}{2}}, \quad (1.3-12)$$

where

$$\begin{aligned} x_{pn} &= f/F_{pn}; & Q_{pn} &= F_{pn}/B_{pn}; \\ x_{zn} &= f/F_{zn}; & Q_{zn} &= F_{zn}/B_{zn}. \end{aligned}$$

The nasal pole is thus specified by its frequency F_{pn} and its bandwidth B_{pn} . Similarly, the zero is specified by F_{zn} and B_{zn} .

Obviously, if the zero has the same frequency and bandwidth as the pole, the product $|N_{pn}(f) N_{zn}(f)|$ obtains the value unity, i.e., there is complete cancellation. If there is only a small difference in frequency or bandwidth or both, the net effect will still be a smooth curve except for a local small maximum or minimum or both in the neighborhood of the critical frequencies. This has been exemplified in *Fig. 1.3-7* showing the contribution $N_n(f)$ from a pole and a zero that differ in frequency, but have the same bandwidth. Besides the hump plus valley, it can be seen that the level of the curve at high frequencies approaches a limiting value which according to *Eq. 1.3-12* is

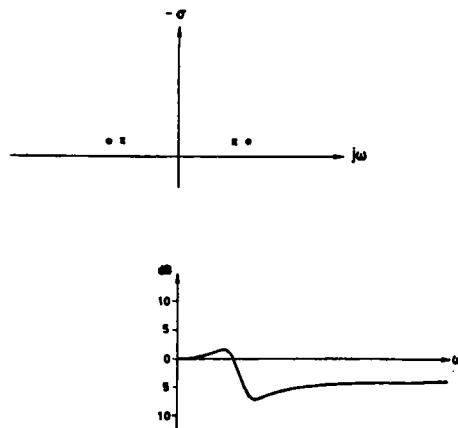


Fig. 1.3-7. Complex frequency representation of a pole-zero pair and its contribution to a frequency response curve. As the pole and zero approach they cancel each other.

$$20 \log_{10} |N_{pn}(f)| |N_{zn}(f)| \simeq 20 \log_{10} (F_{pn}/F_{zn})^2 \text{ dB.} \quad (1.3-13)$$

$$f >> F_{pn}.$$

The general shape of a vowel spectrum cannot change radically by a distortion of this kind. As the coupling to the shunting cavity system increases, the pole and zero within each pair separate and cause a more prominent spectral effect. In addition, some of the poles of the *oral* system $H(f)$ become somewhat shifted in frequency due to the tuning from both cavity systems. These effects are described in more detail in connection with the calculations on nasalized vowels; *Section 2.43*.

Statistically, the average spacing between formants from a cavity system of total length l must be $c/2l$ where c is the velocity of sound. Thus the average spacing between formants of oral origin, i.e., those of the F -pattern, is approximately 1000 c/s assuming a vocal tract 17.5 cm long. Similarly the average spacing between additional formants of nasal origin is also $c/2l_n$ where l_n is the total length of the nasal passages from the uvula to the outlet at the nostrils. Also the average spacing between the zeros must be $c/2l_b$ where l_b is the total axial length of the shunting system. The poles and zeros of the shunting cavity must alternate along the frequency scale.

The same general reasoning holds for the logical separation of the influence from cavities in front of and behind the source of fricative or stop sounds. The average spacing of zeros is $c/2l_b$ where l_b is the length of the cavity system from the source to the glottis. If front and back cavities are very well separated, i.e., with small coupling, each of the zeros of the back cavities will nearly coincide with a corresponding pole, thus effectively reducing the level of the formant associated with the pole, that is, one of the back cavity formants. Under these conditions of small coupling, the vocal tract filter function can approximately be specified by the formants of the front cavity system alone. The front cavity system generally includes a narrow passage

at the articulatory constriction. The degree of coupling to the back cavities is frequency dependent so that the coupling gradually increases with increasing frequency. However, the main shape of the vocal tract filter function may still be essentially conditioned by the front cavity resonances, and the back cavity formants, or rather the formants of the total vocal tract, can often be seen as a fine structure within the broader front cavity formant regions. In these instances the formant density, i.e., the average frequency spacing between formants can be used as a criterion for relating formants to the total vocal tract cavity structure or to the front part only. These effects are exemplified in *Section 2.63*.

1.4 THE F-PATTERNS OF COMPOUND TUBE RESONATORS AND HORNS

1.41 *The Twin-Tube Resonator. The Effect of Lip-Rounding*

The graphical procedure for determining the resonance frequencies of a twin-tube resonator is illustrated by *Fig. 1.4-1*. At the boundary between the front section 1 and the back section 2, the sum of the reactances looking right and left shall be zero:

$$Z_1 \operatorname{tg} \omega l_1/c = Z_2 \operatorname{cot} \omega l_2/c. \quad (1.4-1)$$

The resonance frequencies are found at the intersections of these two curves. If a sound source were inserted between the two sections, the zeros would occur at the frequencies of infinite back cavity impedance.

In the absence of any front section, the resonances occur at the zero crossings of the back section reactance curve. However, there is always an end correction to be taken into account. This inductance element plus a short, not-too-narrow, front section causes a linear decrease of the resonance frequencies by the factor $(l_2 + A_2 l_{1e}/A_1)/l_{2e}$, where the effective lengths $l_{1e} = (l_1 + l_{1t})$ and $l_{2e} = (l_2 + l_{2t})$ include the usual end corrections. The impedance of the front section passes through infinity at the frequency where

$$\operatorname{cot} \omega l_1/c = \omega l_{1t}/c, \quad (1.4-2)$$

or approximately at $f = c/4l_{1e}$.

Independent of the actual cavity configuration of the vocal tract, the input reactance, as seen from the lips, must be constantly rising with frequency except for the infinity discontinuities. The negative of this function must be constantly falling. The effect of prolonging the vocal tract or adding an extra section at the lip end must then by necessity be a shift down in all resonance frequencies, even those above the quarter-wavelength frequency $f = c/4l_{1e}$ of the added section. A decrease of the cross-sectional area of the lip section, everything else held constant, will evidently cause the resonance frequencies to approach the frequencies of infinite impedance of the tract behind the lips. This means a lowering of all resonance frequencies that lie below

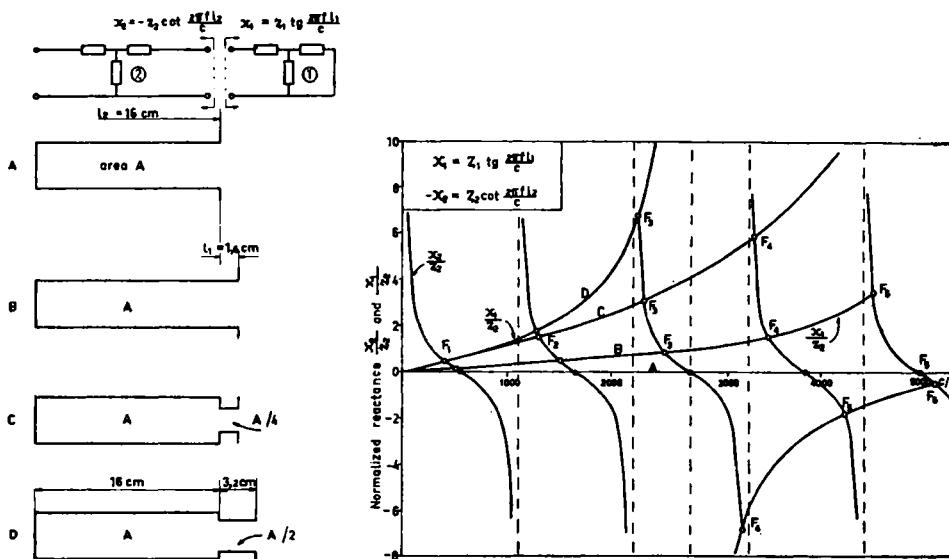


Fig. 1.4-1. The effect of prolonging the front part of a resonator or of decreasing its area illustrated for the case of a single tube resonator. Formants occur at frequencies where $X_1 = -X_2$, i.e., at the intersection between the reactance curves for the lip resonator and the main resonator system. An increased prolongation or rounding of the lip cavity will have the effect of lowering all formant frequencies that lie below a limiting frequency defined by a quarter-wavelength resonance of the lip cavity. This rule holds for any arbitrary configuration of the main resonator.

the quarter-wavelength resonance of the lip section and a slight increase in those above this critical frequency; Eq. 1.4-2.

These are the acoustic correlates to the phonetic term *rounding*, which can be either a protrusion of the lips or a decrease of their area, or both. A quantitative measure of lip-rounding can apparently be specified by the ratio of the effective length l_{1e} to the area A_1 of the lip section.¹ In the frequency range of the first four formants, the effect of lip-rounding will apparently be the same—a lowering of all resonance frequencies. Those that are largely dependent on the cavity immediately behind the lips will be influenced considerably more than those that correspond to standing wave phenomena in the pharynx cavity.

Similarly, the effect of adding an extra section at the far end of the system, i.e., at the bottom of the pharynx cavity, will also be to lower the position of all resonances whose frequencies lie below the critical value $c/4l_g$, where l_g is the length of the added section. This change affects the resonances of the pharynx cavity more than other resonances. The physiological counterpart of this cavity change is the lowering of the whole larynx.

¹ Stevens and House (1955, 1956) utilize the inverse measure A/l .

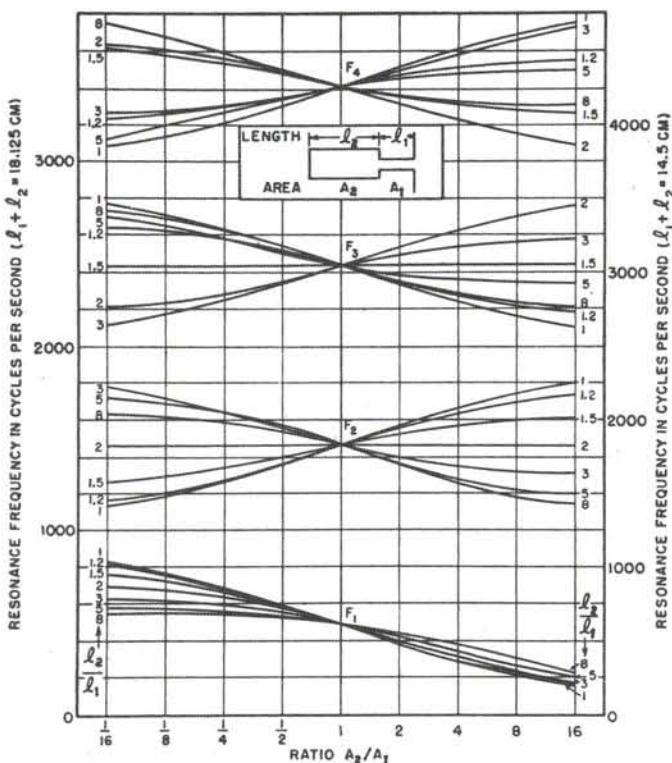


Fig. 1.4-2. Nomogram of the twin-tube resonator. The resonance frequencies F_1 , F_2 , F_3 , and F_4 can be read off from data on the ratio of back tube cross-section area A_2 to front tube cross-section area A_1 and the length ratio l_2/l_1 of the two tubes. The ordinates correspond to two different values of total length 14.5 cm and 18.125 cm. For other values of total length, the resonance frequencies can be obtained by the inverse proportionality rule. The contribution from the reactive load of radiation has been neglected but can approximately be taken into account by defining the length l_1 as including $0.8 \sqrt{A_1/\pi}$, where A_1 is the cross-sectional area at the opening of the front tube.

The nomogram of the twin-tube resonator, Fig. 1.4-2, can be used for an estimate of the frequencies of the first four resonances, given the lengths and the areas of the two tubes. The length l_1 thus includes both the external and the internal end corrections of the front section.

The nomogram covers length ratios l_2/l_1 from 1 to 8 and area ratios A_2/A_1 from 1/16 to 16. From the resonance conditions

$$\frac{A_2}{A_1} \operatorname{tg} \frac{\omega l_1}{c} \operatorname{tg} \frac{\omega l_2}{c} = 1, \quad (1.4-3)$$

it can be seen that the lengths l_1 and l_2 are interchangeable. The ratio l_1/l_2 thus provides exactly the same resonance frequencies as the inverse ratio, i.e., l_2/l_1 . When

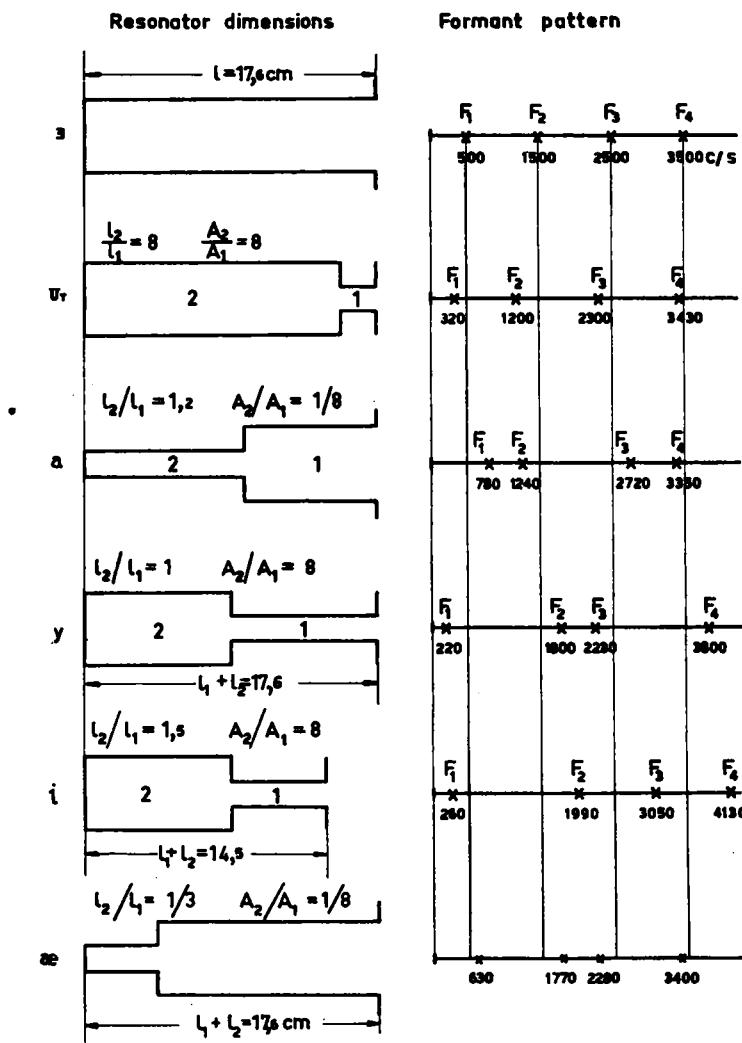


Fig. 1.4-3. Twin-tube resonators that provide formant patterns resembling those of some vowels. The open tube resonator with formants spaced at odd multiples of 500 c/s is the neutral reference. A synthetic vowel with this formant pattern is denoted by the symbol [ə]. All other resonators and their corresponding formant patterns with the exception of the second one from the top, are fair approximations of the essential articulatory and acoustic characteristics of the vowels indicated.

l_2/l_1 is smaller than unity, the inverse value can accordingly be used. This is an indication of the possibility of compensatory forms of articulation. In this class of idealized resonators the compensation is perfect except for the differences due to the finite radiation impedance. The effect of narrowing the front part of the resonator is thus the same as widening a part with the same length at the opposite end of the

resonator system. Conversely, the effect of widening the front part of the resonator is exactly the same as the effect of narrowing a section with the same length at the closed end, providing the area ratio is the same.

From the nomogram it is apparent that a decrease of the ratio of back to front area A_2/A_1 will always be followed by an increase in the frequency of the first formant. Providing the front and back sections do not differ appreciably in length, i.e., $2 > l_2/l_1 > \frac{1}{2}$, the variation of the A_2/A_1 ratio from large to small values is also followed by a decrease in the frequency of the second resonance. Something similar to this happens in the series of vowels [i] [e] [ɛ] [æ] [a] in which the increasing compactness is followed by an increasing gravity; see the discussion in *Section 1.32*.

The maximum rate of increase, of the ratio F_2/F_1 with increasing area ratio A_2/A_1 , is found when $l_1 = l_2$. For $A_2 > A_1$ there is then also a maximum convergency of F_2 and F_3 . It should be observed that the traditional phonetic terms *open/close* are references to relative elevations of the *highest point* of the tongue. A vowel like [a] is generally considered as an open front vowel. The only important aspect for the articulation of the [a] is, however, the relatively narrow pharynx, as pointed out in the more phonetically oriented discussion in *Section 2.33*. A few twin-tube cavity configurations and corresponding resonance frequency patterns are shown in *Fig. 1.4-3*. The phonetic transcriptions included are based on formant frequencies, but there is also some fundamental resemblance in terms of cavity structure.

The twin-tube model of [a], consisting of a narrow back tube and a wide front tube, is equally representative as, or even more so, than the double Helmholtz resonator discussed in the previous section. Observe the quite inverse cavity shape for [a] compared with [y] or [i] and the resulting opposition in terms of the resonance frequency pattern, i.e., in the formant pattern. A shortening of the front part of [y] provides the necessary adjustment towards the higher F_2 and F_3 of [i].

An increase of the front section area A_1 of [i] would have caused a higher F_1 , i.e., a sound shift towards [e]. The [æ]-model may look strange in view of its constricted back part. This is, however, a typical configuration of the articulation of very open front vowels as seen from X-ray pictures. The front orifice at the palatum is more or less an illusion for this sound. Observe the possibility of articulatory compensations, i.e., that the same formant pattern can also be obtained with a homogeneous tube that is widened at its front end. These two modes of articulation may occur in combination.

1.42 Horns as Single Resonators and Connecting Sections

The articulatory correlates to an increase in the frequency of the first formant, other formants held essentially constant, can be described in terms of a single horn representing the entire vocal tract as shown by *Fig. 1.4-4*. The $l = 17.6\text{ cm}$ effective length of a single tube provides the neutral reference pattern of formants at odd integers of 500 c/s . A horn of the same length and of a gradually increasing cross-

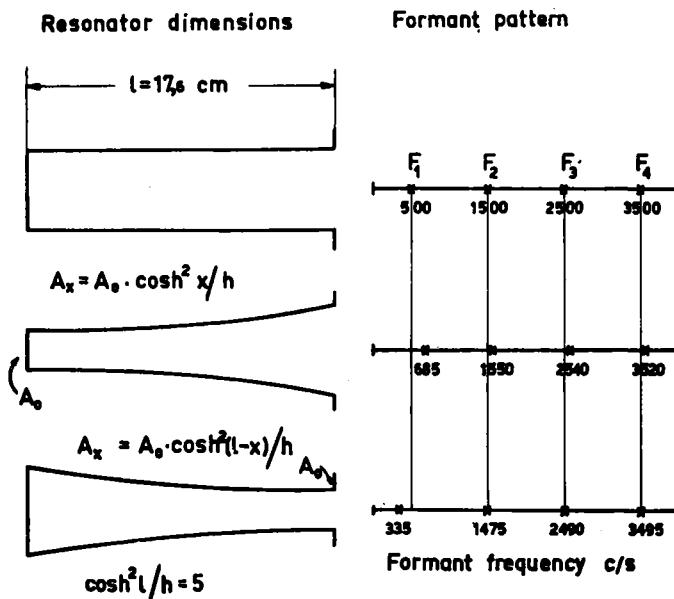


Fig. 1.4-4. The simplest articulatory correlate to a variation in F_1 rather than in other frequencies of the F-pattern. The data have been calculated from the theory of horn resonators. An increasing area of the horn from the throat to the mouth results in a high F_1 and a decreasing area to a low F_1 . The higher formants do not differ appreciably from the neutral reference because of the smooth transitions and thus the freedom from reflections.

sectional area has an almost identical pattern with regard to the higher resonances. The frequency of the first formant is, however, markedly increased. Similarly a horn of continuously decreasing area and the same length has a lower first formant. The data of Fig. 1.4-4 were calculated numerically on the assumption of a catenoidal shape and with the aid of the equivalent circuit data of Section 1.21.

The use of a horn as a part of the cavity system approximating the vocal tract is exemplified in Fig. 1.4-5 illustrating the graphical determination of the resonance frequencies of a model of the vowel [i]. A single equivalent circuit can be utilized for the whole mouth section including the palatal constriction, the front cavity, and the smooth termination to the pharynx cavity. The resonance frequencies are found at the intersection of the rising curve and the falling curve, the former representing the reactance of the front part as seen from the top of the pharynx cavity. The values $F_1 = 250 \text{ c/s}$, $F_2 = 2000 \text{ c/s}$, and $F_3 = 3000 \text{ c/s}$ are found in agreement with spectrographic data and the data from Fig. 1.4-3. The horn representation of the mouth cavity is, however, more natural than the twin-tube approximation. As a further step towards a more natural configuration a larynx tube could be included, as shown by model b) of Fig. 1.4-5. The effect on the [i]-spectrum of this addition at constant

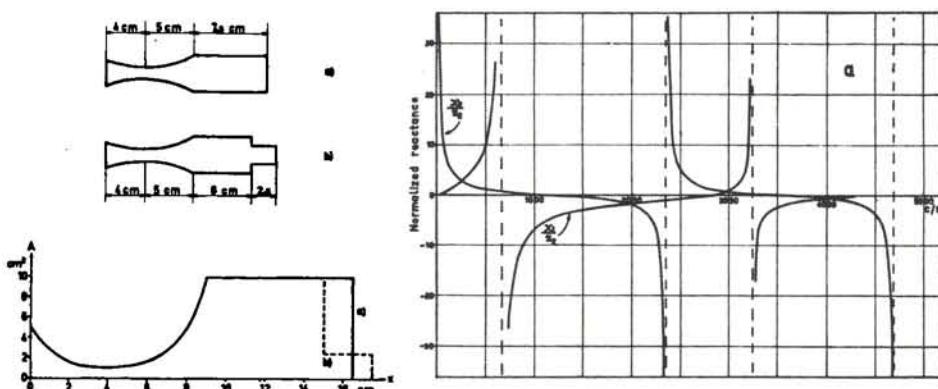


Fig. 1.4-5. Simple resonator model of the vowel [i] based on a catenoidal horn to represent the front, middle, and back part of the mouth cavity and a cylindrical cavity to represent the pharynx. A larynx tube is included in the alternative b). The procedure for calculation of formant frequencies of the simpler model, a), is shown by the reactance diagram. X_1 is the reactance of the mouth as seen from the top of the pharyngeal cavity and X_2 is the reactance of the pharyngeal cavity at the same place. Formants occur at frequencies where $X_1 + X_2 = 0$.

total length of the vocal tract is to cause an increase in the frequency of the second and fourth formants. A further discussion of the larynx and associated cavities will be given in *Chapter 2.2*.

The inverse cavity relations of [i] as compared to [a] can conveniently be described by means of the horn representation. If the mouth opening and the glottis end are interchanged in the model of Fig. 1.4-5 so that $x = 0$ represents the closed end and $x = 16.5$ the open end, the result will be a perfect [a]-model more natural in appearance than the twin-tube model, Fig. 1.4-3. In most speech sounds the parts of the vocal tract containing the tongue constriction and the adjacent transitional regions connecting the constriction to a front and a back cavity may to a fair approximation be simulated by a single catenoidal horn. This is the fundamental feature of one of the three-parameter vocal tract models to be described in *Section 1.43*.

In conformity with the discussions in *Sections 1.11, A.2, 1.31, and 1.21*, the theory of horns cannot provide any information on the structure of sound spectra in excess of that available from the vocal tract system function in terms of a complete pole-zero specification. It is generally simpler to relate the energy output of the vocal tract to the frequency of the first formant than to the cavity shape. A large lip-opening alone is not a necessary guarantee for a high sound level since the tongue articulation also affects the frequency of the first formant. As shown in previous sections, the radiation of sound energy is not increased by an increase of the area of the radiating surface, the formant frequencies held constant. It is thus not possible to increase the acoustic power of a vowel by an articulatory effort (filter function) alone without changing the phonetic quality of the vowel, as specified by the formant frequencies.

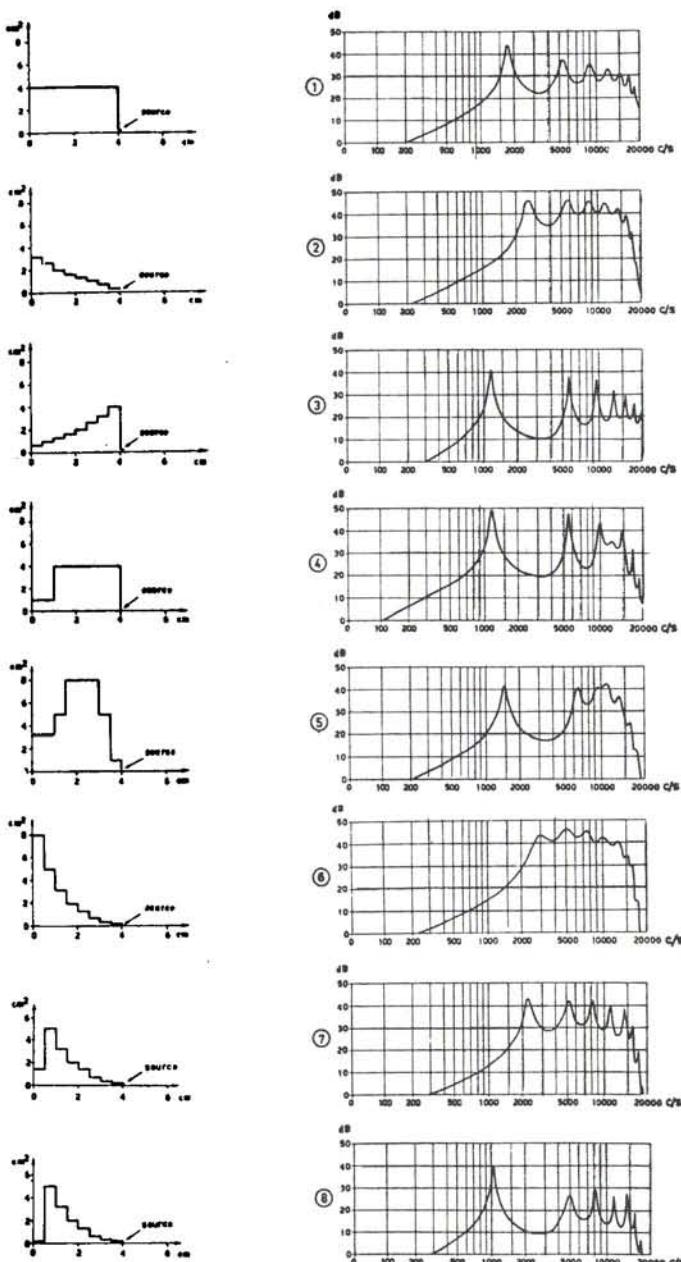


Fig. 1.4-6. Frequency response of some simple resonators of length 4 cm in terms of the ratio of radiated pressure to the volume velocity of a constant current, constant spectrum level source, calculated with the aid of LEA. $X = 0$ cm on the x-axis of the area functions corresponds to the radiating surface of the lips.

A horn resonator can be simulated to any desired degree of accuracy by a stepwise approximation of the area function, i.e., in terms of a set of successive short cylindrical tubes. The acoustic effects of varying the shape of simple resonators that simulate horns are demonstrated in *Fig. 1.4-6*. The resonators were all given the length of 4 cm which is of the same order of magnitude as the length of the front cavity of palatal and retroflex constrictives. The calculations were carried out with the electrical line analog, *LEA*, see further *Chapter 2.2*. *LEA* was fed from a high impedance resistive source of constant level spectrum. Halfway up to the frequency of the first formant the spectrum envelope must thus rise 6 dB/octave owing to the ω -proportionality of radiation.

The first example of *Fig. 1.4-6* is the single tube resonator providing a fundamental resonance at $c/4l_e = 1750$ c/s and higher resonances approximately at the odd integers of this frequency, i.e., 5250 c/s and 8750 c/s. It should be observed that the frequency scale of these spectrum curves and all other curves traced from the *LEA* analog are logarithmic.

In comparison with the neutral reference, model No. 1, model No. 2, and even more model No. 6, display the typical horn effect of increase in the frequency of the first formant, and further the larger damping of the formants, i.e., the increase in their bandwidths and also the increased level between formants which tends to flatten out the peaks. When the mouth-opening area of a horn is increased by an increase in the taper, the spectrum level at the peaks of formants located well above the horn cutoff frequency is affected by two opposing factors. One is the transformer effect of the horn increasing the output volume velocity in proportion to the square root of the area of the mouth-opening. The other is the increased radiation damping which attenuates the peak levels in proportion to the mouth-opening area and frequency squared. The net effect of the area increase when radiation resistance is considered to be small, constant, and to constitute the main dissipative element, is that peak levels decrease and valley levels increase in proportion to the square root of the mouth-opening area.

The narrow peaked formants of the inverse horn are apparent. The effect is similar when lip-rounding is superimposed on a horn or a tube, as seen from models 4, 7, and 8 which also show the lowered position of the first formant frequency. The cavity shape of model No. 5 could pertain to the front cavity of a consonant [k]. Besides the fundamental formant at 1500 c/s there is a formant at 6500 c/s and a formant group at frequencies around 10000 c/s. It should be observed that the level of high frequency formants is exaggerated in *Fig. 1.4-6* as compared with natural speech owing to the particular source characteristics chosen for the calculations; see further *Chapter 2.6*.

1.43 Three-Parameter Models Approximating the Vocal Tract

The twin-tube resonator is not suited for the representation of vocal tract cavity

configurations with a marked tongue constriction. At least three sections are needed, and also a fourth, if any amount of lip-rounding is to be taken into account. Irrespective of the type of resonator sections used, cylindrical, or horn-shaped, it is possible to limit the number of articulatory variables to 3. In phonetic terminology these may be called the place of articulation, the degree of opening, and the amount of lip-rounding. However, it is then understood that the place of articulation refers to the effective center of the main tongue constriction, that is, the part of maximum narrowing in the vocal cavities caused by the proximity of the tongue to a palatal, velar, or pharyngeal area. The degree of opening is the minimum cross-sectional area at this main constriction. The degree of lip-rounding is defined by the ratio l/A of the mouth-opening, that is, the ratio of lip section length to its mean area. In the case of delabialized sounds produced with the upper and lower incisors not very far apart, it is necessary to include the teeth passage in the l/A estimate. As an alternative lip parameter the inverse of the l/A ratio, the conductivity index A/l , may be used.

The vocal tract model can be entirely specified by these three variables.² This is the approach followed by Stevens and House (1955, 1956) who have made extensive investigations with the aid of a model in which the cross-sectional area at the tongue constriction and at the nearest parts of the back and front cavities are defined by a parabolic area function. The total length of the compound system and the maximum area in the back and front cavities are kept constant. A similar model, based on a hyperbolic outline of the area function at the tongue constriction, has been investigated in connection with this work. A simpler model based on a cylindrical section for the tongue constriction has also been analyzed in some detail.

In spite of its less natural shape, it behaves similarly to the model with a horn-shaped tongue section, at least as far as the first three resonances are concerned. The cavity-resonance dependencies can be more readily analyzed in view of the more distinct configurations. They will therefore be treated first.

A. MODELS COMPOSED OF CYLINDRICAL SECTIONS ONLY

Fig. 1.4-7 shows a few 3-section models and their associated F-patterns, that is, their set of resonance frequencies—in technical terminology, their mode spectra. The models are chosen to exemplify the effects of a variation in the position of the tongue constriction and the total length of the system and in particular to show how the fundamental resonance of the cavity next to the mouth-opening may be associated with any of the first four formants. The first model from the top is the single tube resonator providing resonances at odd integers of 500 c/s. The fundamental resonance

² This representation is similar to but not identical with the classical articulatory description in terms of the position of the highest point of the tongue. A velar or postpalatal location of this latter point may be combined with a pharyngeal place of maximum narrowing, as in the vowels [ɑ] and [a]. In classical terminology [ɑ] is open with regard to the mouth cavity. In terms of the area of the effective constriction, [ɑ] may have the same degree of opening as [i]; see further the discussion in *Section 2.32*.

Formant pattern

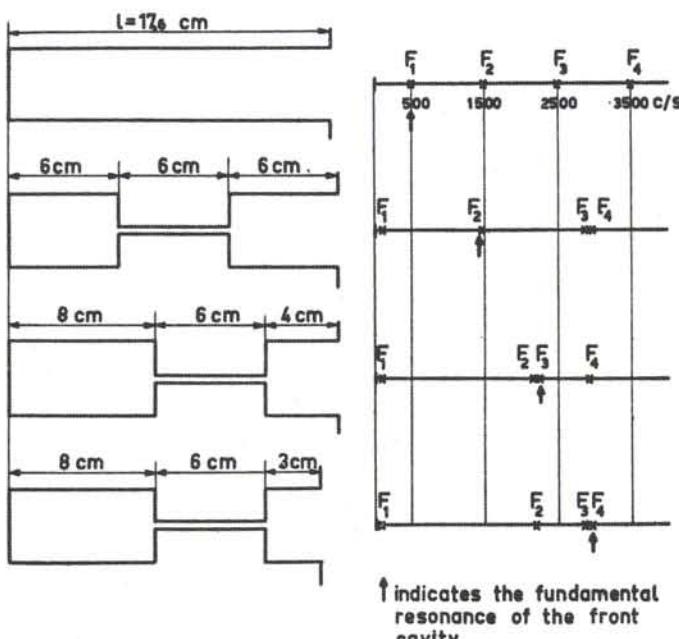
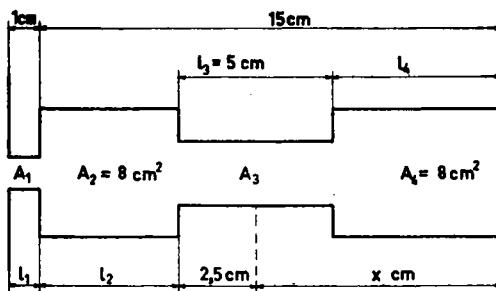


Fig. 1.4-7. Three-tube resonator models and corresponding F-patterns. The dimensions have been chosen so that the frequency of the fundamental resonance of the front cavity corresponds to F_1 , F_2 , F_3 , and F_4 respectively. These idealized resonator systems show some essentials of the articulation of velar and palatal consonants.

of this model originates from a quarter-wavelength standing wave. A front cavity quarter-wavelength resonance is also the determinant of F_2 of the second model, F_3 of the third model, and F_4 of the fourth model.

The proximity of F_2 and F_3 in model No. 3 is due to the identical frequency position of the uncoupled quarter-wavelength resonance $c/4l_1$ of the front cavity and the half-wavelength resonance $c/2l_2$ of the back cavity under the condition $l_2 = 2l_1$. Owing to the finite coupling, the actual resonances cannot coincide. If the tongue hump is advanced slightly more, or if the front section is shortened as in model No. 4, there will result an F_3F_4 -proximity. In this state F_2 is almost entirely influenced by the half-wavelength resonance of the back cavity, and F_3 is determined by the half-wavelength resonance of the tongue constriction section, that is, the intermediate tube. Models Nos. 2, 3, and 4 apply to the production of the consonant [k] or [g] before [a], [æ], and [i] respectively. A superimposed lip-rounding on model No. 2 would effectively lower the frequency of the mouth formant as in [ku].

A detailed investigation has been performed on the basis of 3- and 4-section cylindrical models of this type. Contrary to the models of Fig. 1.4-7, the length of the front tube was defined as the physical length, i.e., the end correction length has not been included. A total length of 15 cm was chosen for the three-tube case representing



	Curve	A_1	l_1	total length	l_3 is decreased
$A_3 = 0,65$	1	8	0	15	for $2,5 > x > 12,5$ to keep total length constant.
	2	4	1	16	
	3	2	1	16	
	4	0,65	1	16	
	5	0,16	1	16	
$A_3 = 2,6$	1	8	0	15	
	2	4	1	16	
	3	2	1	16	
	4	0,65	1	16	
	5	0,16	1	16	

Fig. 1.4-8. Three-parameter vocal tract model based on four homogeneous tubes, simulating the lip section (mouth-opening), a front cavity, the tongue section, and a back cavity. Total length exclusive of lip section is 15 cm and the cross-sectional area of the two main cavities is held fixed at 8cm². The length l_3 of the tongue constriction section is 5 cm unless it is located in an extreme front or back position where either the front cavity or the back cavity disappears. The three parameters are (1) location and (2) cross-sectional area of the tongue passage and (3) the length over area ratio of the mouth-opening. The tabulated parameter data refer to the nomograms of Fig. 1.4-9.

unrounded conditions. The additional lip section was given a constant length of one centimeter. The cross-sectional area of the tubes on both sides of the tongue constriction tube of this model is 8 cm², as may be seen from Fig. 1.4-8. The length of this latter tube section is kept constant at 5 cm, provided it is terminated by cavities on both sides, but its place and cross-sectional area are varied. The place variation is carried far enough towards the front and back ends of the coordinate scale to allow for configurations where the length l_2 or l_4 vanishes and the length l_3 of the tongue hump section varies in order to satisfy the condition of a constant total length of the system. At extreme front or back tongue position coordinates, the resonator system is a single tube of 15 cm length plus the associated lip section of variable area.

Results from calculations performed with the line analog LEA are shown in Fig. 1.4-9a and b. Lip section areas of 4, 1, 0.65, and 0.16 cm² are represented by the curves 2, 3, 4, and 5 respectively. Curve 1 pertains to the case where the lip section is removed. The two separate diagrams a and b pertain to tongue section areas of 0.65 cm² and 2.6 cm² respectively.

The effect of varying the place of constriction as displayed by the curves is qualitatively the same for the two different tongue section areas. The differences can be described as a larger deviation from the neutral, single tube F-pattern when the

tongue passage is narrower. The essential observable pattern changes caused by a shift of the tongue section from a back to a front position are from a high F_1 -, low F_2 -position as in back vowels, to a position of F_2 closer to F_3 as in high front vowels. The place of articulation providing minimum F_2 is found 1-2 cm in front of the coordinate of maximum F_1 . Similarly, the maximum of F_2 is located at approximately 0.5 cm in front of a coordinate of minimum F_3 . This F_2F_3 -proximity occurs at a region of mid-palatal tongue position coordinates.

There is also an F_2 -maximum and F_3 -minimum posterior to the F_2 -minimum, i.e., in a laryngeal region of articulation. In-between the two F_2F_3 -proximity regions there is an F_3 -maximum and another in front of the mid-palatal region. The extent of the F_3 -variations, however, is not very great.

If an advance of the tongue causes a resonance frequency to rise it can be concluded that the resonance is mainly influenced by a cavity of decreasing length, in this instance, the front cavity. Similarly, it must be concluded that those portions of the nomogram curves that slope upwards from left to right, i.e., as a result of tongue retraction, are mainly influenced by the back cavity. It is thus apparent that a formant is equally dependent on two separate cavities at the maxima and minima points and thus changes cavity affiliation at these locations of the tongue passage. Increasing lip-rounding has the effect of lowering all resonance frequencies. The effect is apparently largest on those parts of the curves that represent front cavity resonances. Lip-rounding also shifts the proximity regions forward.

As the tongue is shifted past a proximity region the higher formant will rise in the same direction as the lower formant before its maximum. The common basis of this apparent continuity is an uncoupled resonance frequency of the cavity that is reduced in length owing to the tongue shift. Two uncoupled resonance frequencies cross in the center of a proximity region, but two formant frequencies, i.e., resonance frequencies of the complete coupled system, cannot coincide unless the constriction is complete.

The uncoupled resonance frequencies of interest for the discussion of the origin of the first four resonances, i.e., F_1 , F_2 , F_3 , and F_4 , are the following:

I. Helmholtz resonances, Eq. A.31-4;

- a) The front resonator, composed of the cavity of length l_2 in front of the tongue passage, and the lip section orifice;
- b) The back resonator, containing the cavity of length l_4 behind the tongue passage, tuned by the tongue section.

II. Standing wave resonances in any cavity or section that is terminated differently at the two ends, that is, one end approximately closed and the other end approximately opened. The frequency of these quarter-wavelength resonances is $nc/4l$, where l is the length of the section and $n = 1, 3, 5$, etc.;

- a) The front cavity under the unrounded conditions, i.e., with the lip section removed;
- b) The tongue passage when the length l_2 is reduced to zero owing to extreme front positions of the tongue. The lip-opening must be much smaller than the

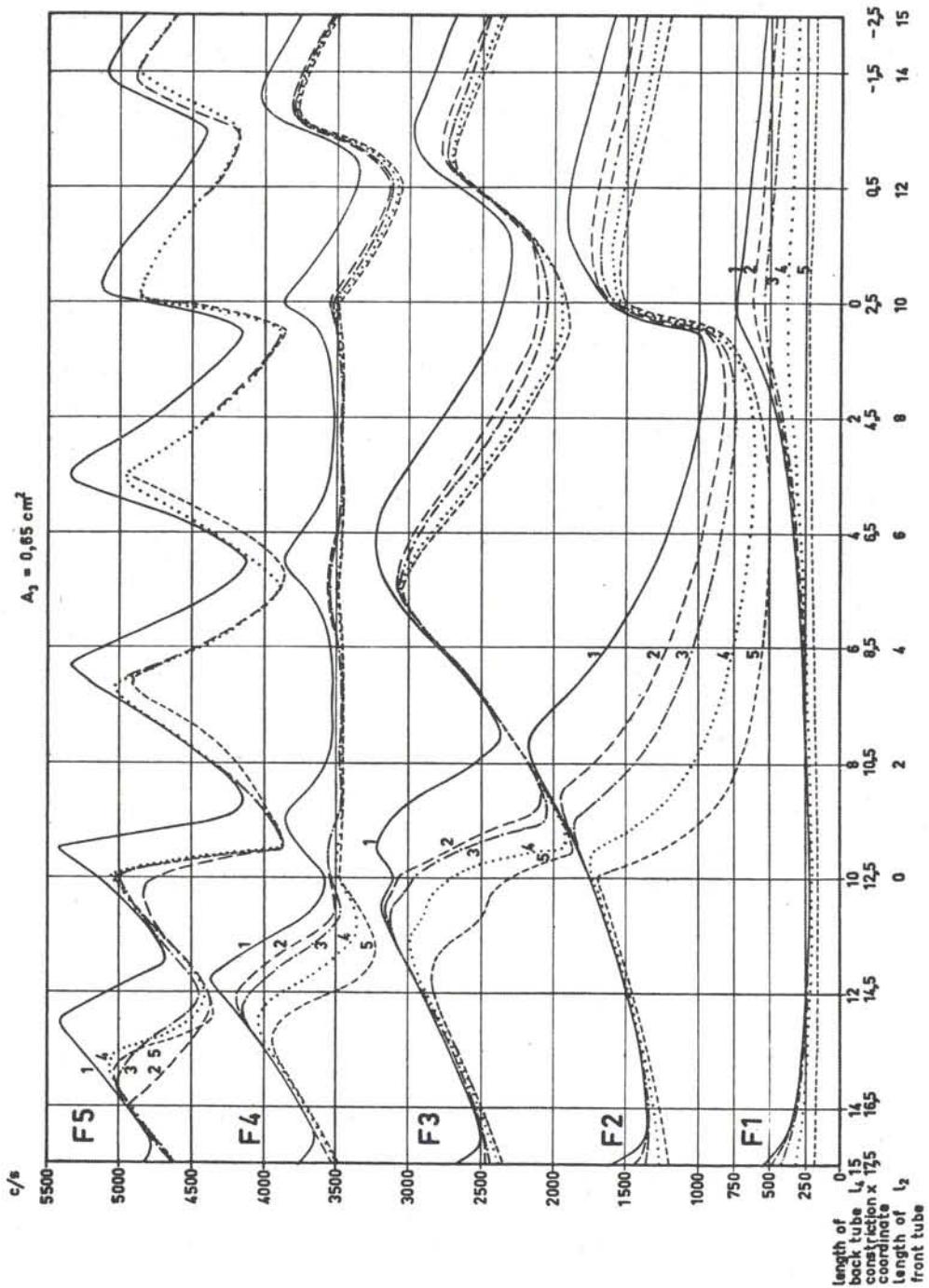
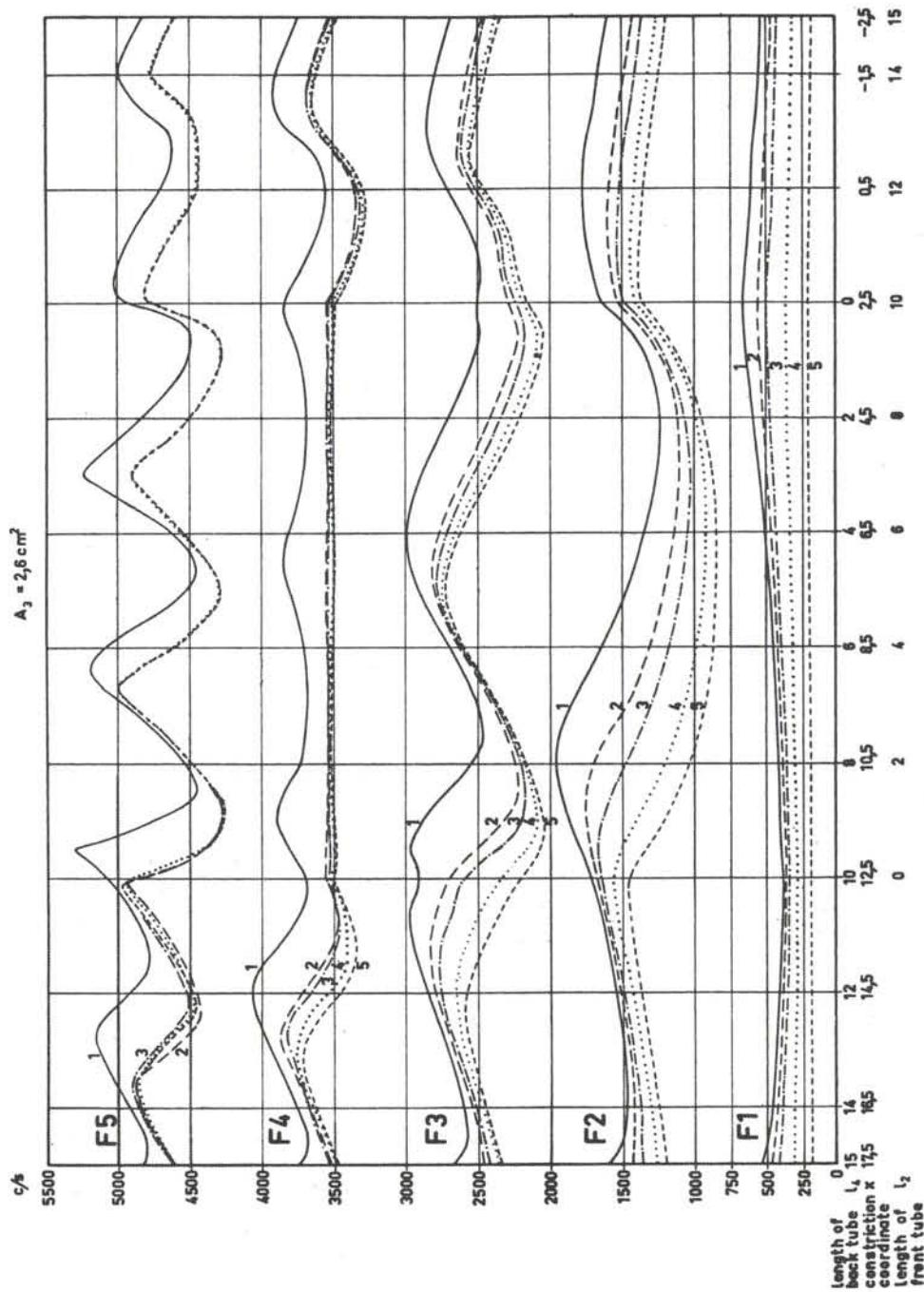


Fig. 1.4-9. Nomograms relating F_1 , F_2 , F_3 , F_4 , and F_5 of the three-parameter model Fig. 1.4-8 to the location of the tongue constriction: a) Tongue constriction cross-sectional area $A_3 = 0.65 \text{ cm}^2$,

Fig. 1.4-9. b) Tongue constriction cross-sectional area $A_3 = 2.6 \text{ cm}^2$.

- cross-sectional area of the tongue passage, which in turn should be definitely smaller than the area of the back cavity;
- c) The tongue passage when the length l_4 is reduced to zero owing to extreme retracted tongue positions.
- III. Standing wave resonances in any cavity or section that has equal terminating conditions at the two ends, i.e., either both ends approximately closed or both ends approximately open. The frequencies of these half-wavelength resonances are $nc/2l$, where $n = 1, 2, 3$ etc.;
- a) The front cavity under conditions of appreciable lip-rounding;
 - b) The back cavity;
 - c) The tongue passage when terminated by larger cavities at both ends.
- IV. Resonances of a character intermediate between any of the three main types I, II and III above, that is, one end of the resonating system half-open;
- a) The fundamental resonance of the front or the back cavity in a state intermediate between IIa and IIa;
 - b) Next higher resonance intermediate between the second resonance of type II, that is, of frequency $3c/4l$, and the first resonance of type III, i.e., of frequency $c/2l$. Similarly, the next resonance will occur between $5c/4l$ and $2c/2l$ but relatively closer to the latter, see further the graphical procedure illustrated by *Fig. 1.4-1*.

The cavity-resonance conditions can be studied in detail by following one of the curves of *Fig. 1.4-9a* from the far right end to the left end of a diagram. The data of *Fig. 1.4-9b* are not equally useful since the wider tongue passage makes all formants less dependent on one section of the system only. At constriction coordinates posterior to the F_1 -maximum, curve No. 1 of the F_1 -group in *Fig. 1.4-9a* is associated with the quarter-wavelength resonance of the front cavity. As lip-rounding increases within the series of curves 1, 2, 3, 4, 5, the front cavity resonance changes character from type IIa to IVa and finally to Ia. As the tongue constriction is shifted ahead of the F_1 -maximum, F_1 will be more dependent on the back cavity resonance. The region of divided dependency is, however, greater than for any other formant and extends over a larger part of the constriction coordinate range at least under labialized conditions. Ahead of the F_2 -minimum and up to the F_2 -maximum the fundamental resonance of the front cavity is associated with F_2 . At more advanced tongue positions it conditions F_3 and then F_4 , as already illustrated by the stepwise changes in *Fig. 1.4-7*.

It can be of some interest to follow the half-wavelength resonance of the back cavity, i.e., the resonance type IIIb. As the tongue constriction is shifted backwards past the coordinate of F_2 -maximum it becomes apparent that this resonance shifts from an association with F_2 to F_3 . The fundamental resonance of the back cavity, type Ib or IVa, will similarly shift from an F_1 - to an F_2 -affiliation as the tongue constriction is moved back past the F_2 -minimum.

The portion of the F_3 -curve between the intermediate F_3 -maximum and the posterior F_2F_3 -proximity is connected with the three-quarter-wavelength resonance of the front cavity in the unrounded state and approaches the two-quarter-wavelength resonance when lip-rounding increases, as stated in paragraph IVb above. At constriction coordinates from 0-1.5 cm the third resonance frequency, F_3 , coincides with the quarter-wavelength resonance of the constriction passage, as described in IIc above. At the very extreme back end and front end of the coordinate system, the whole model is merely a single tube with or without lip-rounding. In the unrounded state the resonances occur at odd integers of 540 c/s, i.e., the quarter-wavelength frequency corresponding to the basic length of 15 cm plus end correction, $0.8(8/\pi)^{1/2}$ cm. The maximum effect of lip-rounding is to lower one of these resonances by the quarter-wavelength frequency of the corresponding physical length 15 cm, that is, 590 c/s.

F_4 is essentially determined by the half-wavelength resonance of the tongue constriction and is thus fairly independent of any articulatory variable, except at a restricted frontal region where it is associated with the front cavity. F_5 is associated with various standing wave resonances of higher order. The data for F_6 are of no considerable phonetic interest since they are more indicative of this particular model than of natural speech. The same is partly true of F_4 . In real speech F_4 may be influenced to a smaller or a larger degree by the quarter-wavelength resonance of the larynx tube. A larynx tube, however, is not necessary for the creation of a fourth formant, see further *Section 2.32*.

If the vocal tract model behaved entirely like two coupled Helmholtz resonators, there would be only two resonances and only one proximity region, containing the maximum in the resonance of lower frequency and the minimum in the resonance of higher frequency. The coordinate of equal uncoupled resonance frequency of the back and front systems would come close to the center of the proximity region.

B. MODELS WITH HORN-SHAPED TONGUE SECTION

The three-parameter vocal tract model shown in *Fig. 1.4-10* is more useful than that of *Fig. 1.4-8* as a reference for estimates of formant frequencies from articulatory data, since it has a more natural shape. The internal constriction is simulated by a catenoidal horn. When calculations are performed with the aid of a vocal tract transmission line analog like *LEA*, it becomes necessary to perform a step wise approximation of the area function of the horn, as indicated in the figure. Numerical calculations can conveniently be based on the complete equivalent network described in *Section 1.21*.

Apart from the lip-rounding, all dimensions within the model are uniquely related to a specification of the center coordinate x_{min} of the horn and the cross-sectional area A_{min} at this point. The cross-sectional area of the horn increases symmetrically on both sides of the center coordinate.

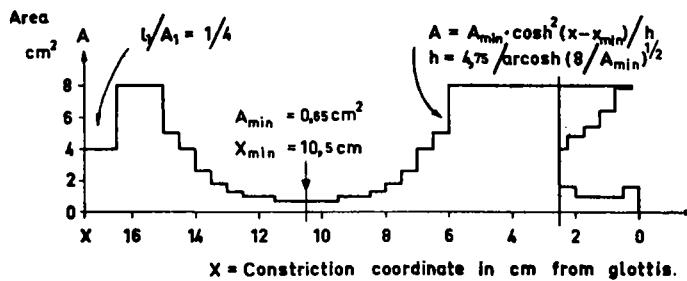


Fig. 1.4-10. Three-parameter vocal tract model based on a horn-shaped tongue section, the area function of which is stepwise approximated to fit the quantized area scale of the electrical line analog LEA. A larynx tube as well as the sinus piriformis cavities surrounding the larynx tube have been included as fixed cavities.

$$A_x = A_{\min} \cosh^2 (x - x_{\min})/h, \quad (1.4-4)$$

where A_x is the value of the area function at the coordinate x , and h is a constant which pertains to the taper of the horn and is determined by the condition that the area shall always reach the limiting value 8 cm^2 at $\pm 4.75 \text{ cm}$ from the center coordinate and then remain constant at this value in both the front and the back cavity.

$$h/4.75 = \operatorname{arcosh}(8/A_{\min})^{1/2}. \quad (1.4-5)$$

The fixed larynx tube and the shunting cavity simulating the sinus piriformis are held constant in all calculations and thus always add to the back cavity. The axial distance from the bottom of the larynx cavity to the front end at the lips is 16.5 cm in the completely delabialized state and 17.5 cm with the extra lip section of length $l_1 = 1 \text{ cm}$ added. An effective length of the constricted tongue section may be defined from the length of a cylindrical tube of area A_{\min} that has the same low frequency impedance as the horn. From Eq. 1.4-4 this length l_e is

$$l_e = A_{\min} \int_{x_{\min}-4.75}^{x_{\min}+4.75} \frac{dx}{A(x)} = 2h \cdot \operatorname{tgh}(4.75/h). \quad (1.4-6)$$

The following tabulation shows the effective length as a function of the minimum area A_{\min}

A_{\min} cm^2	l_e cm
0.16	3.6
0.32	4.1
0.64	4.7
2.0	6.3
4.0	7.6
8.0	9.5

The effects of a variation of the *place of articulation* x_{min} at a fixed small degree of *opening* $A_{min} = 0.65 \text{ cm}^2$ under five different conditions of lip-rounding are displayed in *Fig. 1.4-11a* which is thus comparable with *Fig. 1.4-9a*. Similarly, *Fig. 1.4-11b* pertaining to $A_{min} = 2.6 \text{ cm}^2$, is to be compared with *Fig. 1.4-9b*. It is found that the main topology of the nomograms from the two models, e.g., the coordinates of the proximity regions, is the same. They also compare well with the data of Stevens and House (1955, 1956). This agreement implies that the formant-cavity relations are not changed radically when the cylindrical tongue constriction section is substituted for a horn-shaped section.

The effect of a variation of cross-sectional area of the tongue section as a function of its location is shown in *Fig. 1.4-11c*. The lips are in the maximally unrounded state.

An increase of the constriction area has the effect of increasing F_1 , providing the constriction is located in the front half of the model. In the back vowel region the relation is more complex. There is an optimum constriction area providing maximally high F_1 . At the maximum area value $A_{min} = 8$ where the constriction vanishes altogether, F_1 approaches the neutral position, and at extreme small areas F_1 approaches zero value. The F_1 -variations in the back vowel region are smaller than the F_2 -variations. The major effect of a *centralization* of a back vowel is accordingly a shift up in F_2 .

The general rule is that a reduction of area contrasts within the vocal tract shifts the F-pattern towards that of a neutral vowel. It may thus be seen that F_2 increases with increasing constriction area, provided the center of the constriction is located closer to the glottis than to the lips. When the place of articulation is in the anterior half of the model, the second formant is found to approach the neutral position by a downward shift associated with increasing opening.

At a constriction coordinate in the middle of the vocal tract model there is thus no effect on F_2 from a variation of A_{min} . It might seem more natural to expect this coordinate of constant F_2 to appear at a distance of one-third of the total length of the cavity model from the lips at which place a single tube resonator has a velocity minimum node for F_2 . However, this would imply that a thin plate was utilized as an internal constriction. In our two models and also in that of Stevens and House (1955, 1956), the effective length of the tongue section is of the order of one-third of the total length. The resonance conditions for a 3-section cylindrical system composed of three tubes of equal length l can be approximated by the expression

$$\begin{cases} \cot\varphi = \frac{A_3}{A_2} \operatorname{tg}(\varphi + \operatorname{artg} \frac{A_2}{A_1} \operatorname{tg}\varphi), \\ \varphi = \omega l/c \end{cases} \quad (1.4-7)$$

which is based on the condition that the sum of the reactances to the right and the

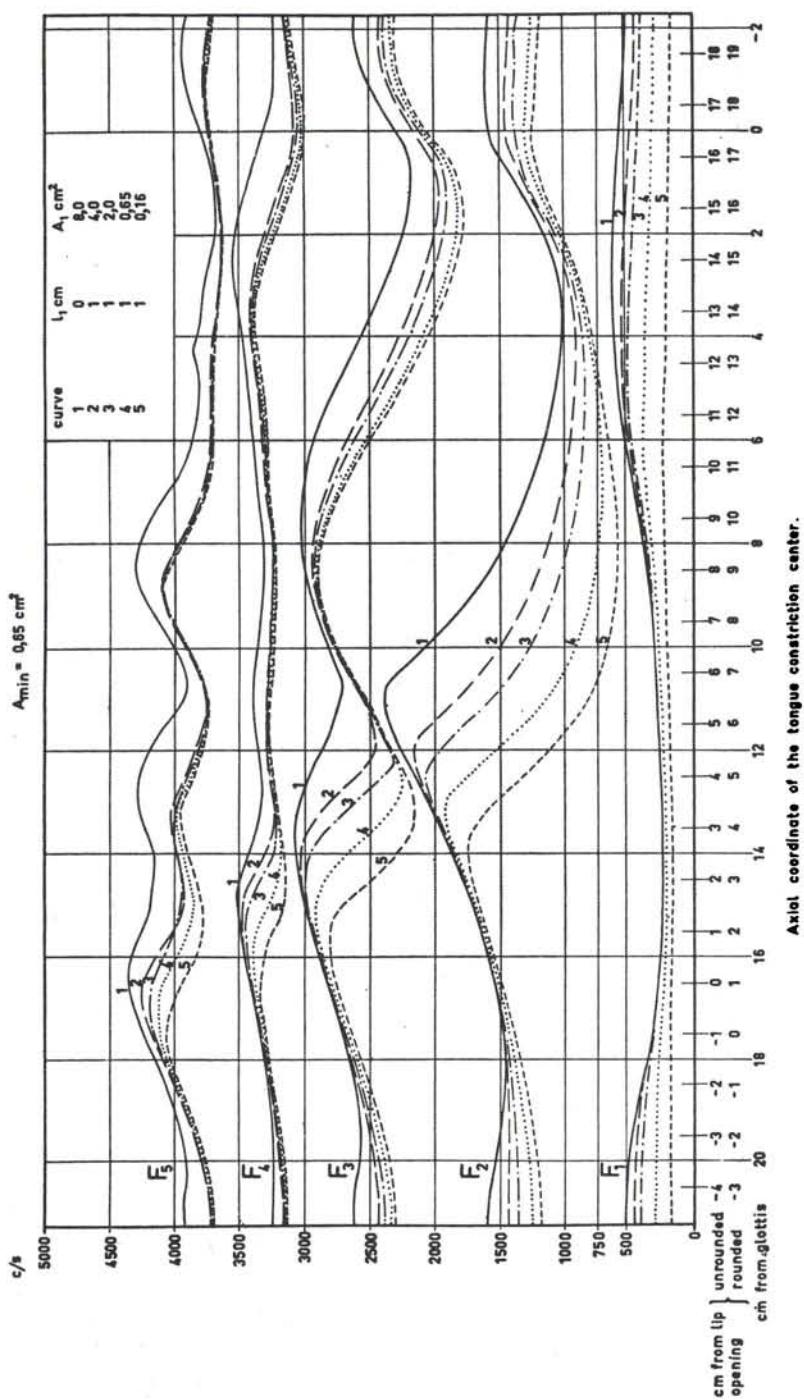


Fig. 1.4-11 a.
Axial coordinate of the tongue constriction center.

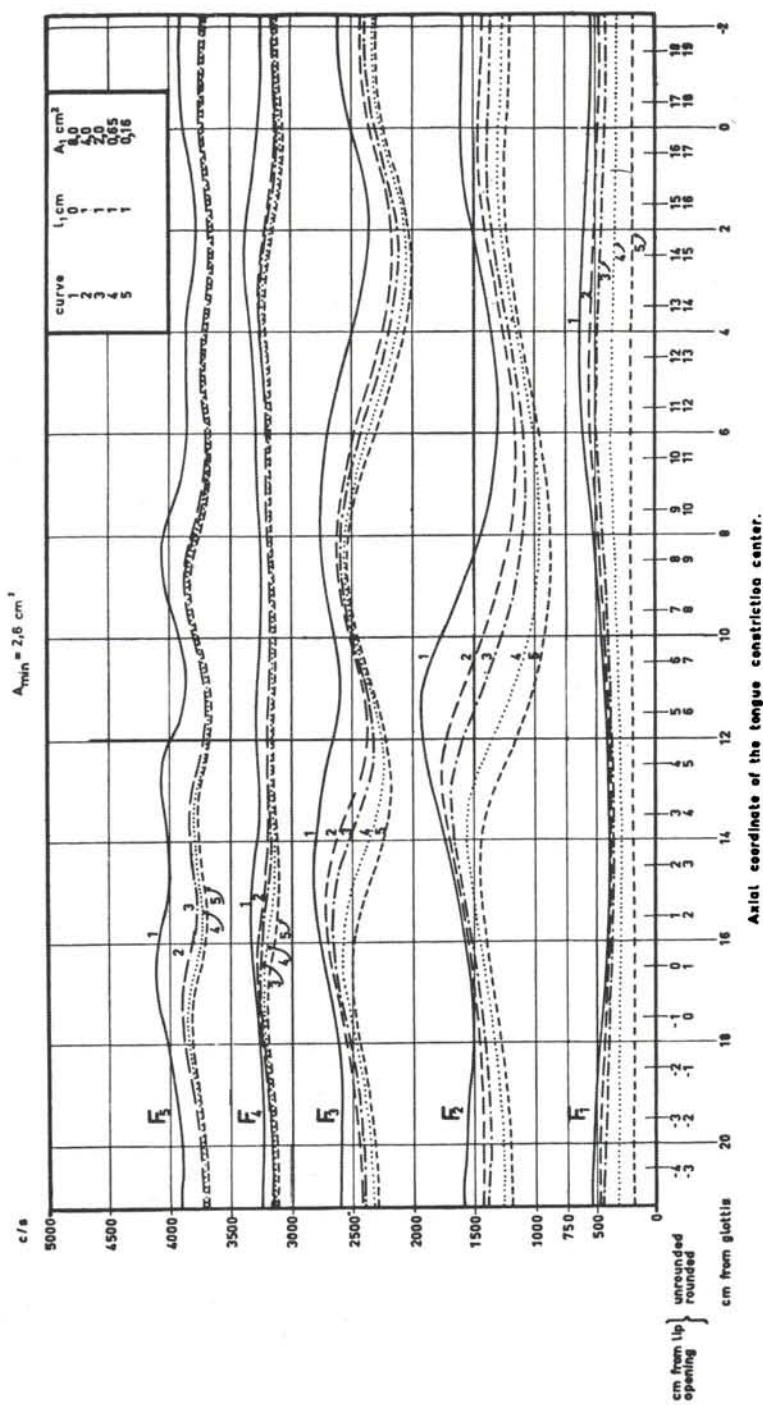


Fig. 1.4-11 b.

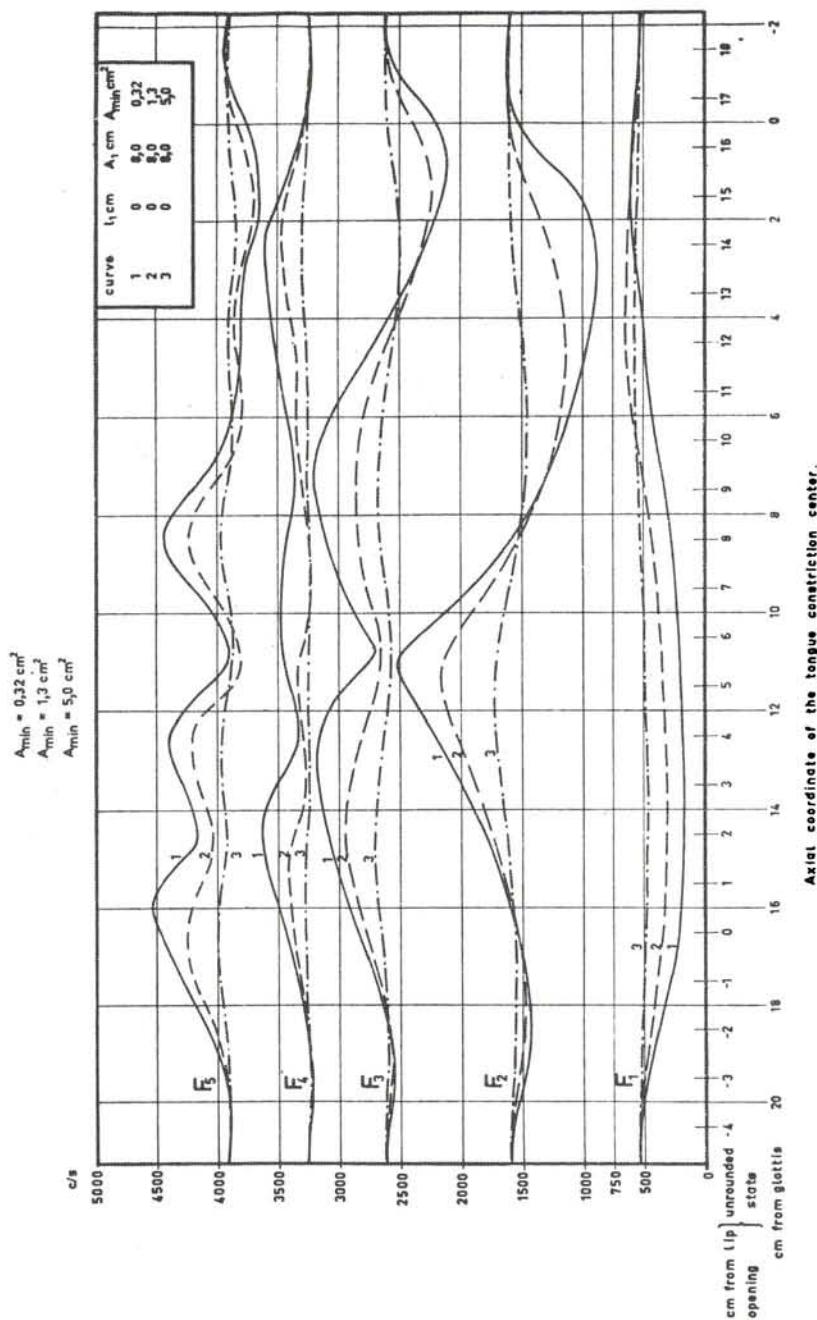


Fig. 1.4-11 c.

Fig. 1.4-11. Nomograms relating F_1 , F_2 , F_3 , F_4 , and F_5 of the horn-shaped three-parameter model, Fig. 1.4-10, to the location of the tongue constriction:

- Minimum tongue constriction area of 0.65 cm^2 . Five degrees of lip-rounding, curve 1 with no lip section and curves 2-4 according to the tabulation;
- Same as a) but $A_{min} = 2.6 \text{ cm}^2$;
- No lip section. Three degrees of opening at the tongue constriction.

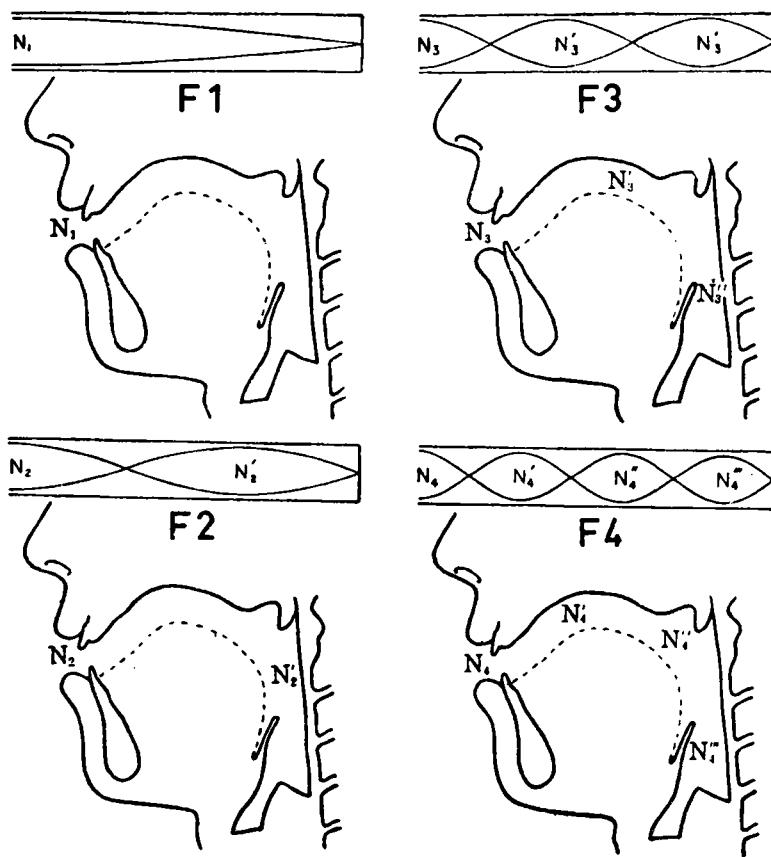


Fig. 1.4-12. Standing wave distributional pattern of volume velocity for each of the first four resonances within a single tube representation of the vocal tract according to Chiba and Kajiyama (1941).

left of the boundary between the back tube and the middle tube, shall equal zero. At the frequency where $\varphi = 90^\circ$ in each section, both the right and the left members of the equation above approach the value zero. This is independent of the particular area relations. The frequency of this resonance is $F_2 = c/4l = 3c/4l_{tot}$ where $l_{tot} = 3l$ is the total length of the model, and it apparently coincides with a neutral position of F_2 . The same effect is found at those constriction coordinates where a vocal tract model approximates a twin-tube of which the back or the front section has a length of one-third the total effective length of the system. Independent of the area relations, F_2 retains the single tube neutral position, as may be derived from the twin-tube formula

$$\cot\varphi_2 = \frac{A_2}{A_1} \operatorname{tg}\varphi_1, \quad (1.4-8)$$

where $\varphi_1 = 2\varphi_2 = \pi$ or $\varphi_2 = 2\varphi_1 = \pi$ which explains why the F_2 -curves of the separate A_{min} conditions meet at a constriction coordinate just behind the lips. The effective length of the front section includes the radiation end correction plus one-half of the tongue section length. The corresponding laryngeal coordinate of constant F_2 is not so apparent owing to the additional cavities there.

Some of the essentials of the relations between articulatory variations and formant frequency variations may be memorized by reference to the standing wave pattern within a single tube model of the vocal tract. Thus Chiba (1941) states *When a pipe is constricted its resonance frequency becomes low or high according as the constricted part is near the maximum point of the volume current (N) or of the excess pressure (P).* The volume velocity (current) distribution for each of the first four resonances of the neutral reference model is shown by Fig. 1.4-12. The spatial pressure variations are inverse to the volume velocity variations. Thus a volume velocity minimum is associated with a pressure maximum. Pressure and volume velocity variations in cavity models of specific vowels are discussed in Section 2.34 B.

At the frequency of the first resonance there is a volume velocity maximum at the lips and the volume velocity is high in the whole front part of the mouth. A constriction at the lips or in the mouth cavity thus lowers the first formant. The volume velocity maximum at the lips is a common feature of all resonances, and any formant frequency is thus lowered by a decrease of the lip-opening.

The second resonance has an additional volume velocity minimum, i.e., a pressure maximum, at two-thirds of the vocal tract length above the glottis. This is also the coordinate of maximum F_2 at varying tongue pass location of the two three-parameter models described above. The third resonance of the neutral tube has two spatial volume velocity minima, at four-fifths and two-fifths of the distance from the glottis to the lips which is in fair agreement with the evidence from the three-parameter nomograms of Fig. 1.4-9 and 1.4-11.

The spatial minimum of the frequency of the second resonance according to these nomograms conforms with what could be predicted from the standing wave pattern of the neutral tube, but only for moderate degrees of tongue passage constrictions. The coordinate of spatial F_2 -minimum retracts towards the glottis as the tongue passage area is decreased. This is due to the combined volume and aperture changes.

The rules for relating formant frequency variations to localized constrictions or expansions within a single tube may be derived from simple impedance considerations. A reduction of the tube cross-sectional area at the place of a volume velocity maximum is equal to the insertion of a lumped series inductance since the capacitance of the section may be neglected in view of the state of pressure minimum. If, on the other hand, a change in tube cross-sectional area is made near a volume velocity minimum, i.e., at the place of a pressure maximum, it is possible to disregard the distributed inductance at this place and take into consideration its capacitance only.

The effect of the increased lumped inductance is to lower the resonance frequency, and the effect of the decreased capacitance is to increase the resonance frequency.

This follows immediately from the basic circuit theorems that the reactance (in this instance comprising the sum of the reactance of the parts in front of and behind the point of inspection) shall be zero at the frequency of resonance and must be rising with frequency.

The effect of a local constriction or expansion of the vocal tract may thus be predicted on the basis of the knowledge of the standing wave pattern of the system before the articulatory change is introduced. It should be observed that the use of the standing wave pattern of the neutral tube as a reference is applicable for unrounded conditions only and that the locations of the maxima or minima are to some extent functions of the degree of constriction. The theory is less accurate for resonances above the third owing to the relatively large dimensions of the tongue section. If the initial conditions include lip-rounding, the reference standing wave pattern will have to be chosen accordingly. The effect of constricting the opening of the reference tube is thus to shift the spatial minimum of volume velocity of the second resonance from a location at two-thirds of the vocal tract length above the glottis to a location closer to the lip section.

This is reflected by the forward shift at increasing degree of lip-rounding of the spatial F_2 -maximum of the three-parameter vocal tract nomograms; see *Fig. 1.4-9* and *1.4-11*.

It is of some interest to estimate how much the formant frequency data of our horn model are influenced by the fixed cavities at the glottis end, i.e., the larynx tube and the sinus piriformis. It may also be anticipated that different ratios of the length of tongue constriction to the total length of the model might provide somewhat different results. In the following tabulation the frequencies of the first three formants are given for four separate model configurations as indicated.

The effect of the removal of the constant cavities at the glottis end appears as an increase in the maximum value of both F_1 and F_2 of which the F_1 -increase is the more apparent. The data of group C and D are representative of female and small children's vocal cavities respectively, because of the smaller lengths. Females have a much shorter larynx tube than men, and shorter throats. According to Chiba and Kajiyama (1941) the average male pharynx is 25 per cent longer than the average female pharynx, but the difference in mouth cavity length is only of the order of 10 per cent.

Stevens and House (1955, 1956), in the following referred to as *SH*, provide a more detailed nomographic material of curves relating the three main model parameters to the frequencies of the first three formants. Their model has a shorter overall length in its maximally delabialized state and a relatively longer tongue constriction section than our model of *Fig. 1.4-10* and *1.4-11*. The range of constriction coordinates of the *SH*-model is restricted to that between 4 and 13 cm from the glottis, which corresponds to the region of 3 to 12 cm of our all-cylindrical model and the range of 4.5 to 13.5 cm for our horn model. If the constriction area of our data is converted to an equivalent radius, it is found to be larger than the radius of the *SH*-model

TABLE 1.43-1

The effect of omitting laryngeal cavities and of shortening the overall length of the horn resonator, Fig. 1.4-10. Constriction area $A_{min} = 0.65 \text{ cm}^2$. $x = \text{distance of constriction center from the back end of the tube in cm}$.

A. Numerical data of curve 1 Fig. 1.4-11 a: No additional lip section, $A_{min} = 0.65 \text{ cm}^2$; total length of model 16.5 cm; both larynx tube and sinus piriformis incorporated.

B. Resonator configuration same as in A except for the removal of the sinus piriformis cavities.

C. Same as B except that also the larynx has been removed. Total length of the model is thus 14 cm. D. Same as C except for a reduction in the total length of the tube has been reduced to the value 11 cm. No scale reduction of the horn section.

Distance of constriction center from the front end in cm	A					B					C					D			
	cm	x	F_1	F_2	F_3	x	F_1	F_2	F_3	x	F_1	F_2	F_3	x	F_1	F_2	F_3		
-4	20.5	530	1590	2620	20.5	560	1695	2740	18	580	1740	2930	15	725	2185	3690			
-3	19.5	500	1520	2565	19.5	530	1630	2685	17	540	1680	2885	14	670	2100	3670			
-2	18.5	435	1465	2600	18.5	460	1570	2715	16	480	1625	2930	13	580	2085	3810			
-1	17.5	340	1460	2680	17.5	385	1580	2780	15	300	1635	3060	12	480	2155	4040			
0	16.5	300	1525	2795	16.5	330	1640	2900	14	340	1710	3240	11	395	2325	4280			
1	15.5	260	1620	2915	15.5	285	1750	3005	13	300	1840	3440	10	350	2570	4310			
2	14.5	245	1745	3015	14.5	265	1900	3080	12	280	2015	3550	9	340	2875	4140			
3	13.5	250	1910	3070	13.5	260	2080	3120	11	275	2220	3450	8	345	3075	4085			
4	12.5	260	2100	3005	12.5	275	2285	3080	10	285	2445	3265	7	380	3035	4260			
5	11.5	280	2310	2815	11.5	300	2470	2935	9	305	2610	3200	6	435	2860	4200			
6	10.5	295	2370	2705	10.5	315	2460	2855	8	330	2470	3290	5	545	2540	3830			
7	9.5	320	1970	2860	9.5	350	2050	2925	7	370	2050	3285	4	750	2110	3720			
8	8.5	355	1640	2975	8.5	400	1710	3020	6	430	1710	3255	3	1030	1815	3760			
9	7.5	410	1410	3015	7.5	475	1460	3055	5	525	1460	3215	2	1195	1860	3805			
10	6.5	480	1245	2990	6.5	600	1295	3030	4	710	1320	3200	1	1135	2215	3735			
11	5.5	555	1125	2895	5.5	715	1200	2945	3	880	1350	3090	0	1025	2560	3620			
12	4.5	600	1050	2710	4.5	770	1180	2780	2	880	1660	2895	-1	930	2650	3720			
13	3.5	625	1015	2520	3.5	760	1235	2575	1	825	2015	2780	-2	850	2540	3980			
14	2.5	635	1055	2345	2.5	720	1375	2410	0	790	2160	2940	-3	790	2380	3945			
15	1.5	625	1205	2215	1.5	680	1595	2305	-1	705	2085	3180	-4	745	2260	3800			
16	0.5	600	1440	2180	0.5	640	1765	2360	-2	660	1980	3240	-5	735	2200	3715			

providing the same F_1 of front vowel configurations. The A/l values of the mouth-opening are directly comparable with our A_1 values since our l_1 was held constant at 1 cm for the curves 2, 3, 4, and 5.

Any of the models discussed above can be utilized for a qualitative discussion of

the relations between articulation and formant patterns.¹ Referring to our horn model *Fig. 1.4-10* and the nomograms representing tongue constriction areas of 0.65 cm^2 , *Fig. 1.4-11a*, it is possible to produce acceptable variants of the vowels [ɑ] and [i] from constriction coordinates 4 and 12 cm respectively from the glottis and no lip-rounding. The vowel [o] could be produced from curve 3 at a coordinate of 6 cm and the vowel [u] from the curve 5 at a coordinate of 7 to 10 cm from the glottis and a lip-opening area of 0.16 cm^2 . Finally the vowel [ɪ] could be produced from curve 3 at a constriction coordinate of 11 cm; see further *Section 2.32*.

The articulation of dental consonants can be simulated in our model at constriction coordinates of 14 to 16 cm, where F_2 varies from 1800 c/s to 1600 c/s and F_3 from 3050 c/s to 2900 c/s. As mentioned by Stevens and House F_2 and F_3 will not be influenced by lip-rounding in the closed state. In case of incomplete closure, however, assuming A_{min} to be comparable with the lip-opening, there will be a decrease of F_3 .

Uvular, velar, and palatal consonants belong to the range of coordinates from 8 to 13 cm. The uvular point of articulation at 8 cm from the glottis coincides with the intermediate F_3 -maximum. In real speech this maximum is counteracted by the larger total axial pathway for the sound propagation from the glottis to the lips past the high back uvular constriction. These length variations should always be kept in mind when the model data are related to actual speech. A correction of about -10 per cent can accordingly be added to all frequencies referring to an uvular articulation. This correction successively vanishes as the place of articulation moves to the front or back end of the vocal tract.

One interesting observation concerning the F-pattern of palatal consonants is that at a mid-palatal tongue position of $x = 11 \text{ cm}$, where F_2 and F_3 meet in the unrounded condition at approximately 2500 c/s, the effect of superimposed lip-rounding of the same magnitude as that appropriate for the vowel [u], i.e., $A_1/l_1 = 0.16$, is to shift F_2 down to 750 c/s. From an inspection of X-ray pictures it is also evident that the varying amount of coarticulation of [k] and [g] with different vowels is more a matter of lip-opening area change than front cavity volume change. The decreasing amount of F_2F_3 -proximity at an increasing palatal tongue-opening, as in the transitional interval following a palatal sound, is apparent from *Fig. 1.6-11c*.

The F-pattern of labials can be extrapolated from that of any vowel configuration with a superimposed high degree of lip-rounding. As long as the lip-opening area is

¹ Comparing the performance of these vocal tract models with the range of F-pattern variations found in human speech, it may be said that F_1 of our model does not reach a sufficiently high maximum value and that F_3 of the American model does not easily reach sufficiently low values. These differences are partly due to the too large shunting effect of the sinus piriformis cavities of our model and the longer tongue constriction passage of the SH-model. It would apparently be more natural to let the sinus piriformis volume of our model decrease as the tongue constriction approaches the laryngeal region. The data of column A and B, *Table 1.43-1*, indicate the maximum effects to be expected. Contrary to our model the length of the horn section in the SH-model is dependent on the cross-sectional area at the point of maximum constriction. This dependency, however, is retained for $A_{min} > 0.5 \text{ cm}^2$ only. An additional local constriction in the center of the horn is introduced in the SH-model for radii $< 0.4 \text{ cm}$, i.e., for $A_{min} > 0.5 \text{ cm}^2$.

appreciably smaller than the tongue constriction area, there will be no appreciable additional lowering of formants even when the lips are completely closed.

It should further be noted that retroflex articulation cannot be simulated with the horn model unless an additional local constriction is made in the mouth part. A superimposed pharyngealization can, however, be simulated by an appropriate retracted location of the horn section which under these circumstances represents a secondary place of articulation.

PART II

CALCULATIONS BASED ON

X-RAY DATA

2.1 X-RAY PROCEDURE, SUBJECT, AND PHONETIC MATERIAL

In 1951 the subject was a 38 year old Russian, native of Moscow, and son of Moscovite intellectuals, actor of the Stanislavskii school, who had left his country ten years before. In order to sample the phonemes of Russian, the subject was instructed to pronounce sounds in the following manner: "Say [s] as in [sat]." In the case of the continuants and the vowels, he held the primary position of the articulators until the X-ray pictures were taken. In the case of the stops two pictures were taken, the first showing the occlusion and the second, the position of the articulators after the release.

The X-ray technique was described by Drs. A. S. MacMillan and G. Kelemen as follows:¹

"The outline of the tongue, the roof of the mouth, and the pharyngeal wall were made visible by giving the subject a mixture of barium and water to which mucilage of acacia was added. With the use of the latter, the barium mixture can be made much denser with the same amount of water. Adhesion to the mucous membranes is very good. This mixture was applied by spatula, and the subject was asked to smooth over the surface with his tongue to get an even distribution.

It is essential that the head should be held in a natural speaking position and that the same position should be maintained throughout the entire procedure. At the same time, it is essential that the head should not be fixed, by a clamp or by any other device, in a strained position. In upright position, the head was placed against a headrest at the occipital region. The lower edge of the cassette rested against the side of the shoulder (against the deltoid) at a distance of about 22 cm from the sagittal midline of the face. To prevent distortion caused by this considerable distance between the subject and the cassette, the X-ray tube was removed to 2 meters. In this way the distortion obtained was negligible. The central ray was directed through the crown of the first upper molar. The central ray being held constant, one was able to duplicate the work from day to day. As a further check to insure exact repetition of the position, a wooden throat stick was placed against the tip of the nose. A rotating tube was

¹ A. S. MacMillan and G. Kelemen, "Radiography of the Supraglottic Speech Organs," *A.M.A. Archives of Otolaryngology*, 55 (1952), pp. 681-2.

used, and the exposure time was $1/30$ of a second, with 75 kilovolts and 480 milliamperes. The films used were *Dupont* with detail screens. Development was carried out under the usual standard vacuum technique. The tracings were made on transparent paper over a highly illuminated viewing box. The fix points, such as the occlusal surface of the upper incisors, the roof of the mouth, the position of the lips, of the tongue, the dorsum and base of the tongue, and the position of the soft palate were sketched, together with the outlines of the posterior pharyngeal wall, the epiglottis, and the arytenoids. The accompanying tables offer a synopsis of the main technical points according to the different workers."

To supplement the information from the X-ray pictures, a plaster cast of the subject's buccal cavity was prepared. This was cut in various places and exact measurements of the sections were made.

The shape of the lips was obtained from photographs of the subject's face; see *Fig. 2.1-1*, which were taken simultaneously with the X-ray picture.

The sounds emitted were recorded on magnetic tape. Unfortunately the acoustical properties of the X-ray studio were very poor, so that the tape-recordings could only be used to compare the position of vowel formants. Other recordings² of the subject made in an anechoic chamber have been used for the comparison of spoken and calculated consonant data. Spectrograms of this speech material are shown in the *Appendix*.

With the exception of the /i/-phoneme, pictures were taken only of a single allophone of each phoneme. In the case of the /i/, both [i] and [ɪ] were studied. An X-ray photograph of the latter variant is shown in *Fig. 2.1-2*.

² The recordings were made at an average level of -10 VU in order to keep non-linear distortions at low values. Similar precautions were taken in the *Sonagraph* analysis and are necessary if clean, distinct spectrograms are to be made.

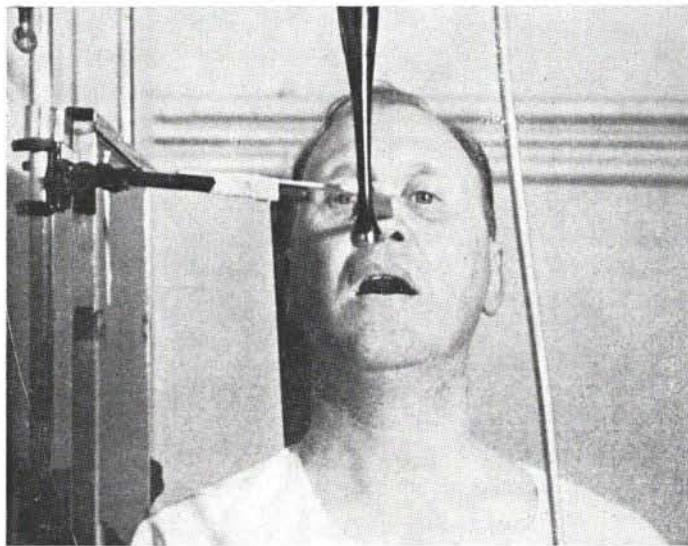


Fig. 2.1-1. Front view of the subject during X-ray photography.

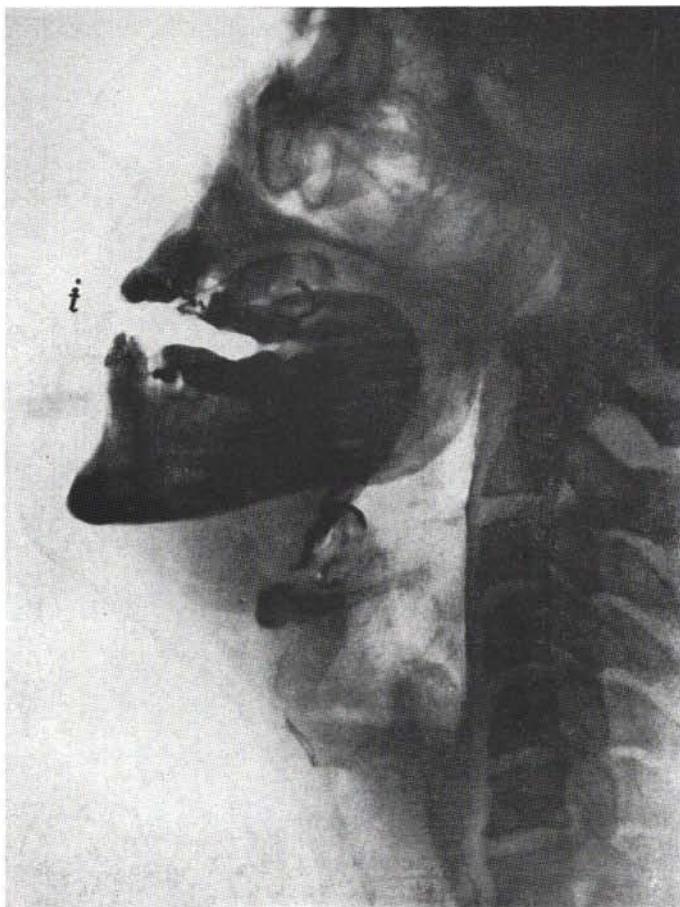


Fig. 2.1-2. X-ray picture during phonation of the vowel [i].

2.2 METHODS AND APPROXIMATIONS

The X-ray investigation was intended to provide the main physiological facts for studying the relations between articulation and speech wave, primarily on a distinctive feature level. Now, six years later, it is felt that the extent of the physiological investigation does not stand in proportion to the considerable amount of calculation work that has been devoted to this particular project. The scope of the investigation has grown, and it is evident that it would have been desirable to have more detailed articulatory data especially in the form of palatograms and data from more than one speaker. An extensive study of Russian speech has, however, recently been published by Koneczna and Zawadowski (1956). It contains X-ray and palatographic data of all Russian phonemes produced by several speakers. This is a valuable reference for articulatory data¹ supplementing the more analytically orientated presentation offered here.

Some of the vocal tract dimensions are impossible to measure under natural speaking conditions. Thus there is a considerable amount of guesswork involved in estimating the true shape and lateral dimensions of the pharynx cavity during back tongue articulation. The only investigation besides the present one that has been concerned with the mapping of vocal tract area functions is that of Chiba and Kajiyama (1941), and it was therefore natural to consult their data in doubtful cases. It has thus been postulated that a decrease of the observable distance between the back wall of the pharynx and the tongue as in the back vowels [ɑ] and [ɔ] is followed by a simultaneous decrease of the lateral dimensions thus decreasing the area at a rate faster than in direct proportion to the observed distance. This is naturally the case in the mouth cavity when the tongue approaches the palate.

Other sources of uncertainty are the true dimensions of the air passage in the region of the point of articulation of fricatives and affricates and the release interval of stops, i.e., their fricative and aspirative intervals. The extent of the air volume on both

¹ Other X-ray studies of general interest are those of Polland and Hála (1926); Russel (1928); Parmenter and Treviño (1932); Sovijärvi (1938a); Chiba and Kajiyama (1941); Forchhammer (1942, 1954). For further literature, see A. S. MacMillan and G. Kelemen (1952).

sides of the tongue and below the tip of the tongue is also questionable. Recent X-ray investigations in Sweden² indicate that these measurements show very great individual variations. The estimation of the lateral pathways of [l] has been pure guesswork.

Important data on the dimension of the nasal passages were obtained from direct measurements of a plastic mold from a corpse.³ However, these data should not be regarded as normative since the nasal passages are known to vary considerably in width from one individual to another.

It is hard to distinguish between complete closure and a small opening to the nasal passages even if the outlines of the uvula velum can be clearly seen, which was not always the case. It seems likely that a small degree of nasality is often present in the articulation of open vowels without causing a very noticeable change of quality.

It has been adopted as a convention to mark the central outline of the cavity walls of an X-ray tracing by solid lines and to mark lateral contours of interest by broken lines. The rearmost contour of the larynx tube has also been indicated by a broken line.

The effective center line for the acoustic wave propagation within the vocal tract may be rather dubious in case of sharp bends and complicated cross-sectional shapes. As a rule the center of gravity of successive cross-section slices has been used for defining the x-axis. The x-axis extends from the plane where radiation takes place at the lips to the bottom of the larynx cavity. The former point, or rather surface, has been adopted as the zero point. Exactly where radiation takes place along the x-coordinate is not well definable since the wave front of the sound emitted from the mouth is not confined to a plane surface and may be more or less spread laterally. Lip closure has the effect of screening off the side outlets at the cheeks and will thus cause some amount of prolongation even if the lips are not pursed forward. For delabialized sounds radiation has been assumed to take place not further than 0.5 cm in front of the teeth.

Possible errors due to less clearly defined dimensions are generally not crucial for the acoustic interpretation of the data. Length dimensions are more important than cross-dimensions because of the standing wave character of most of the formants. The existence of uncertainties in the mapping of vocal tract dimensions should not, however, discourage an investigator from making maximal use of his measurements. Errors due to local misjudgments are partially compensated for by the redundancies contained in a more precise evaluation of the remaining parts of the vocal tract.

Vocal sound sources cannot be studied very well by direct methods. The frequency distribution of source energy and the internal impedance of the source are best inferred from calculation on the basis of physiological mapping, flow data, and spectrographic data of the speech sound to be studied. The general procedure is to subtract from the latter the calculated vocal tract filter function in order to extract the source characteristics. As mentioned in *Chapter A.2*, it is highly probable that the voice source, as well as fricative sound sources, are associated with non-linear

² Performed by Dr G. Edholm, Karolinska Sjukhuset, Stockholm.

³ Personal communication of data from Dr Gunnar Bjuggren, Sabbatsbergs Sjukhus, Stockholm.

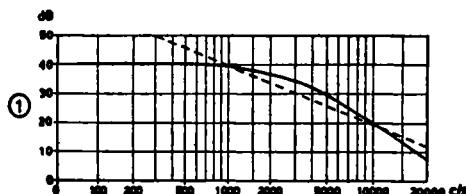


Fig. 2.2-1. Voice source spectrum envelopes with the addition of a $+6\text{dB/octave}$ rise. The solid line refers to the source utilized in an earlier stage of the calculations and included in the complex frequency diagram, Fig. 1.3-3. The broken line refers to the standard voice source of -12dB/octave slope.

impedances. The theoretical basis is not, however, very complete with regard to actual flow conditions within vocal tract constrictions.

In consonant calculations, several different combinations of source spectrum and source constriction impedances have been utilized in order to evaluate alternative basic assumptions. It should be observed that both the source spectrum slope and the source impedance will influence the overall spectrum slope of a fricative consonant. This sets certain limitations on the possibilities of inferring the true source slope from a comparison of measured and calculated data as discussed in more detail in *Section 2.64*.

Most calculations of the spectra of voiced sounds were initially performed on the basis of the voice source slope utilized by Stevens et al. (1953) for the M.I.T. line analog. A comparison with measured spectra showed, however, an apparent lack of level in the low frequency region. It was found that the -12 dB/octave slope gave a much better fit. All calculations were accordingly repeated with the -12 dB/octave source. The two source spectra are shown in *Fig. 2.2-1*. A $+6\text{ dB/octave}$ rise has been added to both curves which therefore represent the frequency characteristics of the voice source plus radiation.

The calculations on the Russian vowel material were carried out by means of the numerical methods described in *Section 1.22* and with the aid of *BESK*,⁴ a high-speed digital computing machine. A linear frequency scale has been utilized for the presentation of spectral data from these series. All other calculations were performed with the electrical line analog *LEA*⁵ and an automatic level recorder for tracing the spectrum curves. The logarithmic frequency scale of the level recorder was retained in the final presentation of the data. The ideal would have been to have some approximation to the mel scale, i.e., an essentially linear frequency scale up to 1000 c/s followed by a logarithmic representation. The logarithmic scale is convenient for displaying the characteristics of fricatives and stops. It causes, however, some over-

⁴ Binary electronic sequential calculator, The Swedish Board of Computing Machinery, Stockholm. In an early stage of the calculations the simpler *BARK* computer was utilized.

⁵ At the *Speech Transmission Laboratory, Division of Telegraphy and Telephony, The Royal Institute of Technology, Stockholm*.

emphasis of the first formant of vowels and voiced consonants. This should be kept in mind when evaluating the spectrum curves; see also *Section 1.22*.

*LEA** (see *Fig. 2.2-2*) contains 45 filter sections each representing a section of the vocal tract of 0.5 cm axial length and one out of 16 possible cross-sectional areas ranging from 0.16 cm² to 16 cm². The area scale is coded as follows:

TABLE 2.2-1

LEA. Standard area values

No.	Area in cm ²
1	16
2	13
3	10.5
4	8.0
5	6.4
6	5.2
7	4.0
8	3.2
9	2.6
10	2.0
11	1.6
12	1.3
13	1.0
14	0.65
15	0.32
16	0.16

Each of the 45 filter sections contains a parallel capacitance C , followed by a series inductance L with an associated loss resistor R_s . Every other C element is paralleled by a conductance loss element $G = 1/R_p$, representing the shunt losses. The elements R_s and R_p are realized in the forms of potentiometers. The elements L and C are varied stepwise by means of controls that are varied in a vertical direction on the panel perpendicular to the direction of transmission through the successive sections as shown in *Fig. 2.2-2*. The knobs of the successive area controls indicate the outline of the area function to be tested. This visible check has proved to be valuable when articulatory movements are simulated.

The L and the C of a unit section of length 0.5 cm are generally controlled simultaneously to keep their product constant according to the phase constant criterion

$$(LC)^{\frac{1}{2}} = 0.5/c, \quad (2.2-1)$$

where c is the velocity of sound. At 35 °C $c = 35300$ cm/s.

The cutoff frequency is

$$F_c = 1/2\pi(LC)^{\frac{1}{2}} = c/\pi = 11250 \text{ c/s.} \quad (2.2-2)$$

It is also possible to change the elements L and C independently for the purpose of making a filter section represent a shorter or a longer axial length than 0.5 cm. The ratio of L to C determines the analog area A by the relation

* The construction was carried out by B. Eliasson and S. Wadfors.

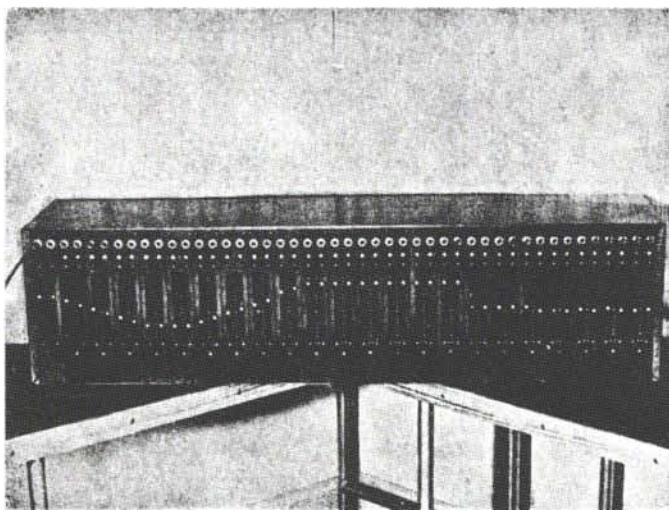


Fig. 2.2-2. The electrical line analog and speech synthesizer *LEA* utilized for most of the calculations in this work. *LEA* is a configurative analog in the sense that the outlines of the vocal tract area functions from the glottis to the lips may be visualized from the contact positions of the corresponding separate filter sections.

$$Z_e = \frac{\rho c}{A} k_e = (L/C)^{\frac{1}{2}}, \quad (2.2-3)$$

where Z_e is the characteristic impedance of the section in electrical ohms, and $k_e = 112.5$ is the conversion factor from acoustical ohms to electrical ohms. Now, if both L and C are doubled, the analog area A is not affected, but the analog length of the section is increased from 0.5 cm to 1 cm. This implies also that the cutoff frequency, F_c , of the section has been decreased to 5625 c/s.

A lengthening of sections by these means was utilized when simulating nasal consonants and nasalized vowels in which case 12 of the *LEA* standard sections were adopted for the nasal cavities. Separate radiation impedance sections were incorporated as a termination of both the oral and the nasal outlets of *LEA*. A mixer circuit for combining the two outputs was also incorporated.

The standard sections of *LEA* are plug-in units. The interconnection between any two successive sections can be opened by means of a three-pole plug. The two joining series branches are thus available separately or in connection as desired. When simulating the production of turbulent sounds the source is inserted in the series branch at the appropriate place. A carefully balanced transformer must be used for this source connection.

A standard value of 200 acoustical ohms, i.e., 22500 electrical ohms, was adopted as the source resistance of voiced sounds. According to the analysis in *Section 2.34*, the damping introduced by this resistance is moderate and fairly representative of actual speech. The losses in the circuit elements of *LEA* are generally small compared with those due to the source resistance and the radiation resistance. The bandwidth of formants from spectra calculated with *LEA* is in some instances smaller than what is generally found in real speech. This is especially true of the first formant. The radiation impedance of *LEA* has a damping effect which is very close to the required value as was shown in *Fig. 1.2-4*.

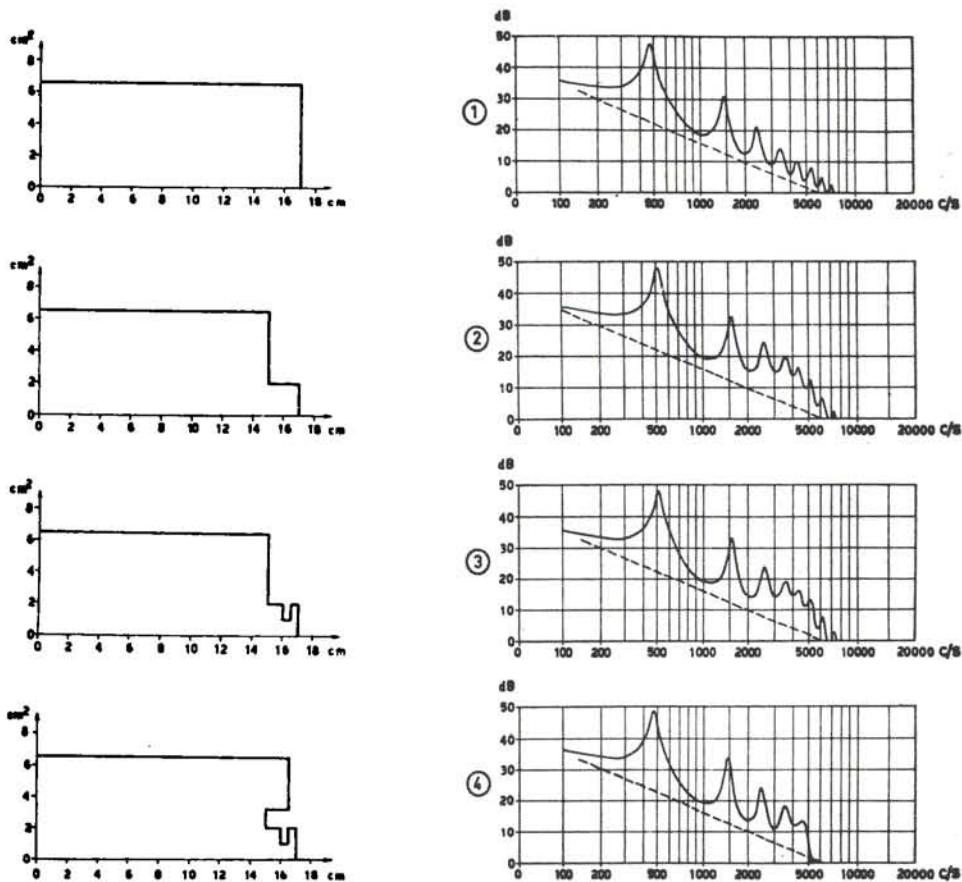


Fig. 2.2-3. Area functions and corresponding spectrum envelopes of vowels produced from a tube resonator with various modifications in the region of the larynx.

- (1) A single tube of 6.5 cm^2 cross-sectional area and physical length 17 cm (effective length 18.2 cm radiation end correction included);
- (2) The area reduced to 2 cm^2 at the first 2 cm of the tube representing *larynx tube*;
- (3) Same as (2) with the addition of *false vocal cords*;
- (4) Same as (3) except for the addition of *sinus piriformis* shunting the larynx tube.

A few other systematic factors influencing the calculations should be mentioned. The sinus piriformis constituting the bottom of the pharynx cavity on both sides of the larynx tube has been neglected since its shunting effect on the sound transmission from the glottis was not very easily taken into account in the numerical calculations. The effect of omitting this air space is demonstrated in Fig. 2.2-3 which also provides information on the effect of the *larynx tube* and the *sinus morgagni*. An open tube neutral vowel configuration was utilized for the remainder of the vocal tract. It can be seen that the single tube resonator provides a spectrum falling off at a rate of

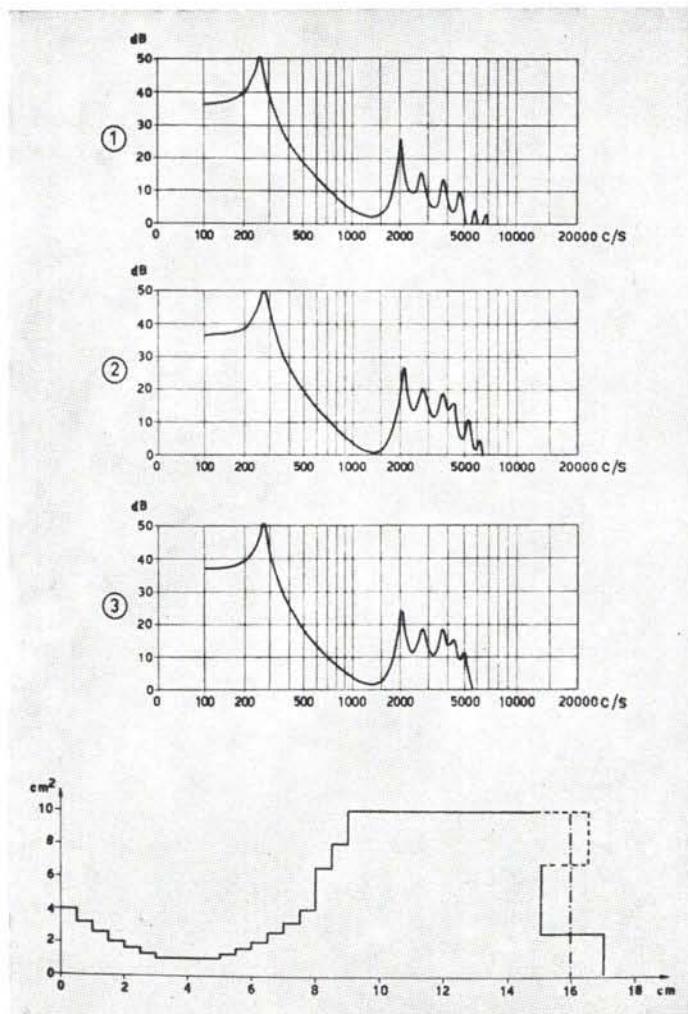


Fig. 2.2-4. Spectrum envelopes of vowels produced from an [i]-model based on a horn-shaped outline for the mouth cavity.

- (1) The pharynx plus larynx represented by a single homogeneous tube; total length of the model 16 cm;
- (2) Same as (1) with the addition of a tube, 2 cm long, representing the larynx and a shortening of the pharynx tube by 1 cm; total length 17 cm;
- (3) Same as (2) with the addition of a *sinus piriformis* shunt.

-6 dB/octave due to the particular source slope of -12 dB/octave . The formants occur at odd integers of 475 c/s corresponding to the quarter-wavelength resonance frequency, $c/4l_e$, of the tube.

The effect of reducing the area at the bottom of the tube so that a terminating tube of 2 cm length and 2 cm^2 area simulating the larynx tube is created, is to increase the density of poles in the frequency region of $3000\text{-}5000 \text{ c/s}$ by the introduction of one additional pole and a slight readjustment of the locations of adjacent poles. There is thus some sense in affiliating F_4 with the larynx tube resonance as has been proposed by Sovijärvi (1938a,b), and Chiba and Kajiyama (1941); see further *Section 2.33*. This frequency region is raised somewhat in level, but the spectrum slopes off faster above 5000 c/s . The introduction of the false vocal cords, as seen in curve 3, has a very small additional low-pass filtering effect. Finally, it can be seen from curve 4 that the addition of the sinus piriformis shunt sharpens appreciably the cutoff at 5000 c/s owing to the appearance of a zero, i.e., an anti-resonance, just above this frequency. It can also be seen that the slight rise in all formant frequencies below 3500 c/s caused by the formation of the larynx tube is counteracted by the addition of the sinus piriformis shunt.

The effect of the terminating cavity systems at the larynx is thus essentially to cause a pronounced low-pass filtering effect with a cutoff at approximately 5000 c/s . As pointed out by van den Berg (1955b), an effect of this general type⁷ can be thought of as part of the source characteristics.

A stepwise approximation to the [i]-model of *Fig. 1.4-5* which was composed of a horn for the mouth cavity and a cylinder for the pharynx cavity is shown in *Fig. 2.2-4*. It can be seen that the formant frequencies in curve 1 come close to those previously calculated and that the formation of a larynx tube, curve 2, and the addition of a sinus piriformis, curve 3, have the same low-pass filter effect as observed in the single tube case, *Fig. 2.2-3*. Curve 1 pertains to a back cavity enlarged to include the volumes of the laryngeal cavities. The total length of this simpler model was 16 cm .

The theoretical effect of omitting the sinus piriformis from the calculations, as will be done in all other calculations reported in this work, is thus to cause somewhat too high estimates of those formant frequencies that are primarily dependent on the pharynx cavity. The effect is probably greatest in back vowels, see *Table 1.4-1* of *Section 1.42*.⁸ In addition the suppression of the spectrum region above 5000 c/s is less pronounced.

The procedure for arriving at the area data needed for the calculations is demonstrated in *Fig. 2.2-5* which illustrates the vowel [i]; see the X-ray picture of *Fig. 2.1-2*.

⁷ The filtering, however, is primarily to be ascribed to the combination of the larynx tube and the sinus piriformis and not so much to the sinus Morgagni as postulated by van den Berg.

⁸ It is of some interest to note that the calculated vowel data fit the measured data better when the sinus piriformis is omitted than when it is taken into account. This can possibly be explained by a compensatory increase in F_1 caused by a small nasal or glottal shunt or a pharynx cavity shunt but other systematic errors might also be involved. This tendency, already mentioned in *Section 1.42*, has further support in results from recent calculations on vocal tract models of Swedish vowels.

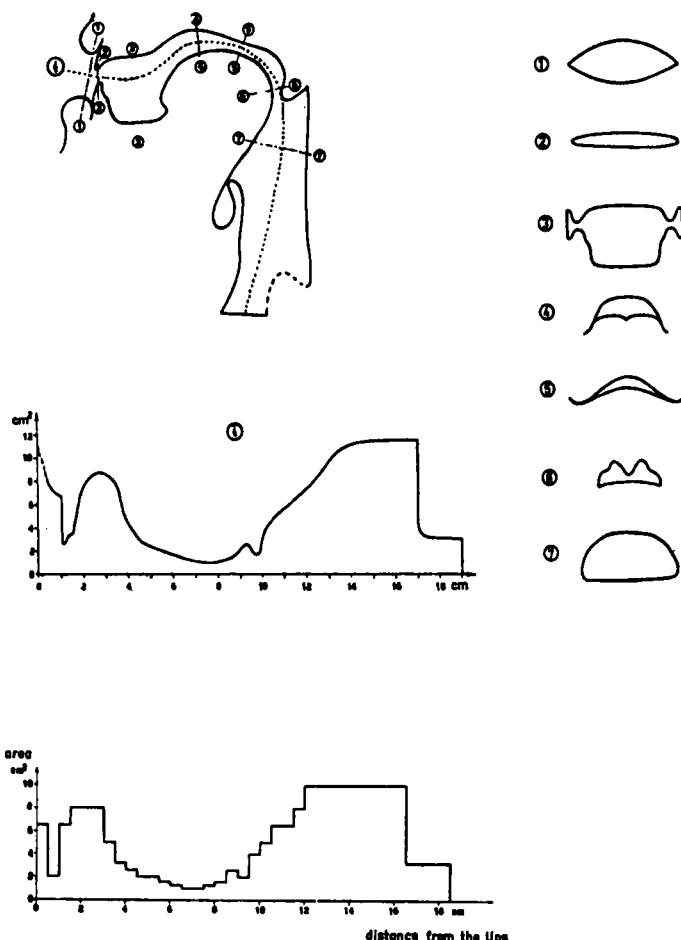


Fig. 2.2-5. The procedure for arriving at a stepwise approximation of the vocal tract area function from the X-ray data. Cross-section figures are made for a number of planes perpendicular to the central line from the glottis to the lips. A continuous outline of the area function is drawn and finally the area data within successive 0.5 cm sections are quantized according to the requirements of either the numerical calculations or those performed with *LEA*.

The first step is to perform closest possible estimates of the shape and area of the air pathways at a number of cross-sectional cuts perpendicular to the axial line. Some readjustments of the axial line may have to be done in the preliminary stage of this process. The area evaluations are performed on the basis of all available data, including dental molds of the mouth and frontal X-ray pictures of the pharynx. These measurements provide the first points of the area function curve. Intermediate values are estimated from the medial X-ray picture and from continuity considerations. Finally, the area function is approximated by a stepwise curve to fit the encoding requirements for the calculations with *LEA* or by means of numerical methods.

2.3 A STUDY OF VOWELS

2.31 Calculations of Formant Frequencies and Spectrum Envelopes

The X-ray tracings, *Fig. 2.3-1*, and the corresponding area functions, *Fig. 2.3-2*, of the six vowels [a], [o], [u], [i], [ɪ], [e] provide the physiological basis for the calculations. Before the production of these vowels is discussed, the results of the calculations will be reviewed briefly.

A comparison of the formant frequencies of the sounds actually sustained by the subject during the X-ray exposure and the formant frequencies numerically calculated from the area functions with the aid of the high-speed digital computer *BESK* is brought out in *Fig. 2.3-3*. The agreement is better than had been anticipated; note for

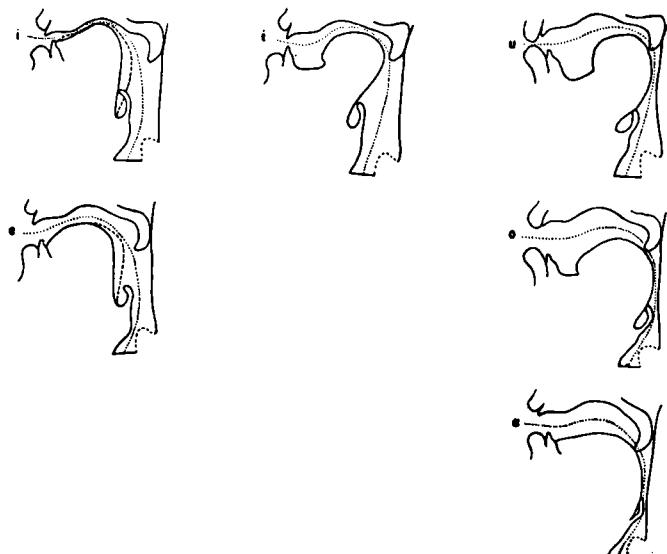


Fig. 2.3-1. Tracings of the vocal tract median sections from X-ray pictures of the Russian vowels [a], [o], [u], [i], [ɪ], [e].

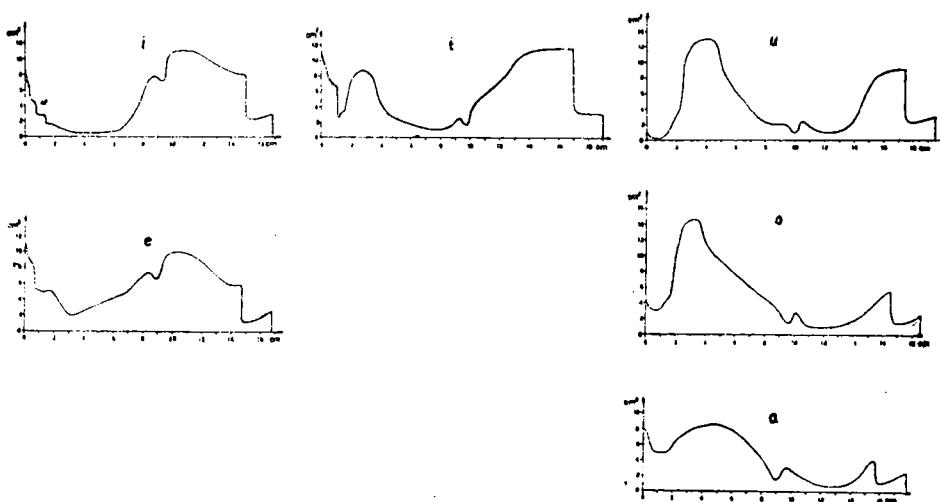


Fig. 2.3-2. Area functions derived from X-ray pictures for the six vowels shown in Fig. 2.3-1.

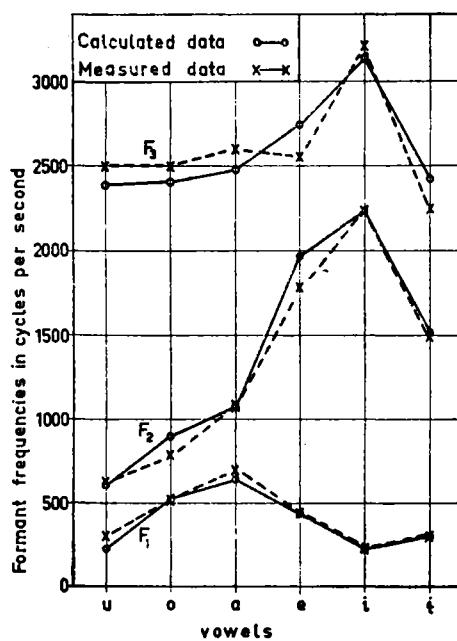


Fig. 2.3-3. Calculated and measured formant frequencies for the six Russian vowels. A high-speed digital computer (BESK) was utilized.

instance the almost perfect fit for F_1 and F_2 of [i] and [i]. The average deviation of calculated data from spoken data is of the order of 5 per cent in F_2 and F_3 and 10 per cent in F_1 . The maximum error in F_1 is that of [ɑ] and amounts to 84 c/s. The greatest absolute deviation between spoken and calculated data is for the F_2 and F_3 of [e] which are 200 c/s too high in the calculated data.

The calculations were repeated with *LEA* to check the accuracy of this device for further routine work. The standard deviations comparing the *BESK* and the *LEA* calculations were 3 per cent in F_1 , 1.3 per cent in F_2 , and 1.6 per cent in F_3 . Compared with the *BESK* calculations the *LEA* calculations gave on the average 0.3 per cent higher F_1 , 0.8 per cent lower F_2 , and 1.8 per cent lower F_3 . The spread is partially due to different types of approximations. When *LEA* is used, the area data are quantized in 16 steps; as shown by *Table 2.21-1*. The accuracy of the *LEA* calculations is quite acceptable for practical purposes.

The complete spectrum envelopes based on the *BESK* calculations are shown in *Fig. 2.3-4*. They have the same general appearance as the spectrum envelopes of the hypothetical sounds derived in *Section 1.32* from formant frequency data only. The level of F_3 and the higher formants is somewhat raised owing to the additional effect of the larynx tube resonance. The formant bandwidths are not the same. These will be

TABLE 2.31-1

Comparison of calculated and measured formant frequencies

BESK: Numerical calculations with a high-speed digital computer. The vocal tract was approximated by 20 uniform area sections, each specified by a cross-sectional area and a length measure.

LEA: Measurements on the configurative electrical analog *LEA*. Every successive section of 0.5 cm axial length is represented by a series inductance and shunt capacitance quantized in equivalent area steps of a factor 2⁴.

SUBJECT: Measured values of formant frequencies from spectrographic analysis of the sounds sustained by the subject during the X-ray sessions.

Vowel	F_1			F_2			F_3			F_4			F_6	
	<i>BESK</i>	<i>LEA</i>	<i>SUBJ</i>	<i>BESK</i>	<i>LEA</i>									
u	231	240	300	615	610	625	2375	2370	2500	3320	3400	4000	3950	
o	510	500	535	900	860	780	2400	2320	2500	3220	3500	3920	3800	
ɑ	616	630	700	1072	1072	1080	2470	2400	2600	3410	3550	3820	4000	
e	432	420	440	1959	1960	1800	2722	2750	2550	3500	3410	4400	4150	
i	222	230	240	2244	2220	2250	3140	2970	3200	3700	3570	4655	4400	
ɪ	296	285	300	1517	1480	1480	2413	2320	2230	3450	3200	4150	4200	

commented on in *Section 2.33*. Compared with the spectrum envelope of actual speech these calculated spectra show too high a level of the spectrum above F_3 , and they lack the extra *voice bar* formant below F_1 found in measured data of this subject.

The acoustic relations between the Russian vowels as revealed by a sequence

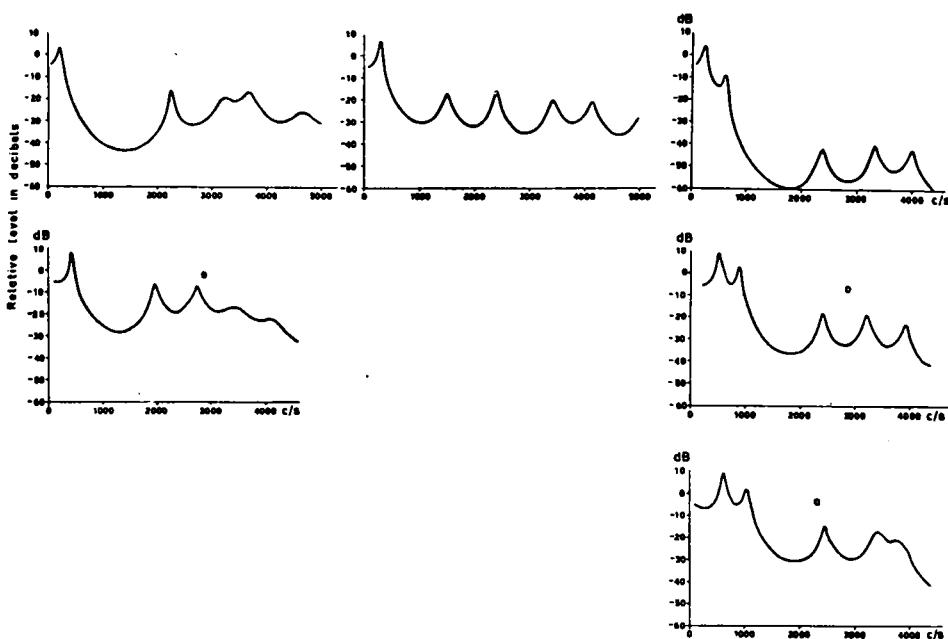


Fig. 2.3-4. Spectrum envelopes for the six Russian vowels calculated numerically (BESK). The spectrum level at frequencies above 3000 c/s is unnaturally high, indicating that the voice source spectrum envelope should slope off faster than -12 dB/octave at higher frequencies.

diagram of the type utilized in Fig. 2.3-3 have been discussed by L. G. Jones (1952). The third formant is of importance as a normalizing reference for F_1 and F_2 and as an effective contribution to the group of F_2 plus higher formants of acute vowels.

The first two formants are, however, generally sufficient for stating the acoustic relations. The *gravity* of the phonemes /u/, /o/, and /ɑ/ is correlated with a relatively lower F_2 compared with other vowels. The simplest way of describing the relation between /e/ and /i/ is in terms of F_1 which is higher in the former than in the latter. This *compactness* criterion could also be applied within the series /ɑ/, /o/, /u/. The *centralization* of the /u/, /o/, and /ɑ/ phonemes in positions between two *sharp* (palatalized) consonants resulting in the allophones [ü], [ø], and [æ]¹ is manifested by a higher F_2 . These are not treated here. Similarly the element of *palatalization* of the allophone [i], occurring next to a sharp consonant, and of [ɛ], followed by a sharp consonant, as compared with [i] and [ɛ] respectively, is attached to a higher F_2 and F_3 and also a lower F_1 .

2.32 Articulatory and Acoustic Vowel Diagrams

There are two basically different methods of making articulatory descriptions of vowels. One is the classical description in terms of the position of the highest point of the tongue in the back-front and low-high dimensions, and the other is

¹ See for instance Boyanus (1944).

a more or less approximate description of the dimensions of successive parts of the air chambers within the vocal tract from the glottis to the lips, in the most complete form by means of an area function. The latter system includes the data necessary for a mathematical prediction of the resonance frequencies of the vocal tract and the former method is merely a conventional phonetic reference frame. The highest point of the tongue is well correlated with the relevant acoustic data but does not uniquely define the resonator dimensions.

One approximate specification of the vocal tract dimensions is by means of the three-parameter model introduced by Stevens and House (1955, 1956) and adopted in a revised form here; *Section 1.43*. The place and cross-sectional area of the tongue constriction of this model together with the degree of lip-rounding uniquely define the resonator dimensions, but the first two of these parameters are not related in a simple fashion to the movements of the highest point of the tongue, as will be shown in *Section 2.33*. The double resonator interpretation of the vocal tract dimensional data has been widely used, and will therefore be discussed in some detail in *Section 2.32*. It has a more restricted applicability than the three-parameter models. This section is devoted to a comparison of the classical tongue height diagram and the F_1 -versus F_2 -diagram. The causal resonator-formant relations will be discussed in the next section.

From the X-ray tracings, *Fig. 2.3-1*, the tongue positions of a vowel may be measured as the highest points on the upper surface of the tongue in the median (sagittal) plane. These points have been determined for each vowel and then plotted within the frame of the X-ray tracing of the vowel [u], as can be seen in *Fig. 2.3-5*.

The resulting figure reflects the traditional relations of [u] and [o] occupying back positions, [u] being situated higher than [o]. Similarly [i] and [e] occupy front positions, [i] being higher than [e]. The vowel [ɪ] occupies a position between [u] and [i]; and [a] occupies a low position more fronted than [o] and also somewhat lower. The tongue height diagram is also shown in *Fig. 2.3-5*, upper right, after a rotation of coordinates 60 degrees clockwise and a linear scale expansion. This transformation results in a remarkable resemblance of the articulatory diagram to the F_1 versus F_2 plot drawn with an F_1 -scale of increasing values to the left, which is a figure in the *second quadrant* of the generalized coordinate system.

The resemblance is even closer than what could be expected in view of a similar study reported by Joos (1948). It should be noted that a tongue displacement leading to a vertical shift down of the reference point is correlated not only with an increase in F_1 but also with some decrease in F_2 . Similarly, the horizontal fronting of the reference point is correlated with an F_1 -increase in addition to the more apparent F_2 -increase. These secondary effects, the F_2 -decrease following tongue-lowering and the F_1 -increase following tongue-fronting, are somewhat exaggerated when a speaker bends his head back. Our subject's head was bent slightly backward which should be taken into account when comparing these data with those of others, for instance D. Jones (1934).

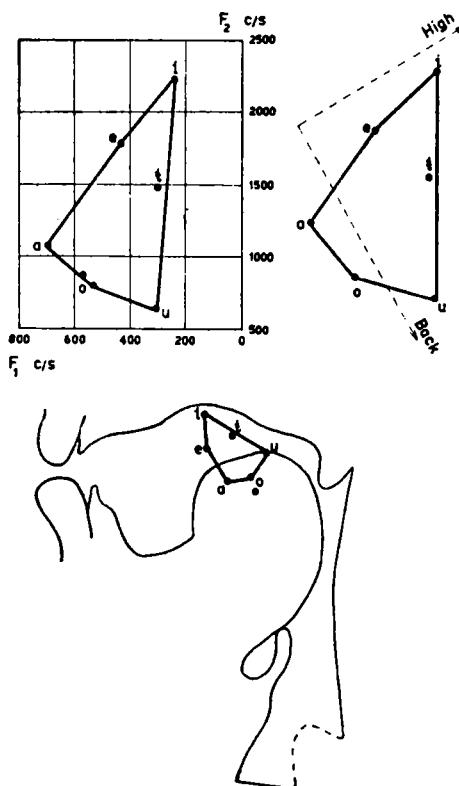


Fig. 2.3-5. Acoustic and articulatory specification of vowels in two-dimensional diagrams. The vowel figure of the F_1 -versus F_2 -diagram resembles the figure constructed from the highest points of the tongue of the separate vowels. The vocal tract median picture is that of the vowel [u].

The above-mentioned secondary effects of the tongue shift are reduced if the low-high dimension is defined not by a vertical line but by a line perpendicular to the roof of the back part of the mouth in the region of the front part of the velum. A tongue elevation or more precisely, a shift of the reference point perpendicular to this surface, provides an F_1 -decrease² but a fairly constant F_2 , and a shift of the reference point in a direction parallel to this surface causes an F_2 -increase at fairly constant F_1 .

The extent of this correlation is, of course, dependent on the degree of lip-rounding and other variations besides tongue movements that affect the configuration of the air chambers, such as the position of the larynx and the velum and the degree of jaw-opening. An additional tongue position point for the vowel [o] is included in Fig. 2.3-5. This sound was produced with a lower tongue position, a larger jaw-

* Conforming with earlier statements by Delattre (1951).

opening, and a greater lip-opening than the regular [o] included in the figure. The acoustic difference between the two variants is smaller in the F_1F_2 -diagram than in the articulatory diagram and of a different direction within the diagram, as can be seen.

2.33 The Relations Between Resonator Dimensions and Formant Frequencies

Three common misconceptions, which have been mentioned earlier in this work, are found in the current theories of speech production. One is concerned with the role of the pharynx. Classical articulatory schematics, such as those shown by Fletcher (1929) and D. Jones (1934), indicate a fixed pharynx cavity supporting a number of different profiles of the upper part of the tongue. This description is linked together with a second misconception, namely, that the *highest point of the tongue* should consistently coincide with a region of *minimum cross-sectional area* in the vocal tract separating a front cavity from a back cavity.

Available X-ray data in the phonetic literature do no support these hypotheses. The large pharynx cavity variations and the existence of a pharyngeal region of maximum narrowing in back vowels have been reported by several investigators, e.g., Russel (1928); Heffner (1949); Dunn (1950); Malmberg (1952).

The third misconception pertains to the inevitable oversimplification of the acoustic function of a compound resonator system. The popular F_1 -back-cavity, F_2 -front-cavity affiliation is seemingly supported by the correlation between the increasingly smaller mouth cavity and wider pharynx cavity within the series [a], [æ], [e], [i], and the associated decrease of F_1 and increase of F_2 . In spite of the fundamental contributions to the understanding of the origin of vowel formants made by Dunn (1950), this incomplete and partially erroneous theory has prevailed in the phonetic literature.³

³ The historical development of the acoustic theory of vowel resonances is an interesting field of study, but a lengthy review would be of doubtful value owing to the incompleteness with regard to instrumental resources and acoustic theory in earlier works. Our results support most of Dunn's (1950) statements concerning the origin of the first three formants, but the present investigation has a wider scope and is based on a more complete model. The standing wave character of some of the vocal resonances, e.g., F_3 , does not seem to have been understood by some phoneticians that have quoted Dunn. The recent book by Hálá (1956) contains a wealth of references to the older European literature supplementing the historical notes made by Dunn (1950) and Chiba and Kajiyama (1941). Hálá's physiological-acoustic interpretations of resonances higher than the second and of intermediate resonances are incorrect and reflect the tendency of vowel theorists of the older school to anticipate a separate cavity, or a sound-generating object for each formant. It is, however, interesting to see that he considers the classical F_1 -back-cavity, F_2 -front-cavity affiliations reversed for [u] and [o], which, comparing back and front vowels, in a relational sense, is supported by our calculations above. The extreme example of an intuitive construction of formant-cavity affiliations, in part compiled from earlier suggestions in the literature, is the physiological classification system of Sovijärvi (1938a,b) involving a large number of formants and corresponding subcavities or regions within the vocal tract (seven variable and eleven fixed). Parts of this system are a product of incomplete acoustic theory but some statements, e.g., concerning the larynx tube resonance, are supported by the present work. Other statements, e.g., concerning nasal resonances and the trachea resonance, are in part correct or deserve a closer investigation. A strikingly simple theory for the interrelation of F_1 , F_2 , and F_3 has been proposed by Ganeshsundaram (1957). His *cascade modulation* theory suggests that $F_3 - F_2$ is constantly equal to $2F_1$, so that F_3 would be an upper and F_2 a lower side-band located plus and minus F_1 relative to a mouth cavity resonance F_2' , the latter suppressed in amplitude. Quite apart from the unrealistic physical theory, which is upset by the basic principle of

It is the purpose of this section to supplement the more general theories of vocal tract models developed in *Chapters A.3 and 1.4* with a study of the main physiological and acoustic facts concerning the production of the six Russian vowels studied in this work. Among questions of particular interest are the validity of the double Helmholtz resonator theory and the applicability of the three-parameter nomograms of *Section 1.43B*. Quantitative expressions for the relative role of any particular part of the vocal tract as a determinant of each of the formants will be dealt with in some detail.

Table 2.33-1B summarizes the articulatory data extracted from the vocal tract area functions of *Fig. 2.3-2*. The good agreement between calculated and actually measured formant frequencies from tape-recordings of the subject shown in *Fig. 2.3-3* guarantees a fair degree of reliability of the essentials of the area functions, but does not exclude misjudgments concerning details, as stated in *Chapter 2.2*.

Among the most apparent features of the area functions is the inverse configuration of [ɑ] as compared with [i], verifying earlier statements in *Chapter 1.6*. Within the series [ɑ], [o], [u], [i], [ɪ], [e], the place of articulation, defined from the effective center of the tongue constriction, moves up and forward in the vocal tract from a place 4 cm above the glottis to a place 4 cm behind the radiating surface at the lips. The resonator system has a maximum length of 19.5 cm in [u], on account of the longer pathway of the air stream past the high back tongue hump. The cross-sectional area A_{\min} at the true place of articulation does not show great changes from one sound to the next of the tabulated series except in the step from [i] to [e]. On the average A_{\min} is of the order of 1 cm². When speaking in terms of the highest point of the tongue, there is an increasing degree of mouth cavity *opening* in the vowels [u], [o], [ɑ]. In terms of the major internal constriction, on the other hand, there is an approximately constant degree of articulatory opening within this series and a slight tendency of progressive closure found for this particular speaker.

There is a clear positive correlation between tongue height and the position of the velum indicating an incomplete closure to the nasal cavities in very *open* vowels, definite in the case of [ɑ] but doubtful for [e] and [o]. The effects of various degrees of nasalization on the characteristics of vowel spectra are discussed in *Section 2.41*.

A systematization of the resonator dimensions can be made according to several different methods, as mentioned in *Section 2.32*. One variant of the dimensional

linearity of the vocal tract network, it should be acknowledged that his basic observations hold approximately for a restricted range of vowels. As seen from the twin-tube and the three-parameter models of *Chapter 1.4*, there is an apparent tendency for F_1 to vary inversely to F_2 and for F_3 to vary opposite to F_2 . This is especially the case for a twin-tube configuration in which the two parts are of equal length. Here the rule proposed by Ganeshsundaram holds exactly. The partial applicability to speech can be ascribed to the tendency of the tongue to alternate between a palatal and a pharyngeal point of maximum narrowing and the tendency of the constriction to occupy one-half of the vocal tract length. However, there are marked exceptions from the $F_3 - F_2 = 2F_1$ rule (or the same rule with F_1 and F_2 reversed as proposed for back vowels) such as the large separation of F_3 from F_2 at a very low F_1 of [i] and conversely the high F_1 combined with a location of F_3 close to F_2 , as in [æ] and in retroflex vowels.

TABLE 2.33-1

A. Cross-sectional area A versus distance x from the front end of the vocal tract. The data on the six vowels are quantized in area steps of $\sqrt[3]{2}$ according to the coding requirements of the electrical line analog LEA.

	[a]	[o]	[u]	[i]	[ɪ]	[e]
x cm	A cm^2	A cm^2	A cm^2	A cm^2	A cm^2	A cm^2
0	5	3.2	0.65	6.5	4	8
0.5	5	3.2	0.65	6.5	4	8
1	5	3.2	0.32	2	3.2	5
1.5	5	3.2	0.32	6.5	1.6	5
2	6.5	6.5	2	8	1.3	4
2.5	8	13	5	8	1	2.6
3	8	13	10.5	8	0.65	2
3.5	8	16	13	5	0.65	2.6
4	8	13	13	3.2	0.65	2.6
4.5	8	10.5	13	2.6	0.65	3.2
5	8	10.5	13	2	0.65	4
5.5	8	8	10.5	2	0.65	4
6	8	8	8	1.6	0.65	4
6.5	6.5	6.5	6.5	1.3	1.3	5
7	5	6.5	5	1	2.6	5
7.5	4	5	3.2	1	4	6.5
8	3.2	5	2.6	1.3	6.5	8
8.5	1.6	4	2	1.6	8	6.5
9	2.6	3.2	2	2.6	8	8
9.5	2.6	2	2	2	10.5	10.5
10	2	1.6	1.6	4	10.5	10.5
10.5	1.6	2.6	1.3	5	10.5	10.5
11	1.3	1.3	2	6.5	10.5	10.5
11.5	1	0.65	1.6	6.5	10.5	10.5
12	0.65	0.65	1	8	10.5	8
12.5	0.65	1	1	10.5	10.5	8
13	0.65	1	1	10.5	10.5	6.5
13.5	1	1.3	1.3	10.5	10.5	6.5
14	1.6	1.6	1.6	10.5	8	6.5
14.5	2.6	2	3.2	10.5	8	6.5
15	4	3.2	5	13	2	1.3
15.5	1	4	8	13	2	1.6
16	1.3	5	8	10.5	2.6	2.0
16.5	1.6	5	10.5	10.5	3.2	2.6
17	2.6	1.3	10.5	—	—	—
17.5	—	1.3	10.5	3.2	—	—
18	—	1.6	2	3.2	—	—
18.5	—	2.6	2	3.2	—	—
19	—	—	2.6	3.2	—	—
19.5	—	—	2.6	—	—	—

B. Resonator data extracted from the quantized area functions of Table A. (The vowels are ordered according to the location of the tongue constriction.)

Vowel	Total length cm	Location of tongue constriction center		Minimum tongue constriction area $A_{2\min}$	Front cavity volume V_1	Back cavity volume V_2	Length over area of resonator neck		Volume ratio $\frac{V_1}{V_2}$	Resonator-formant coefficient $R = \frac{V_1 l_1 A_2}{V_2 l_2 A_1}$
		From the front end cm	From the back end cm				cm^2	cm^3	cm^3	
[ɑ]	17	13	4	0.7	59	8.4	0.75	5.2	7.0	1.0
[o]	18.5	12	6.5	0.8	70	13	0.9	4.2	5.4	1.2
[u]	19.5	11	8.5	1.0	54	31	5.3	4.2	1.7	2.2
[i]	18.5	7.5	11	1.1	27	72	0.7	3.5	0.38	0.076
[ɪ]	16.5	4	12.5	0.5	6	73	0.5	7.5	0.08	0.053
[e]	16.5	4	12.5	2.2	11	76	0.6	1.4	0.14	0.06

mapping of the vocal cavities is to refer to the mouth cavity dimensions and pharynx cavity dimensions by means of a fixed physiological division. This implies that the border between the two cavities is defined without regard to the articulation, e.g., at a fixed plane in the region of the uvula. The back vowels [u], [o], [ɑ] are then characterized by a larger mouth volume than the front vowels [i], [ɪ], [e].

If the true point of articulation is utilized as the dividing point for the front and back cavities it may be seen that the front to back cavity volume ratio V_1/V_2 is useful both for separating [u], [o], [ɑ] from the rest of the vowels and for relating [o] to [u] and [ɑ]. In the front vowels the volume ratio separates [e] from [ɪ], but we lack a hard variant of /e/ to compare with [i], the hard variant of /ɪ/. The data for [i] show a larger volume ratio than for [e], the latter being a *soft*, i.e., palatalized variant. If, however, the mouth volume to pharynx volume ratio is utilized as the criterion, it is found that both [i] and [ɪ] were produced with smaller volume ratios than the [e].

One important question is whether there is any higher amount of precision in the estimation of the center coordinate of the place of articulation. Referring back to the area functions of Fig. 2.3-2, it is apparent that such a center coordinate can easily be specified. The correct manner of evaluation of this place and the associated $A_{2\min}$ is to consider the entire tongue hump and decide how a horn connecting section should be placed in order to fit best the detailed area function. In the case of asymmetric constrictions, this coordinate does not coincide exactly with the absolute minimum of the area function. Also it cannot be expected that $A_{2\min}$ will coincide exactly with the minimum area value of the measured data.

In open vowels of the [æ] and [œ] type on the other hand, the place of articulation may be an illusive concept and thus worthless as the basis for the specification of

volume ratios. In these instances there is no physical foundation for the utilization of the double Helmholtz resonator as a mathematical model.

A specification in terms of the place of constriction and its cross-sectional area in accordance with the three-parameter method retains, however, a sufficient degree of accuracy for the determination of the F-pattern since the error involved in the estimation of the place of the constriction center decreases rapidly as the vocal tract configuration approaches that of a fairly homogeneous tube in which case the formant pattern is determined solely by the total length and the degree of lip-rounding.

It should be observed that the double Helmholtz resonator interpretation of the vocal tract data implies that the whole of the internal constriction constitutes the neck of the back resonator. The division between the two resonators will then not fall in the center of the constriction. The relevant data on the dimensions of a neck are contained in the l/A ratios of the tabulation. The inverse of an l/A ratio, i.e., A/l , is the conductivity index of the neck. The l/A ratio multiplied by ρ , the density of the air, constitutes the inductance of the neck. It contains some contribution from the two adjacent cavities and may be computed from a summation over successive sections of length 0.5 cm

$$L = \frac{\rho}{2} \left(\frac{1}{A_a} + \frac{1}{A_b} + \dots + \frac{1}{A_q} \right), \quad (2.3-1)$$

which is a practical approximation of the integration

$$L = \rho \int_{x_a}^{x_b} \frac{dx}{A(x)}. \quad (2.3-2)$$

If the differences in area, between the constriction and the cavities, are not very great, or if the area variations are gradual, there will result an uncertainty as to where to set the limits of the integration. The rule followed here has been to carry the integration almost into the center of the adjoining cavities. This choice is based on the low frequency approximation of the input impedance to a cylindrical cavity closed at the back end. The mass element contribution is equivalent to that of one-third of the tube length. It should be noted that the corresponding measure of the degree of lip-rounding l/A , must be defined differently for the double resonator theory and the three-parameter model. When dealing with the latter model the inductance contribution from the front part of the mouth cavity is excluded and only the lip section, except for end corrections, is included.

For the sake of consistency, the area functions of [i] and [e] will next be forced into the frame of the double Helmholtz resonator theory. This is partly a violation of the physical facts. The elements of arbitrariness involved are large since in these sounds the whole tongue constriction plus mouth cavity act as an inductance which tunes the main cavity to the frequency F_1 . The mouth cavity and the tongue constriction

act as a single tube which gives rise to a standing wave resonance that is more or less coupled to standing wave resonances of the back cavity.

The procedure for converting resonator dimensions into resonance frequencies with the aid of the double resonator formula is first to compute the resonance frequencies of the front and back resonators separately. The true F_1 and F_2 are then found with the aid of Eq. A.32-2, or with the nomograms of Fig. A.3-2. As previously stated, the two *uncoupled* resonance frequencies must fall between F_1 and F_2 , and their geometrical mean is the same as the geometrical mean of F_1 and F_2 .

The results of the calculations are summarized in the following tabulation. The double resonator theory has evidently failed for the prediction of F_2 of [e] and [i]. The calculated values lie closer to F_3 . The first formant of these two vowels and the first two formants of the remaining vowels [a], [o], [u], [i] can apparently be estimated with a reasonable degree of accuracy. The tendency is towards too high frequency values. This deviation is, of course, dependent on the particular convention adopted for integrating the inductance elements.

TABLE 2.33-2

Results from application of double Helmholtz resonator theory compared with the data obtained with an electric analog (LEA).

Formant frequencies in cycles per second

Vowel	Electrical line analog			Double resonator	
	F_1	F_2	F_3	F_1	F_2
a	630	1070	2400	705	1015
o	500	860	2320	535	905
u	240	610	2370	235	640
i	285	1480	2320	322	1470
i	230	2220	2970	232	3350
e	420	1960	2750	460	3150

As one possible means of expressing resonator-formant relations within the frame of double resonator theory the parameter $R = V_1 l_1 A_2 / V_2 l_2 A_1$ has been included in the tabulation. According to the investigation in Section A.32 an R -value smaller than unity signifies that the back resonator is more closely connected with F_1 , than with F_2 and that the front resonator is more closely connected with F_2 than with F_1 . The reverse relation holds for R -values greater than 1. It should be remembered that resonator dependency here implies the sum of the influence of the neck and of the volume of a resonator. It was shown that at R -values close to 1, the neck of the front resonator influences F_1 slightly more than F_2 while the front cavity is more closely connected with F_2 . Similarly, under these conditions the neck of the back resonator, i.e., the connecting section between the two resonators, influences F_2 more than F_1 while the back cavity volume will influence F_1 slightly more than F_2 .

The following tabulation summarizes the results from calculations of these relational factors on the basis of the double Helmholtz resonator theory and provides for comparison the true measured values obtained from the complete electrical line analog representation. These relational coefficients are defined as the percentage increase in a specific formant frequency due to a one-per cent decrease of a specific resonator element. The coefficient $\frac{\Delta F_1}{F_1} \frac{C_1}{\Delta C_1}$, for example, thus indicates the degree of dependency of the first formant on the front cavity volume, and $\frac{\Delta F_1}{F_1} \frac{L_1}{\Delta L_1}$ the dependency of the same formant on the front resonator neck. According to the double resonator theory, these coefficients range between 0 and 0.5 representing no and maximum dependency. Other rules for the ideal conditions are that the sum of the coefficients for any element with regard to the two formants is a constant equal to 0.5 and that the coefficient relating, for instance, the front cavity to F_1 is the same as the one relating the back cavity to F_2 , and that the coefficient relating F_2 to the front cavity is the same as the coefficient relating F_1 to the back cavity. The coefficients relating the resonator necks to the two formants are subject to the same symmetrical relations.

The practical means for introducing differential changes in the electrical line model was to shorten the cavities by the elimination of a unit length section at the place of maximum cross-sectional area. The inductance variations were introduced by a narrowing at a section located in the center of a resonator neck, i.e., at the lips or at the tongue constriction.

The tabulation shows a reasonable agreement between calculated and measured values of [u] and [i] in support of the double resonator treatment of these sounds.

TABLE 2.33-3

*Formant-resonator element coefficients from double resonator theory (C) and from line analog measurements (M). All coefficients are negative except those marked *.*

For [o] and [d], however, there are some apparent differences in results, for instance, the larger measured effect of the shortening of the front cavity on F_2 . This is due to the considerable wave propagation. The maximum value of a formant dependency coefficient is increased to 1 when complete wave propagation is considered. The front cavity of [a] acts more like a single tube than a Helmholtz resonator, and the shortening of this part is thus equivalent to a removal of both a capacitance and an inductance.

Because of the irregular shapes of the vocal cavities it is not very meaningful to normalize the dependency factors of standing wave resonances with regard to cavity volumes. Length measures would be more significant except for the arbitrariness easily involved in a particular length division of a smoothly varying cavity structure. It has thus been considered desirable to state the percentage shift in formant frequencies without any normalization with regard to resonator dimensions. The following tabulation summarizes the effects of removing a section of unit length from one of the two main cavities:

TABLE 2.33-4

The percentage increase in formant frequencies due to the removal of a section of 0.5 cm length at various coordinates representing places of maximum and minimum cross-sectional area within the vocal tract as revealed by experiments with LEA:

Vowel	Front cavity					Back cavity				
	$\frac{\Delta F_1}{F_1}$	$\frac{\Delta F_2}{F_2}$	$\frac{\Delta F_3}{F_3}$	$\frac{\Delta F_4}{F_4}$	$\frac{\Delta F_5}{F_5}$	$\frac{\Delta F_1}{F_1}$	$\frac{\Delta F_2}{F_2}$	$\frac{\Delta F_3}{F_3}$	$\frac{\Delta F_4}{F_4}$	$\frac{\Delta F_5}{F_5}$
[a]	2.2	3.4	2.4	4.2	1.2	5.0	3.1	0.8	0.9	6.5
[o]	1.6	4.9	4.1	0.7	2.2	5.0	2.5	0.4	2.8	1.7
[u]	2.1	3.9	1.7	0.6	4.2	4.2	3.8	0.7	3.0	0.8
[i]	0.7	9.0	0.9	13.7	1.4	3.6	0.5	5.1	1.8	1.8
[ɪ]	1.3	0.2	6.1	1.3	7.4	3.5	4.7	0.5	3.0	0.2
[e]	2.2	1.6	3.3	6.6	2.6	3.0	3.4	2.2	1.5	3.2

Vowel	Lips					Tongue constriction					Larynx tube				
	$\frac{\Delta F_1}{F_1}$	$\frac{\Delta F_2}{F_2}$	$\frac{\Delta F_3}{F_3}$	$\frac{\Delta F_4}{F_4}$	$\frac{\Delta F_5}{F_5}$	$\frac{\Delta F_1}{F_1}$	$\frac{\Delta F_2}{F_2}$	$\frac{\Delta F_3}{F_3}$	$\frac{\Delta F_4}{F_4}$	$\frac{\Delta F_5}{F_5}$	$\frac{\Delta F_1}{F_1}$	$\frac{\Delta F_2}{F_2}$	$\frac{\Delta F_3}{F_3}$	$\frac{\Delta F_4}{F_4}$	$\frac{\Delta F_5}{F_5}$
[a]	2.5	3.7	1.3	3.0	0.6	3.2	3.7	2.5	1.1	1.9	2.2	1.4	0.6	0.6	14.5
[o]	3.6	5.7	1.7	0	0.4	3.6	4.5	2.1	1.6	1.8	1.8	4.3	0.4	7.0	9.2
[u]	6.7	2.1	0	0.3	0	0.5	3.5	4.5	0.7	3.7	0.5	0.8	0.2	8.0	11.8
[i]	1.0	0.4	0.4	0.6	0.5	5.8	1.3	1.5	2.0	2.2	1.7	0	2.5	2.8	5.0
[ɪ]	0.9	0.2	5.6	1.1	6.6	5.2	0.4	5.2	0.3	4.0	0.9	2.9	0.8	11.5	1.3
[e]	1.4	1.0	2.4	5.3	4.3	6.0	1.6	1.8	2.2	2.6	0.5	1.8	4.2	3.9	8.2

From these data the following conclusions can be drawn:

The first formant. The frequency of the first formant F_1 is generally dependent more on the back cavity volume than on the volume of other cavities. An exception is the vowel [ɑ], where F_1 is affected equally on a percentage basis by a change in the front cavity volume, and by a change of the back cavity volume. Since the back cavity of [ɑ] is much shorter than the front cavity, the percentage increase of F_1 due to the removal of a small unit length section of the back cavity is larger than the shift caused by a removal of a section of the same length in the middle of the front cavity.

F_1 of the vowels [e], [i], and [ɪ] is almost completely determined by the back cavity volume and the narrowest section of the mouth cavity. In the vowel [u] F_1 is tuned by the lip section much more than by the tongue constriction section. The contribution to F_1 of [u] from the back cavity volume is somewhat larger than that from the front cavity.

The second formant. Only in the case of the vowel [i] was the mouth cavity with associated orifices found to be the essential determinant of F_2 . F_2 of [i] is clearly a half-wavelength resonance of the back cavity. There is a similar but not at all so apparent tendency for F_2 of [e] to be influenced more by the back than by the front cavity. The second formant of the back vowels [u], [o], and [ɑ] is somewhat more dependent on the front cavity than on the back cavity. Providing the cavity volume changes are introduced on a constant percentage basis, this tendency is apparent, but if the volume changes are performed by means of a constant length reduction, there is found an equal dependency of F_2 on the two cavities for [u] and also for [ɑ]. In the case of [u], F_2 is dependent much more on the relative dimensions of the tongue pass than on the lip section. These two parts of the compound resonator system have about the same effect on F_2 of both [ɑ] and [o]. The lip section is of practically no importance for F_2 of [i] and does not have a very marked influence on F_2 of [e] either.

The third formant. According to the tabulated data, F_3 of [u], [o], and [ɑ] is chiefly dependent on the parts in front of the tongue constriction. This is also true of F_3 of [i]. The back cavity system is the main determinant of F_3 of [i]. In the case of [e], the dependency is more equally divided on all sections of the vocal tract. More specifically, it can be concluded from the physical boundary conditions of the separate cavity systems that F_3 of [i] is essentially a quarter-wavelength resonance on each side of the tongue constriction, which can also be interpreted as a half-wavelength resonance of the tongue passage plus front cavity. F_3 of [i] is a half-wavelength resonance of the back cavity and F_3 of [u] is a half-wavelength resonance of the front cavity. In [ɑ] and [o] F_3 is associated with a three-quarter-wavelength resonance of the cavity system in front of the tongue constriction.

The fourth and fifth formants. Besides the marked influence of the larynx cavity on F_4 of [o], [u], and [i], it can be seen that the cavity system in front of the tongue constriction has an appreciable effect on [ɑ], [i], and [e], while the back cavities have a greater influence on F_4 of [u], [o], and [i]. F_5 is also substantially influenced by the

larynx tube, except in [i], where the front cavity is the main determinant. The front cavity of [u] and the back cavity of [ɑ] also have a close connection with F_1 .

It is of some interest to investigate the effect of a volume change caused by an increase of the cross-sectional area of a unit section at a specific place within the model. Such area variations were introduced at the places of maximum cross-sectional area in each of the two main cavities.

TABLE 2.33-5

*Percentage decrease in formant frequencies due to a one per cent increase in front cavity volume and separately in back cavity volume by means of an area increase of a unit section at the places of maximum cross-sectional area. Negative values marked by *.*

Vowel	$\frac{\Delta F_1}{F_1} \frac{C_1}{\Delta C_1}$	$\frac{\Delta F_2}{F_2} \frac{C_1}{\Delta C_1}$	$\frac{\Delta F_1}{F_1} \frac{C_2}{\Delta C_2}$	$\frac{\Delta F_2}{F_2} \frac{C_2}{\Delta C_2}$	$\frac{\Delta F_3}{F_3} \frac{C_1}{\Delta C_1}$	$\frac{\Delta F_3}{F_3} \frac{C_2}{\Delta C_2}$
[ɑ]	0.07	0.19	0.23	0.11	0.18	0
[o]	0.05	0.33	0.37	0.22	0.25	0
[u]	0.18	0.28	0.20	0.15	0.06	0
[i]	0.02*	0.39	0.49	0.02*	0.04	0.45
[ɪ]	0.01*	0	0.53	0.39*	0.04	0
[e]	0.08*	0.01*	0.42	0.26*	0.01	0.23

These data should coincide with those of *Table 2.33-3*, provided the volume changes caused by a local area increase and those caused by a length increase had the same effect. This is apparently not the case on account of the finite wave propagation. A local area increase in the back cavity of [u] has a smaller effect on F_1 here than in the data of *Table 2.33-3*. The influence of the local front cavity area change on F_1 of [o] and [ɑ] is relatively small.

Apparently the appreciable shift of F_1 of [ɑ] caused by a shortening of the front cavity length is due to the fact that the removed section had a function more like an inductance than like a capacitance. This effect should be compared with the change in resonance frequency of a single tube closed at one end. A small area change in the middle of the tube has no effect on the resonance frequency. An area increase at the closed end has a maximal lowering effect on the resonance frequency and a similar operation at the open end has a maximal increasing effect, compare *Fig. 1.4-4*. It can also be shown that a doubling of the cross-sectional area localized at a place somewhat anterior to the middle of the tube has no effect on F_1 .

As a general conclusion of these differential investigations, it can be stated that although the double Helmholtz resonator theory provides fairly satisfactory results for the prediction of F_1 and F_2 of the vowels [u], [o], [ɑ], and [ɪ] and of F_1 of [i] and [e], the treatment of the cavities as lumped element capacitances leads to an underestimation of the F_2 -front-cavity affiliation and to some extent also of the F_1 -back-cavity affiliation. However, in [u], [o], and [ɑ], both cavities contribute substantially to both F_1 and F_2 and they have an approximately equal share in determining F_2 and

F_3 of [e]. All the variational investigations performed here indicate the very marked back cavity dependency of F_2 of [i] and the front cavity affiliation of F_3 of this vowel. The same is true of the back cavity affiliation of F_3 of [i]. The lip orifice has an especially strong influence on F_1 of [u], F_2 of [i], and F_3 of [i] and has approximately the same relative influence on F_1 and F_2 of both [o] and [a]. The main tongue constriction, i.e., the back cavity neck, influences F_1 of the front vowels [i], [i], and [e] more than the other formants. In the case of [u] and [o] the degree of tongue constriction is more critical for F_2 than for F_1 and in the case of [o] F_1 and F_2 are equally dependent on this articulatory variable.

Double resonator theory should be used with great caution, and it is not possible to utilize this or any other theory for developing formulas for cavity-formant relations which satisfy the requirements of both simplicity and general validity. The greatest objection to the use of the double resonator theory is that it is worthless for calculating F_2 of high front vowels.

Some of the restrictions encountered in the use of double resonator theory may be avoided if the concept of formant is reinterpreted with the aid of perceptive criteria. It is known from synthesis experiments⁴ that almost all vowels can be satisfactorily synthesized on the basis of two formants. In back vowels and other sounds where F_2 is closer to F_1 than to F_3 , the synthesis can be performed on the basis of the natural position of F_1 and F_2 . In typical front vowels, on the other hand, in which F_2 is closer to F_3 the upper formant of the synthetic sound should be placed closer to F_3 and in case of a very high [i] even above F_3 . In view of the previous analysis it can therefore be stated that the mouth cavity thus exerts a marked influence on F_2 or to the perceptive mean of F_2 and all higher formants, cf., the front cavity affiliation of F_3 of [i]; Fant (1957). These facts supporting a dual resonance theory of vowel color cannot, of course, fully reestablish the rule of low formant—back resonator, and high formant—front resonator affiliation. The low formant is always dependent on the entire vocal tract, and the same is true of the second formant of back vowels. In this connection it should be observed that those vowels where F_2 is very close to F_1 can be synthesized on the basis of a single formant representing their perceptive average which lies approximately halfway between F_1 and F_2 of [a] but closer to F_1 of [u]. This single formant frequency is close to the uncoupled resonance frequency of the front resonator.

If the available vocal tract data of a sound have been condensed into the four elements entering the double Helmholtz resonator theory (which according to the previous discussion, is far from being an unambiguous operation), it is convenient to utilize the following approximate formula for calculating the two resonance frequencies:

⁴ Delattre et al. (1951, 1952).

$$F_{low} = \frac{c}{2\pi} \left[V_2 \left(\frac{l_1}{A_1} + \frac{l_2}{A_2} \right) \right]^{-\frac{1}{2}} \quad (2.3-3)$$

$$F_{high} = \frac{c}{2\pi} \left[\left(\frac{A_1}{l_1} + \frac{A_2}{l_2} \right) \frac{1}{V_1} \right]^{\frac{1}{2}}$$

These approximations hold well, provided

$$\frac{V_1}{V_2} \ll 1 + l_2 A_1 / l_1 A_2, \quad (2.3-4)$$

which is not the case for back vowels, where 10-20 per cent too low values of F_{high} are obtained. In general, however, Eq. 2.3-3 and 4 above can be interpreted with a greater generality on a variational basis with the understanding that in back vowels both volumes have a noticeable effect on both resonances. The relations are then the following, which have been made much use of by Jakobson et al. (1952).

The first formant increases with a decrease of back cavity volume and with an increase of the conductivity index A/l of any constricted passage. The upper formant increases with a decrease in front cavity volume and with an increase in the conductivity index of any of its two associated resonator necks, i.e., by a delabialization or by the introduction of a larger opening at the internal tongue constriction, i.e., at the neck of the back resonator. If the two resonator necks have conductivities of unequal magnitudes, it is apparent that the upper formant is more influenced by the neck that has the largest conductivity and that the lower formant is more dependent on the neck that has the lowest conductivity, i.e., the largest l/A ratio. If the two necks have an equal conductivity, the lower formant will be more dependent on the one farthest away from the glottis, i.e., the lip section.

The two volumes and the two necks can seldom, however, be varied independently. In front vowels a lowering of the tongue causes not only an increase of the conductivity of the internal resonator neck, but also a front cavity volume increase. This effect may be further enhanced by an increased jaw-opening. These two concomitant but opposing resonator element variations may thus partially cancel each other, tending to keep the frequency of the upper formant constant. Thus at a mid-palatal point of articulation, F_2 decreases and F_1 increases as the result of this articulatory movement. Similarly, in back vowels of the type [ɑ] a narrowing of the pharyngeal tongue passage is generally followed by a decrease in back cavity volume. The result is a partial cancellation of the shift in the lower formant. The upper formant will, however, decrease in frequency since the front cavity volume remains approximately constant.

All the restrictions and difficulties encountered in the application of analytical formulas are avoided if the effects of articulation on the formant pattern are studied on a purely correlational basis, that is in terms of the behavior of three-parameter models of the type described by Stevens and House (1955, 1956) and further developed in Section 1.63 of this work. The procedure is to evaluate the place of articulation and the degree of opening at this place and further the length and the cross-sectional area

of the lip section. Nomograms relating these data to the resonance frequencies of the vocal tract should replace or at least supplement existing analytical resonator formulas in phonetic textbooks. It has been found that the frequencies of the first three formants of the six Russian vowels analyzed here can be evaluated within an accuracy of 5-10 per cent by means of the dimensional data of *Table 2.33-1B* appropriately inserted in the nomograms of *Fig. 1.4-11*.

2.34 The Spatial Distribution of Sound Pressure. Formant Bandwidths

A. CALCULATIONS OF FORMANT LEVELS ON AN ABSOLUTE PRESSURE BASIS

The following sections are devoted to a study of the sound pressure distribution within the vocal cavities and in the radiated speech wave and of the contribution from various dissipative elements to the damping of vowel formants, i.e., to their bandwidths.

Both numerical and analog methods of calculations described in *Section 1.42* and *Chapter 2.2* respectively have been applied to the vocal tract configurations of the six Russian vowels analyzed in this work. The electrical line analog *LEA* is especially useful for such studies since the voltage distribution along the line reflects the pressure distribution within the vocal tract. The effect of introducing or removing a definite dissipative element at a given place within the model can readily be studied. The numerical methods are, however, superior for introducing distributed dissipative elements according to a given formula for the attenuation constant within the vocal tract, and they are feasible, provided a high speed digital computer can be utilized. It was shown in *Section 2.31* that the formant frequencies calculated with the aid of the *BESK* computer and *LEA* compare very well up to the frequency of the third formant. The two methods are thus interchangeable to a certain extent.

Table 2.34-1 contains the data on formant bandwidths and formant sound pressure levels calculated by the numerical method and previously displayed by the spectrum curves of *Fig. 2.3-4*. The only dissipative elements made use of here are the radiation resistance and the distributed losses within the resonator sections. The latter were defined by the attenuation constant

$$\alpha = 0.007(\pi/A)^{1/2} \cdot (1/2 + f/4000), \quad (2.3-5)$$

which was empirically constructed to provide a natural range of formant bandwidths.⁵ The area dependency is the same as for classical resonator wall losses, but the frequency dependency is based on a constant and a linear term instead of the half-power relation required in the theory of hard-walled systems. The linear and constant terms of the formula above are equal at a frequency of 2000 c/s. This is a practical compromise due to the existence of dissipative elements of negative frequency power dependency in the vocal tract, i.e., cavity wall vibration damping, and non-linear orifice damping.

A constant volume velocity voice source was utilized. The DC flow component

⁵ Eq. 2.3-5 is introduced here merely to make calculations of formant levels possible. Stevens (1958) has successfully derived formant bandwidths from data on the impedance of human tissue; see *Section A.36-B.3*.

was assumed to be $150 \text{ cm}^3/\text{sec}$, and the waveform was defined by a double pole at 100 c/s on the negative real frequency axis. The source spectrum envelope can thus be expressed by

$$|U_q(f)| = \frac{150}{1+f^2/100^2} \text{ cm}^3/\text{sec}, \quad (2.3-6)$$

falling -12 dB/octave at frequencies well above the cutoff frequency 100 c/s . This is the standard source adopted for all calculations on voiced sounds in this work.⁶

TABLE 2.34-1

Formant bandwidths in c/s and formant sound pressure levels in dB re 0.0002 dyne/cm². Calculated data.

Vowel						Formant levels								
						At 5 cm in front of the lips			At the lips			At the glottis		
	<i>B</i> ₁	<i>B</i> ₂	<i>B</i> ₃	<i>B</i> ₄	<i>B</i> ₅	<i>L</i> ₁	<i>L</i> ₂	<i>L</i> ₃	<i>L</i> ₁	<i>L</i> ₂	<i>L</i> ₃	<i>L</i> ₁	<i>L</i> ₂	<i>L</i> ₃
[a]	57	72	130	175	200	91	85	70	111	104	87	135	121	97
[o]	54	65	100	135	155	90	86	66	112	107	86	134	121	95
[u]	69	50	110	115	110	84	73	37	114	101	68	137	121	86
[i]	43	125	77	134	140	88	67	69	108	82	85	136	86	104
[ɪ]	60	75	240	230	330	84	68	66	106	86	83	136	108	88
[e]	39	95	170	325	310	90	77	70	108	93	92	132	108	105

The formant bandwidths have been determined on the basis of the *BESK* calculations by means of the numerical method of *Eq. 1.22-21*. The sound pressure level relations between the separate formants of the radiated sound were obtained from the *BESK* calculations and included the $K_T(\omega)$ factor, *Fig. 1.2-6*, for the pressure increase in excess of the ω -proportionality. The relation between sound pressure at the lips and at the point $l = 5 \text{ cm}$ in front of the lips was obtained by means of *Eq. 1.23-13* relating the sound pressure in the radiated wave to the volume velocity through the lips:

$$\frac{P_{lip}}{P_l} = \frac{0.8(A/\pi)^{\frac{1}{2}} 4\pi l}{A K_T(\omega)}. \quad (2.3-7)$$

The lip pressure is that across a radiation inductance of $L_0 = 0.8(A/\pi)^{\frac{1}{2}} \rho/A$.

The transfer function relating sound pressure at the lips to sound pressure at the input of the vocal tract model was determined by the corresponding voltage ratio on *LEA*. Finally, the absolute sound pressure levels were determined from the volume velocity of the glottis input source at the frequency of the first formant of [a] as determined from *Eq. 2.3-6*. This is to be multiplied by the appropriate input impedance as measured on *LEA* to yield the corresponding sound pressure. This

⁶ Compare *Eq. 1.3-2*, where in addition the radiation transfer enters.

impedance must be corrected with respect to the difference in formant bandwidth between *LEA* and *BESK* calculations.

The calculated pressure levels have a fair degree of representability since the particular source function adopted for the calculations has resulted in normal formant levels of the produced sounds. The absolute fixation of all pressure and flow levels is based on the *DC* flow component of $150 \text{ cm}^3/\text{sec}$, which is close to the average air consumption during phonation (Chiba and Kajiyama, 1941).

The only comparable data of measured formant levels are those from a study of Swedish vowels (Fant, 1948). In the following tabulation the measured Swedish data of formant levels are indicated by the symbol *S* and the calculated Russian data by *R*. The data refer to a speaking distance of 5 cm :

TABLE 2.34-2

Vowel		L_1/F_1	L_3/F_3	L_9/F_9
a	<i>R</i>	91/616	85/1070	70/2375
	<i>S</i>	91/680	86/1070	64/2520
e	<i>R</i>	90/432	77/1960	70/2720
	<i>S</i>	91/440	76/1795	73/2385
i	<i>R</i>	84/222	68/2240	66/3140
	<i>S</i>	88/255	64/2065	67/2960
u	<i>R</i>	84/231	73/615	37/2375
	<i>S</i>	89/310	76/730	47/2230

Some of the differences in formant levels can be explained by the differences in formant frequencies. Thus the 10 dB higher L_3 of the measured [u] can be partially accounted for by the residue factor $40 \log_{10}(\frac{310}{251} \cdot \frac{730}{615}) = 8 \text{ dB}$ which follows from Eq. 1.3-9. The measured data for L_1 show rather small variations among the *close* and *open* vowels, i.e., those with low and high F_1 . This can be ascribed to an average source spectrum of the speaker sloping at a faster rate than the function adopted for the calculations, at least in the frequency range between 200 and 1000 c/s . There is also an indication that the human voice source slopes off at a faster rate than -12 dB/octave at frequencies above 2000 c/s .

However, formant pressure amplitudes are also dependent on formant bandwidths by an inverse linear relation, so that the level of a formant decreases 6 dB per doubling of its bandwidth. Reliable data on bandwidths for each separate formant of actual speech are not available, but it appears that the calculated bandwidths do not generally deviate more than 50 per cent from typical actual values. A systematic factor of minor importance is the difference between formant intensity level and formant envelope level, the latter concept having been utilized in the calculations. Provided the formant bandwidth is smaller than the fundamental pitch F_0 , the former is no more

than 1 dB above the latter. In any discussion of formant levels and voice source spectra, it should also be remembered that the average spectrum slope differs with voice effort, high voice efforts being associated with a relatively weak voice fundamental and with a high level of the formants of higher frequencies relative to the level of the first formant which normally carries the main intensity of a voiced sound; see *Section A.21*.

B. SOUND PRESSURE DISTRIBUTION WITHIN THE VOCAL TRACT

The spatial distribution of sound pressure within the vocal cavities from the lips to the glottis can be studied in *Fig. 2.3-6*, which contains a separate curve for each of the first three formants of each of the six vowels. These curves represent absolute sound pressure, phase disregarded, portrayed on a linear amplitude scale that has been calibrated in dB sound pressure level. At the frequency of the first formant the sound pressure rises from a minimum value at the lips to a maximum value at the glottis end. In the case of a typical front vowel such as [i] where the whole vocal tract behaves like a simple resonator for the first formant, the sound pressure shows a fairly linear rise in the mouth part representing the resonator neck and is fairly constant in the pharynx cavity. This distribution is evident from the network representation of the neck of the resonator as an inductance and the main volume as a capacitance. The extent of pressure rise is apparently determined by the ratio of resonator neck inductance to the radiation inductance, and in the case of a uniform neck, the ratio of its physical length to its correction. The L_1 -distribution curve for [u], on the other hand, is characterized by an abrupt pressure increase behind the lips, and the sound pressure is not much smaller in the front than in the back cavity owing to the dominance of the lip-passage inductance.

At the frequency of the second formant the pressure curves must have an additional minimum, and at the third formant there should be two pressure minima between the front and the back ends of the system. The L_2 -minimum generally occurs in the upper half of the pharynx cavity and does not vary much in location. A reciprocal argument therefore leads to the conclusion that a throat microphone should not be placed too high up on the throat and preferably at the level of the larynx cavity (van den Berg, 1953). The sound pressure minima are in some instances damped out to a considerable degree, which is true of the third formant of [i] and the second formant of [i].

In a single tube resonator the pressure response $P(x, \omega)$ at the point x cm from the open front end to a volume velocity source $U_q(\omega)$ located at the opposite end of the l cm long tube is

$$\frac{P(x, \omega)}{U_q(\omega)} = Z \frac{\sinh(j\omega/c+a)x}{\cosh(j\omega/c+a)l}. \quad (2.3-8)$$

Similarly the volume velocity transfer ratio is

$$\frac{U(x, \omega)}{U_q(\omega)} = \frac{\cosh(j\omega/c+a)x}{\cosh(j\omega/c+a)l}. \quad (2.3-9)$$

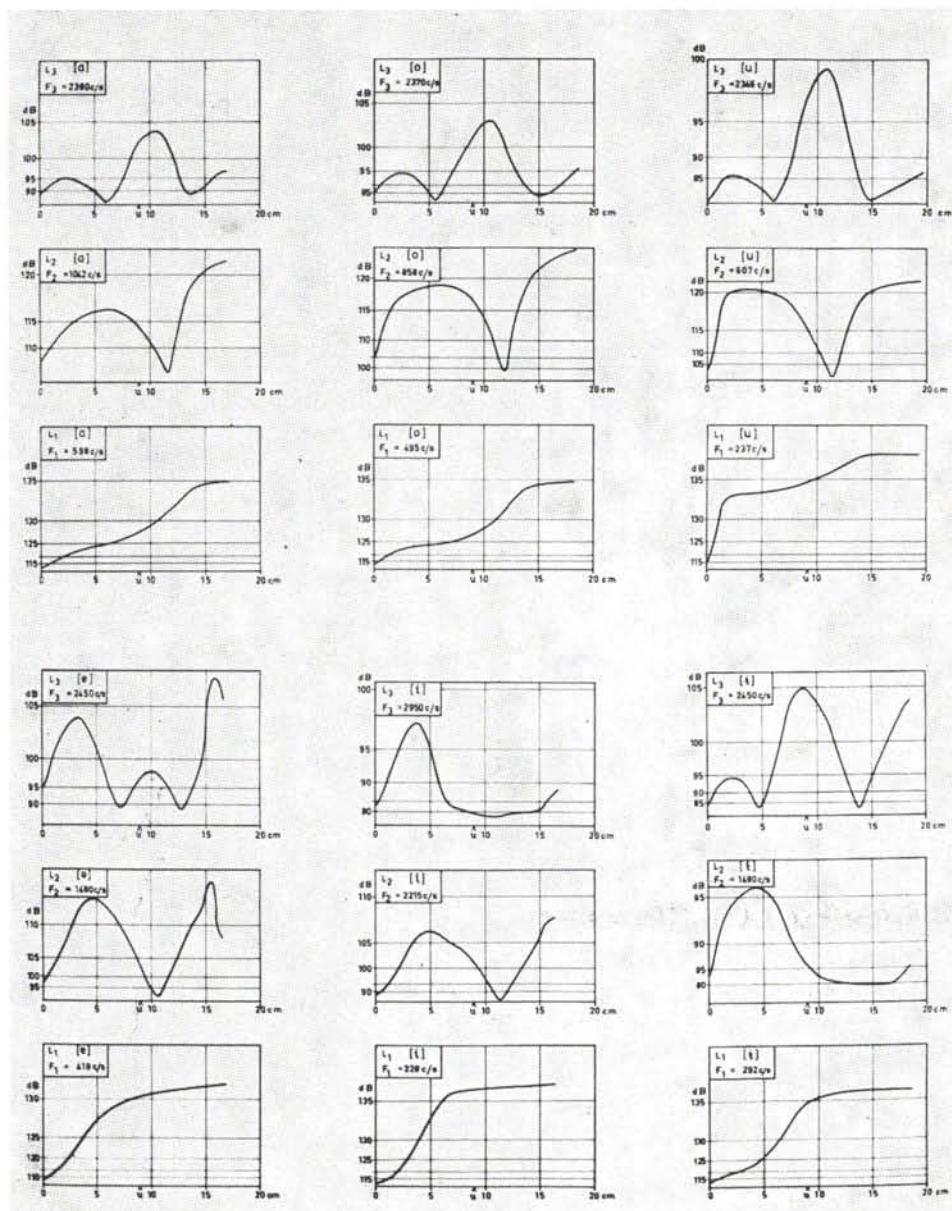


Fig. 2.3-6. Calculated sound pressure level distribution within the vocal tract for each of the first three formants of the Russian vowels specified by the area functions of Fig. 2.3-2. The place coordinate has its zero at the lips.

The resonance frequencies are those at which $\cos \omega l/c = 0$. At the frequency of the second formant, $\omega_2 l/c = 3\pi/2$, the pressure has a minimum when $\sin \omega_2 x/c = 0$, i.e., at $x = 2l/3$, which is the general tendency observable in the F_2 -diagrams of Fig. 2.3-6. The ratio of maximum to minimum pressure response is apparently $1/ax$, and the maximum occurs at $x = l/3$. The volume velocity is displaced 90 degrees versus the pressure in time and space, the maxima of the velocity coinciding with the minima of velocity and vice versa, as implied by the sine and cosine functions.

These relations are well known from the elementary theory of standing waves in tubes and on transmission lines. In general, regardless of the shape of the vocal cavities, the pressure zeros occur at places where the reactance looking towards the lips is zero. Owing to the simultaneous condition of resonance, the reactance looking back towards the glottis must also be zero. Under conditions of heavy damping, the only recognizable cue to the location of the spatial zero may be the phase of the pressure, which is close to an odd integer of 90 degrees and shows a maximum rate of changes at these pressure minima nodes. The pressure and volume velocity curves presented by Chiba and Kajiyama (1941) were calculated for the loss-free case, but show the same general features as our data.

The anterior pressure maximum of the second formant falls somewhere in the palatal region. This has also been predicted by van den Berg (1955a), who suggests that the vibrations of the soft palate at the frequency of the second formant evoke a sensation that plays some role as a feedback signal for adjusting the voice character in singing. From the pressure data of Fig. 2.3-6 it would seem more likely that the effective stimulus is the first formant which is of a level approximately 15 dB higher in this region and lies in a frequency range where the tactile sensitivity is greater; see Rösler (1957). The absolute sound level L_1 at the soft palate is as high as 136 dB for the vowel [i] at a medium-speaking voice effort and at least 10 dB higher values may be expected in singing.

It is of some interest to see how our pressure distribution curves compare with measurements. The probe measurements of sound pressure levels inside and outside the mouth reported on by von Békésy (1949) and the average sound pressure level curve of the first formant of the vowels [o], [ɑ], and [e] of our calculated data may be studied in Fig. 2.3-7. The two sets of data are adjusted to the same level at the edge of the lips. They coincide within 2 dB at other places.

In view of the high-sound pressure levels in the pharynx cavity, the question arises, to what extent the first formant is transmitted directly through the walls of the vocal cavities. One simple estimation is the following. The cavity walls are assumed to be perfectly mass-controlled and will be represented by an area of 50 cm^2 , a thickness of 1 cm, and a density of 1 g/cm^3 . The resulting inductance, $20 \cdot 10^{-3}$ acoustical units, is only four times as large as that of a narrow mouth channel which would tune a pharynx cavity of volume 80 cm^3 to the frequency of 300 c/s typical of F_1 of the vowel [i]. The shunting effect of the vibrating cavity walls is thus to cause a frequency increase of F_1 by the factor $(1+0.25)^4$ which is of the order of 10 per

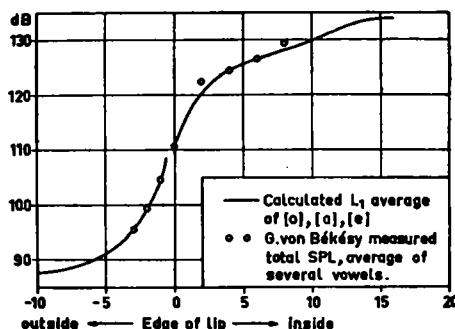


Fig. 2.3-7. Calculated average sound pressure levels outside and inside the mouth for the first formant of the vowels [o], [a], [e] compared with measurements made by von Békésy (1949) on overall sound pressure levels. The absolute level chosen for the latter data is that providing equal sound pressure level at the edge of the lips. The place coordinate is the distance in cm from the lips.

cent and hence of a magnitude large enough to be taken into account in calculations. Also, the volume velocity flow conveyed by the walls would be $1/4$ of that passed by the lips, and the radiated sound pressure from the throat would thus be -12 dB relative to the normal mouth output as measured at a sufficient distance from the speaker. This is only a crude estimate which, however, has some support in the measured data on the first formant of a voiced occlusion preceding a stop sound [b], [d], or [g]. Under these conditions of perfect closure the sound is conveyed by wall vibration only.⁷ The estimated total wall inductance would correspond to a first formant frequency of the order of 150 c/s as the limiting low value of F_1 under any condition. Measurements on the voiced occlusion indicate an F_1 of this order of magnitude and a level about 12 dB below the level of L_1 of [i].

The spatial distribution of sound pressure may not be directly interpreted as a curve of cavity-formant dependency. It should first be normalized with regard to the cross-sectional area. Also, the volume velocity plays an equally important role. The reactive energy per cycle within a section of the vocal tract of cross-sectional area $A(x)$ is the sum of the potential energy which is proportional to $A(x)P^2(x)$ and the kinetic energy which is proportional to $U^2(x)/A(x)$. The sum of these two terms expressed as a function of the place coordinate would be an appropriate measure of the dependency of a specific formant on various parts of the system. Of more immediate interest to the theory of speech production is the conclusion that the effect of shunting the vocal tract system with an extra branch is greatest at those places where the sound pressure is high and that the effect of an added series element is greatest

⁷ There is also some low frequency radiation from the chest to be taken into account, as discussed by van den Berg et al. (1957). This sound originates from that lost to the subglottal system via airborne vibrations. A mapping of the vibrational amplitude at various parts of the human body has been performed by von Békésy (1949).

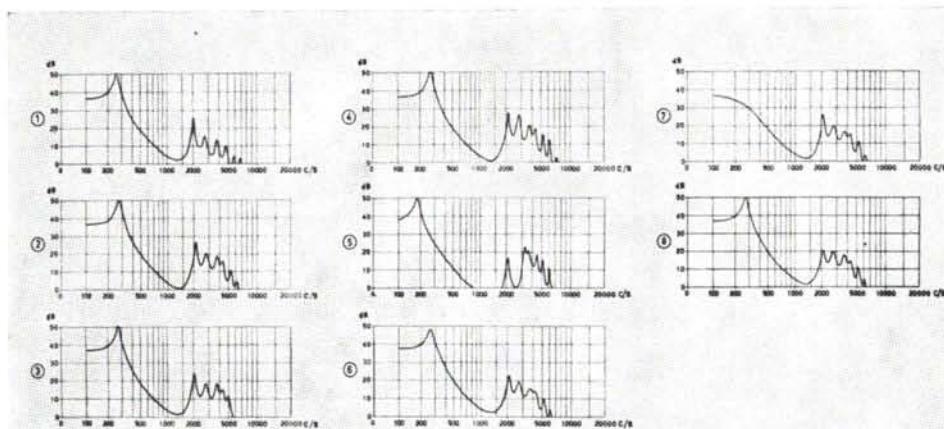


Fig. 2.3-8 Variation of factors affecting the damping of formants of a vowel [i] produced by the horn-shaped model of Fig. 2.2-4.

- (1) (2) (3) Identical with those of Fig. 2.2-4;
- (4) Larynx configuration that of (2), the radiation resistance reduced to one-half its normal value;
- (5) Same as (4) except for a reduction of the tongue constriction cross-sectional area by a factor of 2. The double peak of $F_3 + F_4$ is apparent;
- (6) Same as (2) except for a reduction of the glottis resistance from 5 q.c. to 2.5 q.c. ;
- (7) Same as (2) except for the introduction of a 0.25 q.c. resistance in series with the tongue constriction;
- (8) Same as (2) except for the introduction of a series resistance in the middle of the pharynx.

at those places where the volume velocity is high. The resistive part of the added impedance element adds to the damping of the formant, and the reactive part causes a frequency shift of the formant.

C. THE DEPENDENCIES OF FORMANT BANDWIDTHS ON VARIOUS RESISTIVE ELEMENTS WITHIN THE VOCAL TRACT

The damping effect of various resistive elements within a vocal tract model of the vowel [i] will next be illustrated. The idealized configuration of a horn-shaped mouth cavity and a pharynx of homogeneous cross-sectional area as shown in *Fig. 2.2-4* is utilized in *Fig. 2.3-8*. Curves 1-3 of these figures are identical, representing normal damping conditions within *LEA*, i.e., normal radiation damping, a glottis resistance of $5 \text{ } \mu\text{c}$, and the minimum possible values of losses in the coils and condensers of the analog. Curve 1 pertains to a pharynx without larynx tube. In curve 2 the larynx cavity is included and in curve 3 the sinus piriformis, shunting the outlet of the larynx cavity, have been taken into account.

This shunt was removed from the vocal tract model in the experiments on which curves 4-8 were based. In curve 4, the resistive load of the radiation resistance has been reduced to one-half its normal value. The most apparent effect is the rise in level and the associated decrease in bandwidth of the third and some of the higher formants. Curve 5 was produced with the same radiation resistance as in curve 4, but the width of the palatal constriction was reduced to the extent that F_1 was lowered to 180 c/s and was thus more representative of the consonant [j]. The lowering of F_1 has the usual effect of decreasing the level of higher formants. Because of the frequency rise of F_3 to a position close to F_4 , displayed in curve 5, there is an increase of the levels of both these formants. Here F_3 is the fundamental resonance of the mouth cavity and F_4 of the larynx cavity. In the case of a normal value of the radiation resistance, these two resonances would have formed a single peak. This seems often to be the case with the third formant of [j] containing $F_3 + F_4$.

The conditions for curve 6 differ from those of curve 2 by a reduction of the glottis resistance to $2.5 \text{ } \mu\text{c}$, that is, an increase of its shunting effect by a factor of 2. As seen by the pressure distribution curves, the pressure at the glottis end of the vocal tract is fairly high for all formants of all vowels. Accordingly the glottis resistance affects all formants, as can be seen from curve 6.

The effect of introducing a resistance of $0.22 \text{ } \mu\text{c}$ in series with the tongue constriction may be studied in curve 7, which differs from curve 2 mainly by the larger damping of the first formant. This is to be expected considering the high volume velocity of F_1 in the resonator neck. The almost 100 per cent affiliation of the second formant with the half-wavelength resonance of the back cavity is reflected in the absence of any perceivable damping. The third as well as the second formant has a pressure maximum at the palatal constriction, and thus low volume velocity in this region. A series resistance placed in the middle of the pharynx cavity, on the other hand, has a sub-

stantial effect on the level and bandwidth of the second formant as can be seen from curve 8.

A variational investigation of the influence of radiation resistance, of glottis resistance, and of classical viscosity and heat conduction losses on the bandwidths of each of the first five formants of the six separate vowels was undertaken. The effect of radiation and glottis termination was evaluated by the aid of *LEA*, and the effect of the cavity wall damping was determined from a series of numerical calculations with the *BESK* computer. The results of the direct measurements on *LEA* of the effect of radiation and of glottis termination on the bandwidths of the first three formants were in substantial agreement with numerical calculations carried out with *BESK*, but it was judged that *LEA* gave more accurate data for higher formants.

TABLE 2.34-3

Differential contribution to formant bandwidths in c/s from: A. The radiation resistance; B. A glottis parallel resistance of 5 qc; C. Frictional and heat conduction losses assuming ideal hard-walled conditions as specified by the attenuation constant:

$$a = 2.92(\pi/A)^{1/4} \cdot 10^{-5} f^{1/4} S_A \text{ Nepers/cm, in which } S_A = 2 \text{ is a shape and surface factor.}$$

Vowel	A					B					C			A+B+C		
	<i>B</i> ₁	<i>B</i> ₂	<i>B</i> ₃	<i>B</i> ₄	<i>B</i> ₅	<i>B</i> ₁	<i>B</i> ₂	<i>B</i> ₃	<i>B</i> ₄	<i>B</i> ₅	<i>B</i> ₁	<i>B</i> ₂	<i>B</i> ₃	<i>B</i> ₁	<i>B</i> ₂	<i>B</i> ₃
a	4	13	35	110	29	84	70	24	52	194	17	20	33	107	103	92
o	3	13	14	1	14	52	32	12	230	94	16	19	31	71	64	57
u	0	0	1	0	0	18	16	12	270	48	15	16	40	33	32	53
i	0	50	8	23	24	18	4	44	48	88	11	16	24	29	70	76
ɛ	0	2	190	43	400	16	88	44	172	76	14	22	36	30	112	256
e	3	28	85	240	255	20	50	140	76	112	11	19	27	34	107	252

Under conditions of small damping, i.e., narrow bandwidths, the frequency of the formant peak coincides with the frequency at which the input impedance to the vocal tract network model is purely resistive, in analogy to a parallel resonance circuit. The bandwidth increase caused by the source resistance stands in direct proportion to its conductivity. Conversely, the input conductance at the frequency of a formant is proportional to the bandwidth and is determined by dissipative elements exclusive of the source resistance. The bandwidth data of *Table 2.34-1* may accordingly be utilized for calculations of the corresponding vocal tract input resistances. Computed values range from 1.6-7.3 qc for the first formant, 0.14-5.8 qc for the second formant, and 0.45-4.2 qc for the third formant. These are of the order of magnitude of qc, as previously derived by van den Berg (1953) for an [i]-model.

The damping effect of a glottis resistor is not dependent on its absolute value only, since the impedance level (*L/C*)^{1/4} of a resonance varies with the vocal tract configuration and the type of resonance. Thus the first formant of [a] is affected much

more than the first formant of [u], [i], [ɪ], or [e]. This could be anticipated from the narrow pharynx passage in [ɑ] representing a high characteristic impedance; see *Eq. 1.21-6* and *A.36-14*.

The input impedance of the vocal tract, as seen from the lips, in the frequency range of a formant is analogous to that of a series resonance circuit. Because of the ω^2 -dependency of the radiation resistance, its damping effect on the first formant is negligible; see further the analytical formulas and calculations in *Section A.36*. The damping effect of the radiation resistance and of the glottis resistance increases with the degree of front and back cavity affiliation, respectively, of the formant under consideration, as can be observed from a comparison of the data tabulated above with that discussed in *Section 2.32*. The bandwidth data thus reflect the fairly equal importance of back and front cavities for the first two formants of the back vowels [u], [o], [ɑ], and further, the very definite affiliation of the second formant of [i] to the back cavity, F_2 of [i] to the front cavity, and F_3 of [i] to the front cavity. Because of the transformer effect of the vocal cavities and the relatively small configurational variations in the region of the larynx tube, the variability of the glottis resistance damping is not so great as that of the radiation resistance.

The radiation resistance apparently does not contribute significantly to the damping of any of the formants of [u]. This could be expected from the small lip-opening, providing a large reactance in series with the radiation resistance. Also the effect of the 5 μ c glottis resistance on the damping of any of the first three formants of [u] is relatively small.

The last column of *Table 2.34-3* is intended for an evaluation of the bandwidths arrived at when summing up the influence of those dissipative elements that are analytically predictable. It should, however, be remembered that the inductance in series with the glottis resistance presumably causes a reduction of the glottis damping of the third and higher formants and of a high frequency second formant. The cavity wall frictional and heat conduction losses, as specified by the formula for ideal hard-walled systems, have been introduced in the calculations with the additional shape and surface factor $S_A = 2$; see *Sections 1.21* and *A.36*. Normal deviations of vocal tract cross-sections from the circular shape account for the greater part of this factor. The frictional properties of vocal cavity surfaces are not known, and they are not easily separable from the cavity wall vibrational losses, except that the former dominates at constricted passages which function as the neck of a Helmholtz resonator.

The bandwidth sums $A + B + C$ of *Table 2.34-3* are less representative than the data from *Table 2.34-1*, the latter being arrived at by neglecting the glottis resistance and introducing an empirical formula, *Eq. 2.3-5*, for the attenuation constant of the vocal cavities. The main objections are that the first two formants of [ɑ] have become too heavily damped by the glottis resistance and that the first formants of the other vowels (except [o]) have not been damped quite enough. Preliminary experimental investigations on the damping constants of human vocal tract resonances excited with a closed glottis show, rather consistently, bandwidths of the order of 50 c/s up

to 2000 c/s, and rapidly increasing values at higher frequencies. Resonances in the frequency region of 300-500 c/s may even have smaller bandwidths, down to 25 c/s. The B_1 -values of column C plus those of column A are not large enough to match these measured data. It seems probable that the cavity wall vibrations⁸ contribute somewhat to B_1 of front vowels, as suggested by van den Berg (1953). The damping effect of the glottis during phonation is probably that of a resistance of the order of 5-15 ρc . This resistance is time variable within a voice fundamental cycle.

The DC flow conditions for the vowel [u] fulfill the requirements for turbulent streaming at the lips which, according to Eq. A.36-11, would give rise to a B_1 -bandwidth increment of as much as 100 c/s on the basis of $u = 150 \text{ cm}^3/\text{sec}$, $A_{\min} = 0.2 \text{ cm}^2$, and an F_1 lip section dependency factor of 0.86 for [u], according to Table 2.33-3. The flow dependent bandwidth increment is of the order of 15 c/s only for [i], and the estimated $Re = 1000$ could lie below the critical level for turbulence.

⁸ According to recent investigations by House and Stevens (1958) the vocal tract cavity wall vibrational losses can account for a larger part of the damping of the first formants; see Section A.36-B.3.

2.4 NASAL SOUNDS AND NASALIZATION

2.41 Physiological Data

The theory of nasalization, i.e., the interaction of the nasal and oral cavities in the production of sounds without oral closure is closely related to the theory of nasal consonants proper, i.e., sounds produced with nasal-opening and totally obstructed oral pathway. From an acoustic point of view, and quite apart from phonemic classification, there are one ternary and two binary physiological distinctions of interest. The latter pertains to the function of the velum as a means of guiding the sound either to the mouth alone with the nasal tract closed, as in oral sounds, or the reverse, as in velar nasal consonants, or allowing sound propagation both to the mouth and to the nose. In this intermediate state of the velum, the opening versus closure within the mouth cavity differentiates *nasalized vowels* from the *nasal murmur*. The second binary distinction is a presence versus absence of an obstruction within the nasal cavity system. A blocked nasal passage due to the swelling of the soft tissue during a cold causes an apparent weakening of the nasal murmur and a quality change of the nasalized vowel. So-called *nasal vowels*, e.g., of the French vowel system, are produced with specific tongue positions in addition to the element of nasalization.

Nasalization occurs, as a rule rather than as an exception, in vowels positioned close to, or more often between, two nasal consonants. Conversely, in unstressed positions, the assimilated nasality may become combined with a shortening or complete absence of the sound interval of nasal murmur of a nasal consonant, in which case the nasalization may become the major cue of the presence of a nasal phoneme. A small degree of nasalization is probably a normal attribute of very open vowels. The tendency of a lowering of the velum in open vowels, as can be observed for [a] and [e] of Fig. 2.3-1, is a well-known phenomenon.

Of the entire vocal tract, the nasal tract is less accessible for measurements than the mouth-pharynx-larynx system. The apparent anatomical variability of the width of the nasal passages and of the amount of mucous filling in cavities and constrictions

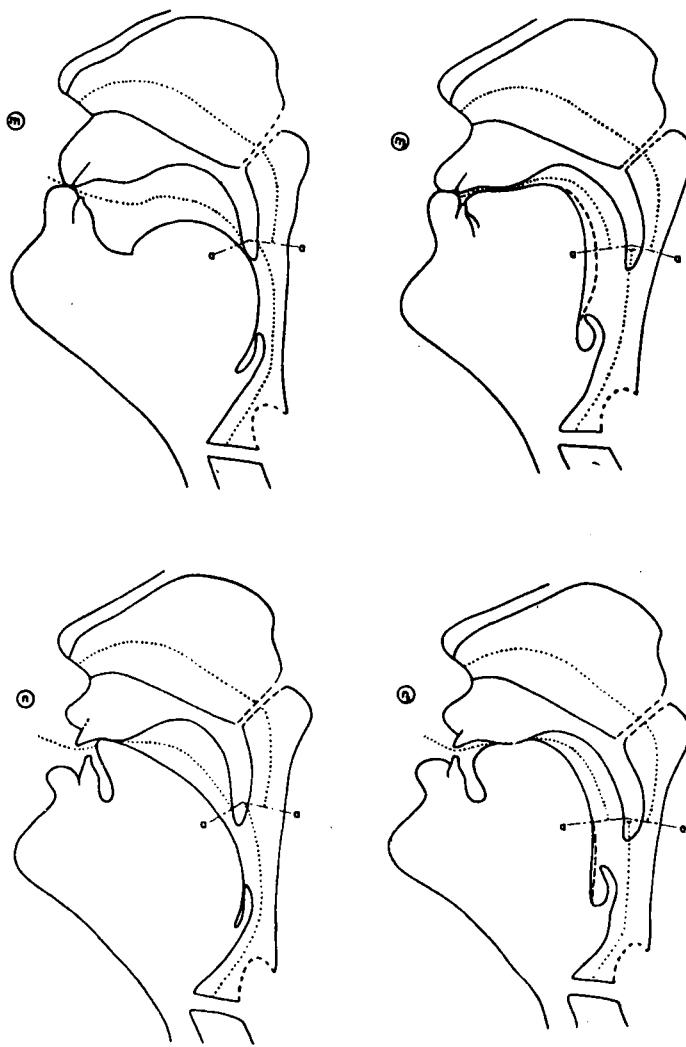


Fig. 2.4-1. X-ray tracings of nasal consonants; non-palatalized [m] and [n] and palatalized [m̪] and [n̪]. The lines marked $\alpha-\alpha$ in the region of the uvula indicate the estimated terminations of the mouth cavity, the nasal cavities, and the pharynx, i.e., the location of their interconnection in each instance.

is reflected in the variability of spectrographic details, when data from different subjects or from a single subject on different occasions are being compared.

Measurements of the dimensions of the nasal cavities involve fairly gross approximations. From the X-ray picture of our subject the overall length of the nasal pathways measured as the shortest distance from the uvula to the outlet at the nostrils was estimated to be 12.5 cm; see Fig. 2.4-1 showing the sagittal section through the head for the non-palatalized [m] and [n] and the palatalized [m̪] and [n̪]. The area

functions relating the cross-sectional area to the coordinate of the nasal pathways were constructed on the basis of measurements performed by G. Bjuggren¹ on a plastic mold of a complete nasal tract. These data were fitted to the appropriate length scale established for our subject.

The left and the right nasal channels run approximately parallel for a distance of about 8 cm from the nose-opening and combine in the naso-pharynx. Each of the frontal halves contains a bottom, middle, and upper branch in full communication at any cross-section. These appear to be too closely coupled to function as independent resonators. Provided the right and the left parts show complete symmetry, they will function acoustically as a single cavity system. This is the ideal configuration underlying the calculations, but it can be expected that asymmetry will cause an additional diffusion of spectral energy owing to the occurrence of formants from the left and the right pathways, and to the particular mixing in the nasal radiation. A greater damping of resonance in the nasal part than in the oral part of the vocal tract can be expected owing to the greater surface outline to area ratio of any cross-section except in the nasopharynx. The shape factor *Eq. A.36-10* was thus found to be of the order of 3. Nostril hairs will also contribute to the damping.

All calculations were performed with the aid of the electrical line analog *LEA*, described in *Chapter 2.2*. A number of unit filter sections, each simulating an acoustic section of effective length 1 cm and a variable area, were combined in cascade to form a nasal branch of the electrical analog. A terminating link simulating the effects of radiation and the mixing-in with the radiated oral output was included. The nasal tract was coupled to the oral system at a plane marked a-a in the X-ray tracings and area functions; *Fig. 2.4-2*. This arrangement is almost identical with that utilized by Stevens and House (1956) and House (1957) for experiments with the M.I.T. analog. Those investigators utilized a fairly large damping distributed in the form of resistance in series with the coils and the condensers of the nasal circuit resulting in a combination of constant and frequency-squared proportional losses. Our experiments were carried out with a nasal damping lumped into a single resistor of 1 μ c shunting the nasal tract at a point halfway along the length coordinate. This simple solution was chosen because of the empirical requirement of a larger damping of the second nasal resonance, i.e., the one at 1000 c/s, than of the first and the higher resonances. The standard voice source of -12 dB/octave spectrum envelope was adopted for the calculations.

The input impedance to the nasal network as a function of frequency, seen from the uvula, may be studied in *Fig. 2.4-3*, which shows the voltage at the input terminals of the nasal tract when connected through a 5 μ c resistor to the variable frequency sine wave source. These data are of particular interest for nasalization. Curve 1 pertains to a 2.6 cm² wide and 1.5 cm long coupling section at the velar-pharyngeal passage. The frequencies of zero impedance are found at zero frequency, 1000 c/s, 2800 c/s, and 4200 c/s. The maxima are intermediate. As discussed in *Section 1.33*, the expected

¹ Personal communication of data.

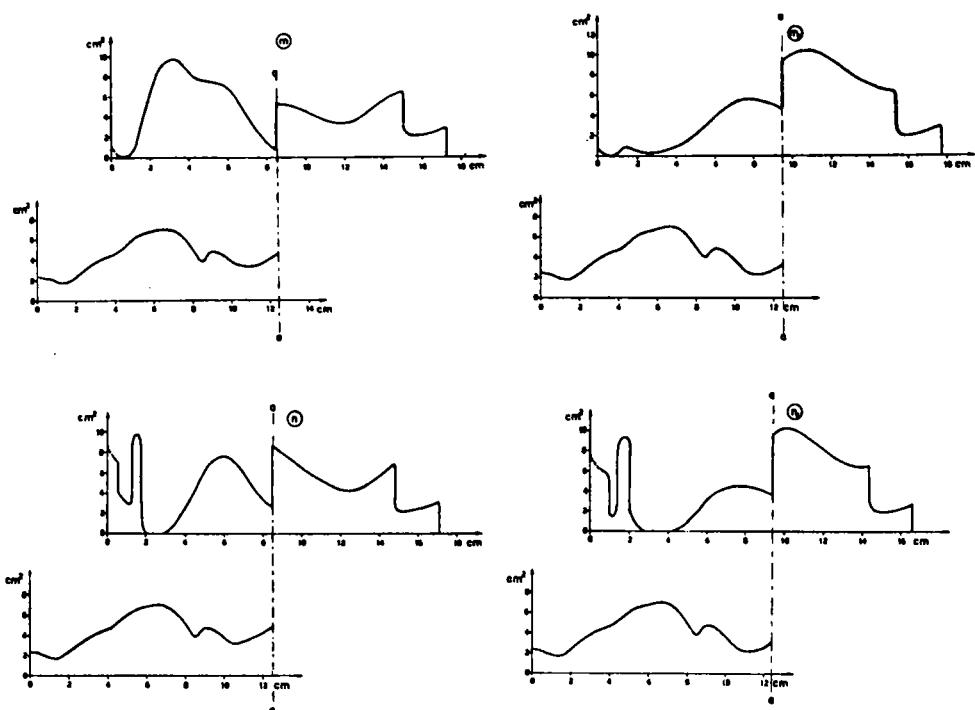


Fig. 2.4-2. Area functions of the nasal consonants [m], [m̄], [n], and [n̄]. The oral and nasal parts of the vocal tract are shown separately. The places of interconnection are indicated by the lines a-a.

average spacing between any two successive minima or maxima would be $c/2l_{tot} = 35300/25 = 1400$ c/s. Curve 2 refers to a decrease of the coupling area to 0.65 cm^2 which causes a decrease of the zero impedance frequencies by a few hundred cycles. Curve 3 shows the effect on the input impedance of reducing the coupling area to 0.16 cm^2 and narrowing the total area at the nose outlet from 2 cm^2 to 0.5 cm^2 . In curve 4 the effect of removing the nasal cavity damping resistor is shown to be a sharpening of the maximum at 300 c/s and the first minimum at 430 c/s.

2.42 Nasal Sounds Produced With Oral Closure

The nasal consonants are easier to discuss than the nasalized vowels since there is only the nasal output to be taken into account in the proper nasal murmur. Spectra of nasal consonants calculated on the basis of the configurative resonator data of Fig. 2.4-2 and the corresponding data from spectrographical analysis of the subject's speech may be studied in Fig. 2.4-4. Curve 1 shows the effect of removing the mouth cavity shunt of [n̄] entirely, so that the pharynx plus nasal tract constitute a single cascaded system without side branches as could be expected in the production of a velar nasal. The first formant is found at 300 c/s and represents the fundamental resonance of the pharynx cavity tuned by the nasal system as a resonator neck.

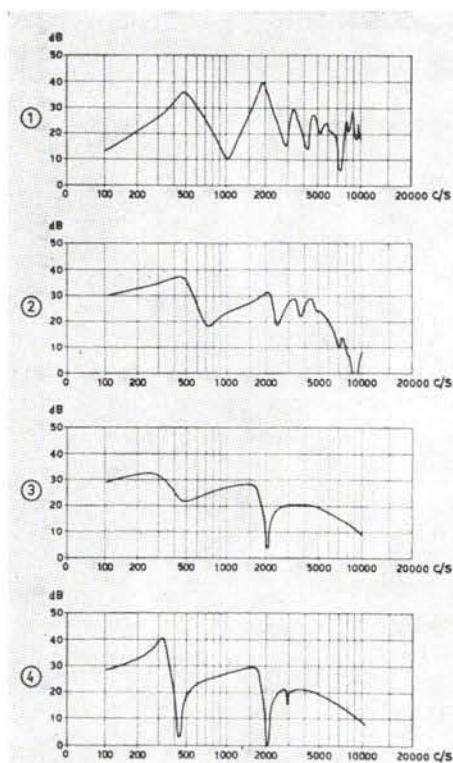


Fig. 2.4-3. Voltage level in dB at the input terminals of the nasal tract network when connected to a variable frequency source through a 5Ω resistor. These curves simulate the input impedance.

- (1) Cross-sectional area at the entrance to the nasal pharynx 2.6 cm^2 . The nasal tract area function is that of $[m]$; see Fig. 2.4-2. Nasal damping introduced by a 1Ω resistance shunting the nose cavity;
- (2) The input area reduced to 0.65 cm^2 ;
- (3) The input area reduced to 0.16 cm^2 and the area at the narrowest part of the nostrils reduced to 0.32 cm^2 over a length of 2 cm ;
- (4) Same as (3) with the nose cavity damping resistor removed.

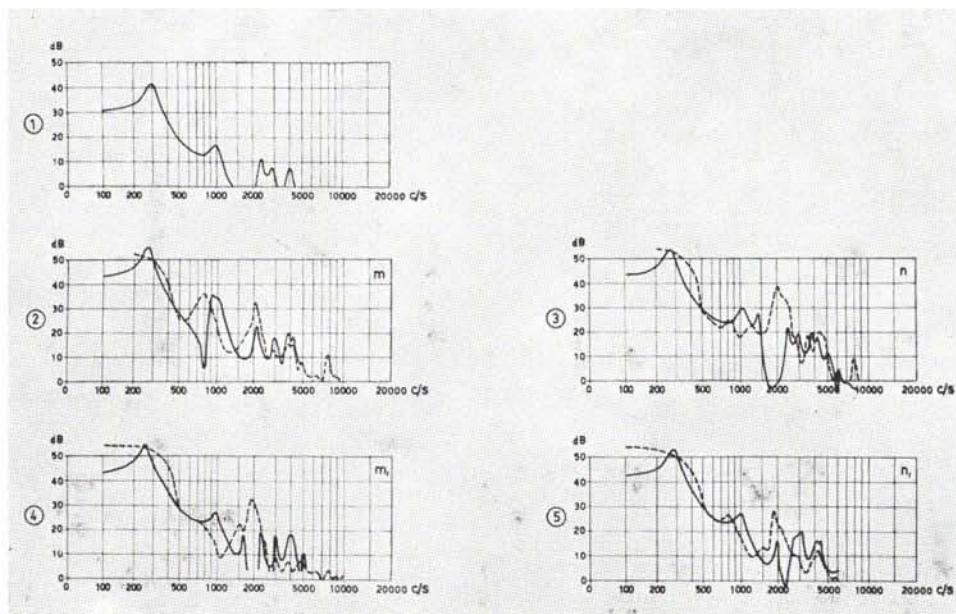


Fig. 2.4-4. Calculated (solid lines) and measured (broken lines) spectrum envelopes of nasal murmur.

(1) Mouth cavity shunt entirely removed from the [n̄]-model;
 (2)–(5) The sounds [m], [m̄], [n], [n̄], respectively.

The second formant of the hypothetical velar nasal occurs at 1000 c/s, which is close to the frequency of the first nasal tract input impedance minimum of *Fig. 2.4-3, curve 1*. The frequency of the third formant at 2200 c/s appears to be mainly related to a half-wavelength resonance of the pharynx cavity, and the fourth at 2900 c/s reflects the second impedance minimum of the nasal system above the uvula. The spectrum shape displayed by curve 1 is similar to that of a vowel [u+]. The larger damping of the second and higher formants marks the difference.

The effect of the mouth cavity as a side chamber, shunting the sound transmission through the pharynx-nose system, is to cause a shift in resonance frequencies and to introduce anti-resonances. The latter occur at the frequencies where the impedance of the mouth shunt is zero, thus trapping all the current delivered by the larynx source. Diagram 2 of *Fig. 2.4-4*, containing calculated and measured spectrum curves of [m], shows this effect clearly. The mouth cavity constitutes a Helmholtz resonator closed at the lip end and connected to the pharynx by the rather narrow passages on both sides of the uvula, the center of which is in contact with the back tongue. One obvious effect of the coupling to the mouth is the increase in total volume, tuning the fundamental resonance to a lower frequency, 250 c/s. The anti-resonance at 800 c/s is located 150 c/s below the second peak of the calculated spectrum. An analysis of the poles and zeros of the system reveals that this peak contains two poles, the 1000 c/s nasal cavity resonance already mentioned and a 900 c/s resonance due to the mouth cavity.

It has been verified by supplementary analog experiments that an increase of the coupling area to the mouth cavity, as in the case of an incomplete lowering of the velum or a lowered tongue position, shifting the mouth cavity anti-resonance up to 1000 c/s, causes a neutralization of the 1000 c/s nasal formant. The remaining peak at 1100 c/s will then be of a 15 dB lower level than the 1000 c/s double peak of diagram 1.

An approximate procedure for predicting the locations of poles and zeros in the frequency scale is as follows. First calculate the nodal point impedance of the naso-pharyngeal system at the uvula, i.e., at the place of coupling. This impedance is that of the nasal cavities above the uvula in parallel with the impedance of the pharyngeal cavity system below. The frequencies of maximum impedance of this network should be identical with the resonance frequencies of the compound naso-pharyngeal system without oral shunt, i.e., those of curve 1. Next, the input impedance to the mouth cavity shunt is plotted as a frequency curve, with reversed sign and neglecting resistance elements. The poles of the complete system are found at the intersection of this reactance curve with that of the naso-pharyngeal reactance curve, as seen from the coupling point and previously derived.²

² A simple procedure is in terms of admittances. With the notations $Y_n = 1/Z_n$ for the admittance of the nasal tract, Y_p for the pharynx, and Y_m for the mouth cavity, the condition for resonance of the complete coupled system is $Y_n + Y_p + Y_m = 0$. The uncoupled resonance frequencies of the pharyngeal system shall satisfy the relation $Y_n + Y_p = 0$.

Since the input impedance of a network composed entirely of reactances must be constantly rising with frequency except for the plus to minus infinity jumps at the maxima, and since further the reactance of the shunting system must start from minus infinity because of the closure at the lip end, it may be concluded that the frequency of zero reactance of the mouth shunt precedes the associated frequency of infinite reactance and that these frequencies come very close as the coupling area is decreased. Provided the rest of the vocal tract behaves like a capacitance at the frequency of the shunt zero, i.e., like a negative reactance, it can be anticipated that the pole frequency, i.e., the resonance frequency of the entire system, will be located between the zero and infinity points of the shunt and 100 c/s above the 800 c/s zero in this particular example. This graphical procedure is basically the same as that illustrated in *Fig. 1.4-1*; see also *Fig. 1.3-7* with regard to the contribution of a pole-zero pair to the spectrum envelope. The next pole-zero pair of the shunting system occurs at approximately 3500 c/s but is of less importance.

The significance of the [m]-spectrum calculation is supported by the close agreement with the experimental data. Because of the 150 c/s broad filter utilized for spectral sectioning, it is evident that both peaks and valleys will be broadened. From the broken line of curve 2, it may thus be concluded that the first zero-pole pair of the spoken [m] lies at frequencies which are 100-250 c/s below the calculated frequencies, the zero at 550 c/s, and the pole at 800 c/s.

These data also agree closely with those found in the extensive investigation of Ochiai, Fukumura and Nakatani (1957). In *Fig. 4* of their work it may be seen that the zero occurs at 550 c/s and the pole at 900 c/s. Our calculated [m]-curve is also in substantial agreement with the corresponding calculated data of House (1957), except for the smaller damping of the first and higher nasal formants in our curve. In actual speech the damping seems to be intermediate between that used in the American and that in the Swedish calculations. Our subject has a prominent 2000 c/s nasal formant typical for some speakers and probably due to a close proximity and thus mutual reinforcement of the first standing wave resonance in the pharynx and the first standing wave resonance of the nose, i.e., those formants found at 2200 c/s and 2800 c/s in diagram 1 of *Fig. 1.4-4*.

This 2000 c/s peak is the main cause of discrepancy between calculated and measured spectra of the other nasals. The fixed nasal cavity formant in the region of 1000 c/s recurs in the calculated spectra of the other nasals, diagrams 3-5. It is found at 800 c/s in the spectra of [n] and [n̄] but is less prominent in the [m]-spectrum. The mouth cavity resonance appears at 1400 c/s in the calculated curve of [n] but at 1200 c/s in the measured curve of the subject's speech. A 200 c/s difference is fairly constant. It should be observed that the calculated data represent sustained and stressed articulation whilst the measured data were taken from connected speech of consonant plus vowel syllables. Broad-band time-frequency-intensity spectrograms of this speech material can be found in the *Appendix*.

The main mouth cavity anti-resonance of the calculated [n]-spectrum is located at

1800 c/s, i.e., above the first resonance frequency of the non-nasal cavity system, whereas in the [m]-spectrum the anti-resonance was of a lower frequency than the mouth resonance. This difference is due to the smaller mouth cavity of [n].

The variable formant found at 900 c/s in the calculated [m]-spectrum and at 1400 c/s in the [n]-spectrum recurs at 1700 c/s in the calculated [m,]-spectrum and at 2000 c/s for [n,]. It is identical to the second formant of the nasalized vowel, which could have been produced by removing the oral closure. In the case of [m,] and [n,] it can no longer be referred to as a mouth formant because of the large coupling area at the velar entrance.

The F_2 -locations of the spoken [m,] and [n,] do not appear very clearly in the spectrograms and sections. From the continuity with the following vowel they are estimated to occur close to 2000 c/s for both sounds. The lower calculated F_2 of [m,] compared to [n,] is due to an apparent lowering of the subject's larynx as seen from the X-ray picture, Fig. 2.4-1. Typical of all palatalized sounds is the [i]-position of the tongue and thus the pharynx standing wave origin of the second formant, which is positioned at approximately 2000 c/s.

The frequency of the first zero in [m,] and [n,] is dependent on the mouth cavity configuration only, in accordance with general rules for shunting networks. The origin of this zero can be explained on a simplified basis by regarding the mouth cavity as a tube of effective length 4 cm closed at the mouth end. The frequency of zero impedance would be $c/4l = 35300/4 \cdot 4 = 2100$ c/s, which is close to the more carefully calculated values of 1800 c/s and 2200 c/s for [m,] and [n,], as may be seen in curves 4 and 5. This anti-resonance thus falls within the 2000 c/s peak region of the calculated [m,]- and [n,]-spectra and causes a selective level reduction. The second anti-resonance of the mouth shunt would occur at a frequency $3c/4l = 6300$ c/s referring to the idealized mouth tube of [m,] and [n,]. The observed values from the LEA calculations are 5600 c/s and 6400 c/s respectively. The second zero of [m] was accordingly found at 3500 c/s and for [n] the value 5600 c/s was obtained.

Summarizing, it may be said that nasal sounds contain fairly fixed formants essentially depending on the nasal tract and the pharynx. These occur at approximately 250 c/s, 1000 c/s, 2000 c/s, 3000 c/s, and 4000 c/s, and the lowest one has a dominating intensity level. The 1000 c/s formant is not always above the threshold of detectability in spectrograms owing to its low intensity in combination with the high frequency emphasis generally utilized for the processing. The formant at 2000 c/s and higher formants can be expected to undergo shifts only when the pharynx is contracted as in nasals coarticulated with back vowels. There are also formants that depend on the oral cavities, but they are severely weakened owing to the fairly close proximity of zeros. The spectrum of the nasal murmur thus contains a reduced second vowel formant F_2 which has a frequency place continuity with the F_2 of an adjacent sound if this is produced with unobstructed oral passage. The break of the oral closure causes a radical shift of zeros and thus a sudden increase of the intensity of the *oral* formants, but it may cause a small F_2 -variation only. Formants of higher

number within the F-pattern, e.g., F_3 , behave similarly.

The formants of the nasal murmur will be labeled N_1 , N_2 , N_3 , etc., in the order they occur in the spectrum. There is thus a correspondence of N_1 to F_1 , N_3 to F_2 , and in some instances N_4 to F_3 . This correspondence is a matter of continuity more than of identity since a formant may be equally dependent on the nasal and oral cavities, e.g., the ones labeled N_1 and F_1 .

As discussed by several authors (Cooper et al., 1952; Malécot, 1956), the formant transitions in the oral sound intervals next to the nasal murmur are the most prominent auditory cues for differentiating the various nasal phonemes. This fact conforms with the highly reduced levels of the F-formants within the nasal murmur. The audible residue of the F-pattern within the nasal is, however, not without importance, as can be demonstrated by singing a nasal murmur and shifting the tongue from an [i]- to an [u]-position (Smith, 1947).

2.43 Nasalization

The theory of nasalization has received a considerable attention in the last decade. Apart from the linguistic and phoniatric interests there are very good reasons for communication engineers specializing in analysis-synthesis-telephony to study nasalization. The distortion of the spectral characteristics of vowels in the form of extra peaks might cause errors in the labeling of formants according to their order numbers unless the formant tracker is designed to ignore the nasal superstructure. As discussed in *Section 1.31*, it is possible to make use of a knowledge of the redundancies within sound spectra for instructing the analyzer, mechanical or human, how to make correct identification.

The spectrographic attributes of nasalization are fairly well known. Joos (1948) mentions an extra formant above F_1 and specifically at 1000 c/s and a tendency of extra formants to occur between all regular vowel formants. He also mentions the presence of an anti-resonance at 900 c/s, which however is not consistent and is therefore regarded as insignificant.

Smith (1951) made the following statements concerning *open nasality*:

- 1) There is no significant change in the voice fundamental;
- 2) F_1 is mostly weakened;
- 3) A formant appears at 1000 c/s;
- 4) F_2 is often weakened and raised a little;
- 5) A formant at 2000 c/s occurs occasionally;
- 6) F_3 is weakened and lowered;
- 7) F_4 is intensified;
- 8) Resonances above F_4 tend to be weakened.

Delattre (1954) has made the following statements as a conclusion from extensive pattern playback control of spectrographic data:

- 1) F_1 is weakened (primary cue);
- 2) A nasal formant appears at 250 c/s (secondary cue);

- 3) There is, except for open vowels, a (non-essential) 2000 c/s formant;
- 4) F_2 is not influenced;
- 5) F_3 descends (non-essential cue);
- 6) F_4 descends on the frequency scale (non-essential cue).

House and Stevens (1956) synthesized nasalized vowels with the aid of an electrical vocal tract employing a technique similar to the one used for the purposes of the present work and supplemented with perceptual studies. The results of the experiments performed by House and Stevens are essentially as follows:

- 1) F_1 is weakened; it has an increased bandwidth and higher center frequency;
- 2) The overall level of the vowel is reduced;
- 3) There are various secondary effects, e.g., anti-resonance at 900-1800 c/s, elimination of F_3 , irregularities in the upper formants, and possibly additional spectral peaks.

Hattori, Yamamoto, and Fujimura (1956) came to the conclusion that the principal characteristics of nasalization are:

- 1) Enforcement of spectral intensity at about 250 c/s;
- 2) Selective reduction of spectral intensity at about 500 c/s (this anti-resonance is ascribed to the shunting effect of the nasal cavity system as a side channel to the vocal tract);
- 3) Weak and diffuse spectral components (ascribed by the authors to the sound emitted from the nostrils).

These observations were made on sonagram records and under various experimental conditions.

Nasalization is not an easy feature to study, if detailed data are required, since the acoustic characteristics vary both with speaker and with the particular sound upon which the nasalization is superimposed and with the type of and degree of nasal coupling. From the earlier investigations reviewed here it is, however, evident that the most consistently reported cue is the intensity reduction of the first formant. The anti-resonance found by some investigators and an increased formant bandwidth are analytically associated with formant intensity reduction. From a circuit theory point of view, any anti-resonance due to a shunting side branch is always paired with a resonance, and it is thus convenient to describe the spectral effect of nasalization in terms of the frequency characteristics of pole-zero pairs, as was discussed in *Section 1.33*; see *Fig. 1.3-7* and *Eq. 1.3-12*.

The pole and the zero of such a pair are well separated if the coupling to the shunt branch is considerable, i.e., if the coupling area is large. Conversely, if the coupling is small the zero will come close to the pole, and the spectral minimum normally associated with a zero might not develop to a visible extent. In the limiting case of complete closure the pole and zero will coincide, and there is complete and mutual cancellation of their effects. It is also quite possible for the first zero, owing to the shunt, to fall very close to the pole corresponding to the formant labeled F_1 .

in the spectrum of nasalized vowels, causing a maximum amount of reduction of the intensity level of the first formant.

These general statements and the transform equation developed in *Section 1.23* are applicable to either the oral output alone or to the nasal output alone. In the following more complete treatment of the theory of nasalization, the effect of mixing the two components will be considered. On the basis of *Eq. 1.23-1* to *1.23-16* the mixture of the oral and nasal speech outputs as picked up by a microphone placed at a relatively large distance from the speaker may be expressed by the following transform equation:

$$P_M(s) + P_N(s) = U_q(s)H(s)R(s)[k_{zM}(s)H_{zM}(s) + k_{zN}(s)H_{zN}(s)], \quad (2.4-1)$$

where $P_M(s)$ and $P_N(s)$ are the Laplace transforms of the oral and nasal components of the nasalized sound. Both components include the source factor $U_q(s)$, the pole factor $H(s)$, and the radiation characteristics $R(s)$. The common pole factor reflects the basic theorem that the natural modes of vibration are the same in any branch of a compound network. The zero function

$$H_{zM}(s) = \prod_{i=1}^{\infty} [1 + (s + \bar{\sigma}_{Mi})^2 / \bar{\omega}_{Mi}^2] \quad (2.4-2)$$

pertaining to the oral output and the zero function

$$H_{zN}(s) = \prod_{j=1}^{\infty} [1 + (s + \bar{\sigma}_{Nj})^2 / \bar{\omega}_{Nj}^2] \quad (2.4-3)$$

pertaining to the nasal output are assumed to contain conjugate complex zeros only. The frequency of the zero number i of the sound radiated from the mouth is thus $\frac{1}{2\pi}(\bar{\omega}_{Mi}^2 - \sigma_{Mi}^2)^{1/2}$. In spite of the relatively large nasal cavity damping, it will be permissible to specify the frequency and the bandwidth of the zero by the high Q approximation:

$$\begin{aligned} F_{zM_i} &= \bar{\omega}_{Mi}/2\pi; \\ B_{zM_i} &= \bar{\sigma}_{Mi}/\pi. \end{aligned} \quad (2.4-4)$$

The zeros of the mixed oral/nasal output are found by equating *Eq. 2.4-1* to zero:

$$k_{zM}(s)H_{zM}(s) + k_{zN}(s)H_{zN}(s) = 0. \quad (2.4-5)$$

The factors $k_{zM}(s)$ and $k_{zN}(s)$, as defined by *Eq. 1.23-5*, determine the proportions of oral and nasal sound transmission at low frequencies, i.e., well below the frequency of the first conjugate zero. Each of these contains a zero and a common pole on the negative real axis of the complex frequency plane which are dependent on the resistance and inductance elements of the two branches; see *Eq. 1.23-6*. The two zeros

add in the sound mixture to a new zero of a frequency equal to the common pole. This follows from the relation $k_{zN}(s) + k_{zM}(s) = 1$. It is thus permissible to substitute these weighting functions for the frequency independent shunt factors:

$$\begin{aligned} k_M &= L_N/(L_M + L_N), \text{ and} \\ k_N &= L_M/(L_M + L_N). \end{aligned} \quad (2.4-6)$$

Eq. 2.4-5 can be solved by the standard technique of substituting $s = \sigma + j\omega$ and equating the real and imaginary parts separately to zero. If only the first zero in each of the mouth and nose outputs is considered, it will be found that the sum contains only one zero of the following frequency and bandwidth

$$\left. \begin{aligned} F_{z1} &= F_{zM1} \cdot k_M^{-1} (1+a)^{-1} \\ B_{z1} &= B_{zM1} \left[\frac{1+aB_{zN1}/B_{zM1}}{1+a} \right]^{-1} \end{aligned} \right\} \quad (2.4-7)$$

where

$$a = k_N F_{zM1}^2 / F_{zN1}^2 \cdot k_M.$$

If the zero in the nasal output is of a higher frequency than the zero of the oral output, which generally is the case in nasalized sounds, it can apparently be expected that the zero of the combined output obtains an intermediate bandwidth and occurs at an intermediate frequency but closer to the position occupied by the mouth output zero, especially if the coupling to the nasal cavities is small. The weighting factors k_M and $k_N = 1 - k_M$ could also include additional factors to take care of possible differences in transmission loss from the mouth and the nose to a microphone. As the microphone is moved in the soundfield from the lips to the nose, it can thus be expected that the first zero will shift continuously from the frequency F_{zM1} to F_{zN1} . Since the coupling from the nose to the mouth-opening via radiation is not infinitely loose, the ideal end positions will not be reached exactly. Similarly, provided k_N is large owing to a constricted oral passage, there is a small base boost in the oral output owing to the greater R/L ratio in the nasal pathway. This is in effect the origin of the uncompensated pole of $k_{zM}(s)$ as previously discussed. The order of magnitude of this bass emphasis is estimated to be 1-2 dB at 100 c/s.

The effect of the second zero of each of the mouth and nose systems on the first zero of the mixed outputs can easily be taken into account in numerical calculations. The effect is generally small since the contribution from a zero to the spectrum at appreciably lower frequencies is small.

The procedure for an approximate prediction of the poles and zeros of the nasalized vowel, given the formant frequencies of the non-nasalized vowel and the poles and zeros of the nasal tract input impedance, is the same as the one earlier discussed for the sounds with nasal opening and oral closure; see *Section 2.41*. First the impedance of the whole oral system including the mouth and the pharynx as seen from the uvula is estimated, disregarding losses so that reactances are considered

only. The infinity points of this reactance curve coincide with the pole frequencies of the ideal denasalized system and thus correspond closely with its formant frequencies. The reactance must start from zero value at zero frequency because of the short-circuit, as seen in the direction towards the lips, and the curve passes zero value wherever the reactance in the direction of the mouth or the pharynx is zero. If the negative of the nasal tract input reactance curve is plotted on this oral curve, the pole frequencies of the combined oral and nasal system may be found from the intersection of the two curves. From a few graphical constructions of this type it is possible to state some general rules.

- 1) The zeros of the nose output F_{zNi} are exclusively those of zero mouth cavity impedance as seen from the uvula, and the zeros of the mouth output F_{zMi} are exclusively those of the nasal tract input impedance. The first zero of the nose output is generally located between F_2 and F_3 of the denasalized system but may come below F_2 in [u] and above F_3 in [i].
- 2) The coupling of the nasal cavities will introduce the same number of extra formants as the number of zeros in the nasal tract input impedance.
- 3) The frequency of the first pole F_{P1} , referring to the order relation only and not necessarily to F_1 , must always fall below the first zero and always between the frequency F_1 , as found without the nasal coupling and the frequency of the first impedance maximum of the nasal tract.
- 4) At small degrees of nasal coupling, the first zero of the mouth output, F_{zM1} , will be positioned just above the first pole of the compound system, F_{P1} , but as the coupling area increases, F_{zM1} will move up along the frequency scale at a faster rate than F_{P1} . This is even more true of the zero of the combined nose and mouth outputs, F_{z1} , owing to the larger proportion of sound radiated through the nose and the generally higher position of the first zero in the spectrum of the nose output. Provided the nasal tract is especially wide and free from constrictions, the first nasal zero may be found as high as 1800 c/s and even above F_2 . The opposite extreme is a complete blocking somewhere in the nasal passages, preferably at the outlet combined with a small coupling area at the velum, in which case the first zero will fall below 500 c/s.
- 5) As the outlet of the nasal tract is narrowed, a decrease in frequency of all poles and zeros of the nasal tract input impedance can be expected. However, if the coupling area at the uvula is large, the decrease of the zeros will be small. A narrowing of the nasal passages halfway between the uvula and the nostrils will cause a substantial shift down in frequency of the second nasal impedance pole and zero in addition to a moderate shift down of the first pair; compare the three-parameter model diagrams in *Section 1.43*.

Next, the production of a few nasalized vowels will be discussed on the basis of this theory and spectrum curves derived from synthesis with the electrical analog *LEA*. The spectral data demonstrated in *Fig. 2.4-5* pertain to different degrees and types of nasalization superimposed on the articulatory configuration of the vowel [a] as specified by the area function tabulated in *Section 2.32*. Curve 1 shows the spectrum

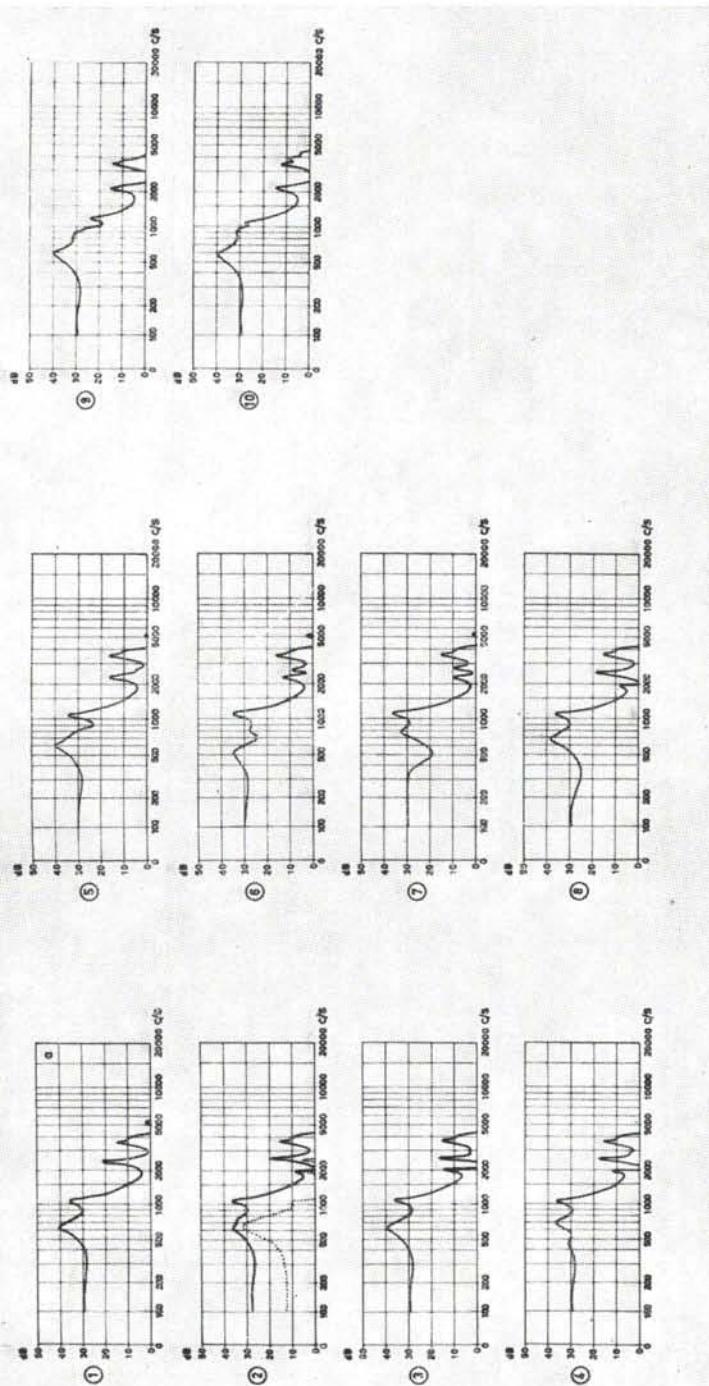


Fig. 2.4-5. The effect of nasal coupling on the calculated spectrum envelope of the vowel [a]. Glottis resistance $\zeta = 0.5$.

- (1) No nasal coupling;
- (2) Mouth output (solid curve), nasal output (dotted curve), assuming a coupling area of 0.16 cm^2 at the naso-pharynx;
- (3) The sum of the mouth output and the nasal output of (2);
- (4) Same as (3) but the area at the narrowest region of the nasal tract outlet reduced to 0.32 cm^2 ;
- (5) Same as (3) but for increase of the naso-pharyngeal coupling area to 0.65 cm^2 ;
- (6) Same as (5) but for a reduction of the nasal tract outlet as in (4);
- (7) Same as (6) except for a complete obstruction at the nostrils;
- (8) Same as (7) but the naso-pharyngeal area reduced to 0.16 cm^2 ;
- (9) Normal nasal tract outlet area. Naso pharyngeal coupling area 2.6 cm^2 ;
- (10) Same as (9) except for a reduction of the area at the entrance to the mouth cavity to 0.65 cm^2 caused by an approach of the velum toward the back of tongue.

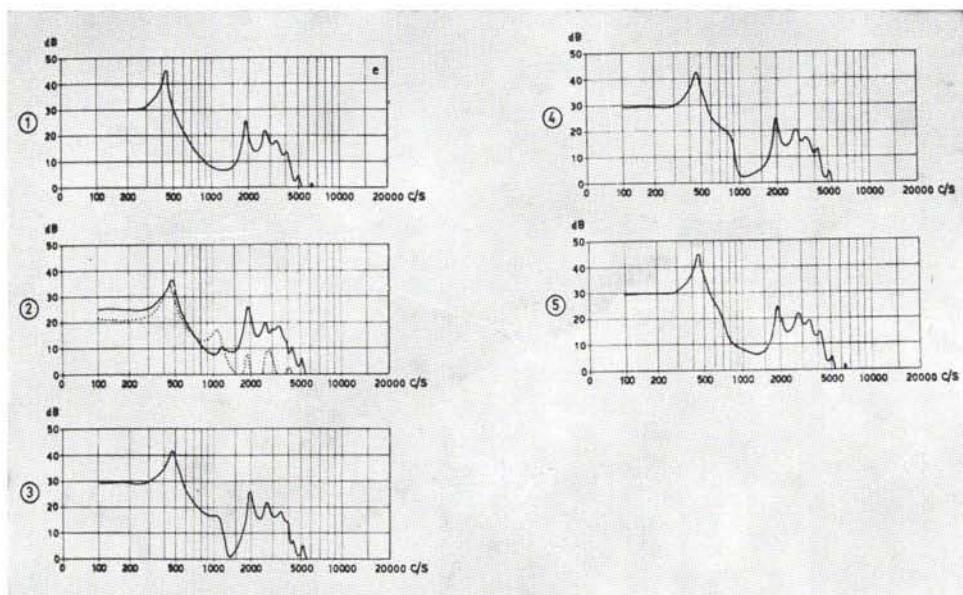


Fig. 2.4-6. The effect of nasal coupling on the spectrum envelope of the vowel [e]:

- (1) No nasal coupling;
- (2) 2.6 cm² naso-pharyngeal coupling area. The solid line represents the mouth output and the dotted line the nose output;
- (3) The sum of the nasal and the oral outputs of (2);
- (4) Same as (3), except for a naso-pharyngeal coupling area of 0.65 cm²;
- (5) The naso-pharyngeal coupling area reduced to 0.16 cm².

without coupling to the nasal tract. The first four formants occur at 630 c/s, 1070 c/s, 2400 c/s, and 3550 c/s respectively. Curve 2 shows the effect of a very small nasal coupling, 0.16 cm^2 at the naso-pharyngeal passage, on the spectra of the mouth and nose outputs, and their sum is represented by curve 3. Here the most obvious effect is the split and level reduction of the third formant, or more precisely, the occurrence of a nasal formant at 2000 c/s and an anti-resonance between this formant and F_3 . This distortion originates from the second pole-zero pair of the input impedance to the nasal tract. The effect of the first pair is somewhat obscured owing to its place between F_1 and F_2 and owing to the relatively large damping of the glottis resistance, providing higher $B1$ and $B2$ than in normal speech; see Section 2.33.

The frequency of the first nasal formant is $F_{p2} = 660 \text{ c/s}$, and the first zero of the mouth output occurs at $F_{zM1} = 700 \text{ c/s}$. The first zero of the nasal output, F_{zN1} , is located at 1200 c/s, i.e., just above $F_2 = F_{p3}$. The nasal output is thus dominated by a double peak containing F_{p1} and F_{p2} , but $F2$ is almost eliminated by F_{zN1} .

The difference between the mouth and the nose output curves at low frequencies reflects the degree of nasal coupling as specified by the inductance ratio $L_N/L_M = 6.4$, from which the shunt factor $k_M = 0.87$ is calculated, and further, a 5 per cent increase of frequency of the zero F_{z1} in the sound mixture as compared to F_{zM1} . Neither F_1 nor F_2 are influenced much by the addition of the nasal cavities at this low degree of coupling. It is of interest to see that the apparent reduction of the level of the first formant peak, as seen from the mouth output curve is counteracted by the addition of the nasal output so that the net reduction of level is only 1 dB. The anti-resonance effect of the first zero F_{z1} is reduced to a 1.5 dB decrease of the level of the valley between $F1$ and $F2$.

Curve 4 of Fig. 2.4-5 illustrates the effect of reducing the cross-sectional area of the nasal passages at a region 1-4 cm from the radiating end of the nostrils whilst in other respects the conditions of curve 3 are retained. The nasal formant F_{p1} is now visible and appears below F_1 at 450 c/s. There is also a 3.5 dB reduction of the level of the first formant as compared with the unnasalized state in curve 1. The split third formant is retained, but its lower part F_{p3} is weaker than in curve 3.

Curves 5, 6, and 7 all pertain to a coupling cross-sectional area of 0.65 cm^2 at the naso-pharyngeal passage, curve 6 with the same narrow nasal outlet as for curve 4. The first zero F_{z1} is shifted to 700 c/s in curve 6, and to 900 c/s in curve 5, representing the wider nostrils. The complete closure of the nostrils brings down the zero to 500 c/s and the first nasal pole F_{p1} to a position of about 300 c/s. There is a rise of F_1 to 780 c/s, and $F4$ dominates over $F3$ and $Fp2$.

The effect of decreasing the nasal coupling to 0.16 cm^2 retaining the nostril closure can be studied from curve 8. The first zero has moved down to 350 c/s but is less apparent, and the decrease of the level of the first formant is 1.5 dB only, compared with 8 dB for the 0.65 cm^2 coupling area of curve 7. Apart from the weak nasal formant, always found below F_3 if there is some degree of nasal coupling, the spectral distortion is small.

Curve 9 refers to the extreme case of a very open nasal tract, both with regard to the area at the outlet and the coupling area. The first zero of the combined output is found just below F_2 , resulting in a very low level of the second formant. This effect is even more pronounced in curve 10, which was produced after an additional decrease of the cross-sectional area at the mouth cavity inlet, as would be the case when the velum drops very low and approaches the tongue.

The effect of nasalization on the vowel [e] is illustrated in Fig. 2.4-6. Only the wider alternative of the front part of the nasal passages was utilized for the spectral calculations. Curve 1 represents the unnasalized conditions with $F_1 = 420 \text{ c/s}$, $F_2 = 1960 \text{ c/s}$, $F_3 = 2750 \text{ c/s}$, and $F_4 = 3410 \text{ c/s}$. Diagram 2 contains the separate spectral curves for the nose and the mouth outputs under the condition of a velar-pharyngeal coupling area of 2.6 cm^2 , and the sum of the two components is given in curve 3. There is a shift up in frequency of the first formant to 490 c/s and a broadening of its bandwidth. The peak level is reduced by 7 dB in the oral output but by only 4 dB in the combined output. No striking changes are observed in the second and higher formants, but there is a clear additional formant at 1100 c/s identified as F_{p2} . It occurs just above the first zero of the mouth output. The latter is shifted up to 1350 c/s in the combined output.

A numerical control of the shift of this nasal zero was performed and includes the following operations: First the shunt factor $k_M = L_N/(L_M + L_N) = 0.62$ is determined from the ratio of mouth output to nasal output at low frequencies which is 1.8 or 5 dB , as seen in diagram 2. This ratio was also checked by an integration of the mouth and nose area functions according to Eq. 2.3-1. Eq. 2.4-5 is next solved, disregarding the bandwidths of the zeros and noting that the first zero of the mouth output is located at 1050 c/s and the second at 2900 c/s and that the first zero of the nasal output has the frequency 2650 c/s . The result is a first zero of 1350 c/s in the combined output which checks with the analog computations. The effect of reducing the coupling area to 0.65 cm^2 is to lower the frequency of the first zero to 1000 c/s , as seen in curve 4, and a further narrowing of the coupling area to 0.16 cm^2 shifts the zero to 800 c/s , as seen from curve 5. The associated nasal formant is just below this zero and is merely seen as a change of slope in the spectrum envelope above F_1 . This effect is often seen in spectrographic records.

Fig. 2.4-7 contains spectra of nasalized [i] and [u]. Diagrams 2 and 3 refer to the larger coupling area of 2.6 cm^2 . It can be seen that the nasal output is of 8.5 dB higher level than the mouth output. The general effects are the same as discussed for the vowel [e] except that the first zero has reached a position of 1800 c/s causing an apparent weakening of the second formant and all higher formants. Since the nasal output component predominates, it is of some interest to study the difference between this spectrum and one obtained with complete closure at the palatal constriction. As seen from a comparison of curves 3 and 4, the effect is small, which implies that curve 3 represents a high degree of nasalization. Because of the narrow tongue passage of the vowel [i], and thus the large mouth-passage inductance L_M , it is obvious that the nasal

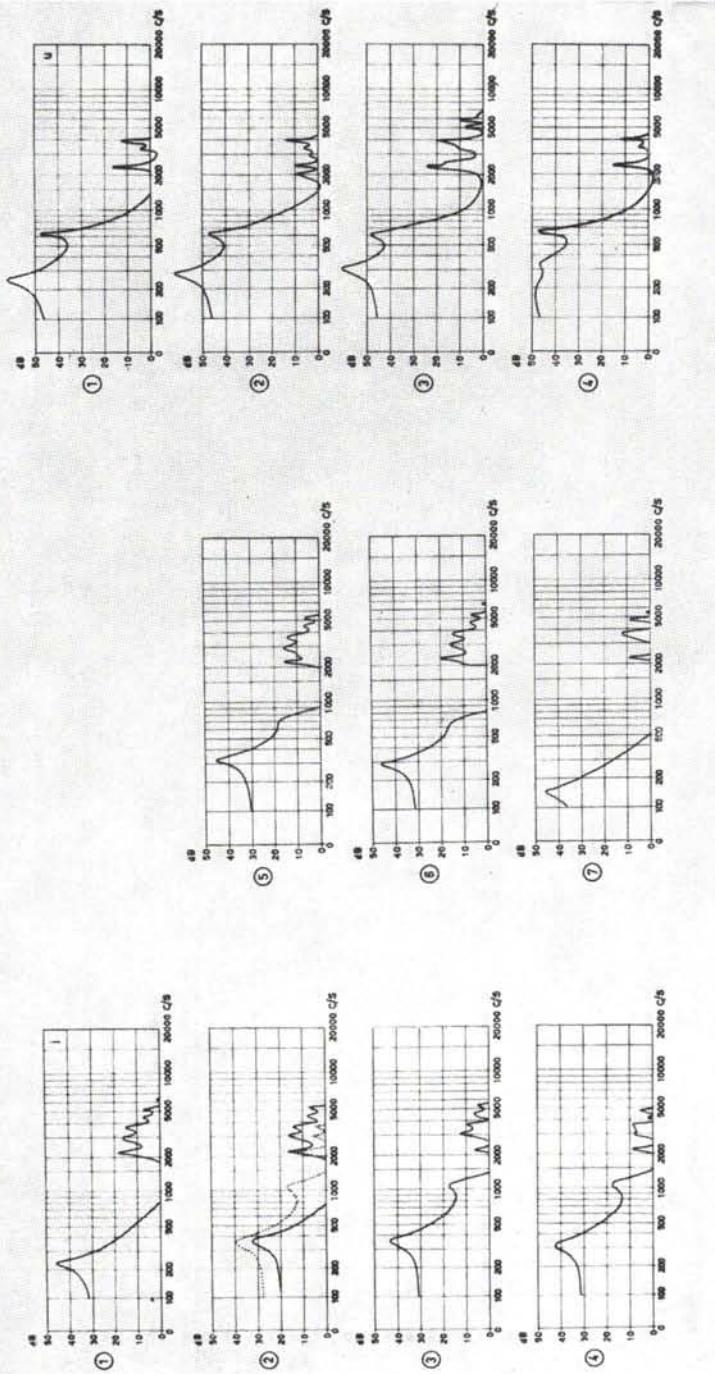


Fig. 2.4.7. The effects of nasal coupling on the spectrum envelopes of [i] (1) - (7) and [u] (1) - (4) to the right of the figure.

- (1) The vowel [i], no nasal coupling;
- (2) 2.6 cm² nasal pharynx coupling area (the solid line represents the mouth output and the dotted line the nose output);
- (3) The sum of the nasal and the oral outputs of (2);
- (4) The palatal tongue passage narrowed to one-quarter of its normal value and shut off completely at its midpoint;
- (5) Same as (3) except for a 0.32 cm² coupling area;
- (6) Same as (3) except for a 0.16 cm² coupling area;
- (7) Same as (1) except for a reduction of the cross-sectional area at the palatal passage to one-quarter of its normal value.
 - (1) The vowel [u], no nasal coupling;
 - (2) The sum of the nose and mouth outputs assuming a coupling area of 0.16 cm² at the nasal pharynx;
 - (3) Same as (2) except for a coupling area of 0.65 cm²;
 - (4) Coupling area 0.32 cm² (the outlet at the nostrils is completely blocked).

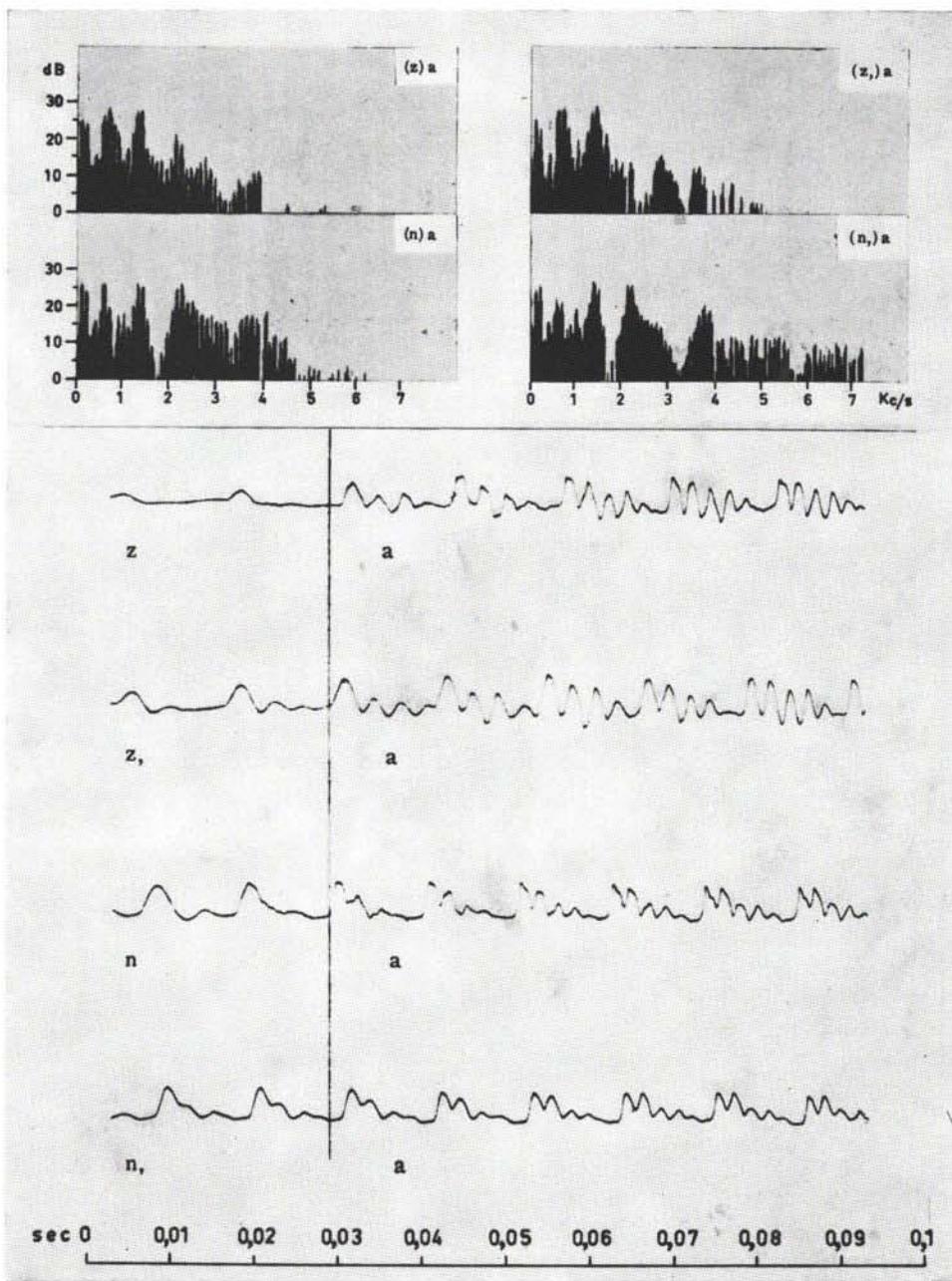


Fig. 2.4-8. Sonagraph sections and Mingograph oscillograms of [za], [z,a], [na], and [n,a]. The section samples originate from the third to the fifth period after the break of the oral closure. The oscillograms are low-pass filtered at a cutoff frequency of 1000 c/s. The suppressed F_1 of the oscillogram is typical of the waveform of a nasalized vowel.

sound transmission can be appreciable even at low degrees of nasal coupling.

Curve 5 refers to a coupling area of 0.32 cm^2 , and curve 6 to a coupling area of 0.16 cm^2 . Even in the latter case, the frequency of the first formant is high enough to cause a shift in sound quality from [i] to [ɪ]. According to the general rules relating formant frequencies to formant levels and to the level of any point on a spectrum envelope, as discussed in *Section 1.32*, there should follow a rise in spectrum level of the higher formants by the amount of $40 \log_{10}(315/220) = 6 \text{ dB}$ when F_1 is shifted from 220 to 315 c/s . The reason why this level increase does not appear when, for instance, curves 1 and 5 of *Fig. 2.4-7* are compared, is the existence of a compensatory spectral term largely due to the first nasal pole and zero providing a high frequency constant level loss of the type shown in *Fig. 1.3-7*. Since the extent of this level correction is $40 \log_{10}(F_{21}/F_{P1})$, it may be concluded that the ratio of the nasal zero frequency to the frequency of the nasal pole is approximately equal to the ratio of F_1 after the nasal coupling to F_1 of the non-nasalized state.

The diagrams 1 to 4 to the right in *Fig. 2.4-7* illustrate the nasalization of the vowel [u]. Curve 2 indicates that even a very small nasal coupling can provide a noticeable damping of all formants except the first. When the nasal coupling is 0.65 cm^2 , there is a shift up of F_1 from 240 c/s in the unnasalized state to 300 c/s . The particular combination of poles and zeros in the region of F_3 to F_5 is by chance such that F_3 and F_5 are emphasized. Finally, it is shown in curve 4 that a complete closure at the nose outlet causes an almost total leveling of the first formant owing to the shift down of the first zero from the position 900 c/s in curve 3 to a position close to F_1 .

After these observations on the behavior of the ideal but lifeless "subject" LEA, it might be worth-while to see what nasalization is like in human speech. *Fig. 2.4-8* contains spectrographic sections and oscillograms of the vowel [a] sampled in the stationary part of the vowel following an initial consonant [n], [n̄], [z], and [z̄], as shown by the spectrograms in the *Appendix*. It was noticed in intensity records of these syllables that the overall intensity level of vowels following a nasal consonant was on the average 2 dB lower than that of the same vowel following other consonants. An inspection of the sections of *Fig. 2.4-8* reveals a fixed frequency anti-resonance at 300 c/s . This is an inherent quality feature of the subject's voice and is probably not due to nasality. Another individual feature of the spectrographic records from this speaker is that the third formant is often weak but considerably emphasized in vowels carrying nasality by assimilation.

The effect of the nasal contact on the frequencies of the first two formants is rather small. Within both pairs, F_1 is lowered about 50 c/s in the vowel following the nasal consonant. From [(z)a] to [(n)a] there is a decrease of the level of the first formant L_1 by 2 dB , and of the second formant L_2 by 1 dB , while the third formant level, L_3 , is increased by 5 dB . Comparing [(n,a)] with [(z,a)], there is a 5 dB decrease of L_1 and 1 dB in L_2 , but L_3 increases as much as 11 dB . The second zero at 1800 c/s is very pronounced in both [(n,a)] and [(n̄,a)]. In the former vowel the first zero is found at

800 c/s, and in the latter it is probably 50 c/s closer to F_1 owing to the higher reduction caused in L_1 . The associated nasal formant is seen at 1000 c/s.

This configuration of poles and zeros between F_1 and F_2 resembles that found in curve 6 of Fig. 2.4-5. This is also applicable to the spectrogram of Fig. 1 of the article by Delattre (1954) which shows a very clear anti-resonance at 900 c/s. However, to the summary given by Delattre it could be added that the weakening of the first formant is largely due to this zero and that the baseline formant at approximately 250 c/s, referred to by Delattre as FN_1 , does not differ very much in the nasalized and unnasalized portion of his spectrogram nor in our Fig. 2.4-8. The automatic volume control of the Sonagraph raises the reproduction gain as the first formant is weakened which raises the marking level of the baseline formant.

As shown in the discussion of Fig. 2.4-5, the nasal passages must be very narrow for a first nasal formant at 250 c/s to superimpose itself on an [a]-spectrum. Such conditions were apparently present in the production of the nasalized [a] of Fig. 1 of the article by Hattori et al. (1956). The spectrum of a nasalized [a] with the nostrils closed demonstrated by these Japanese investigators, shows essentially the same effect in the frequency region up to the second formant, as demonstrated by our synthesized spectrum, curve 7 of Fig. 2.4-5.

The third nasalization cue specified by Delattre and earlier mentioned by Smith (1951) in the form of a 2000 c/s formant was often seen in most of the synthetic spectra of Fig. 2.4-5 as a *split third formant*. The two formants within this group are dependent on the second resonance of the nasal tract and on either of the two possibilities of a half-wavelength resonance of the mouth cavity or of the total oral plus nasal pharynx. If the velum has dropped very low, the total pharynx length may be the main determinant of the lower of the two formants, but if the nasal coupling is very small, this formant may be mainly dependent on the nasal cavities. From the synthesis experiments it is apparent that the effect of nasal coupling on the topology of peaks and valleys in the region of F_3-F_5 varies much and according to rather intricate rules since the effects of varying degrees of coupling depend on the particular nasal tract configuration.

It has become the normal technique in speech analysis to illustrate acoustic sound quality attributes by spectral curves. However, oscillograms produced at a high paper speed may sometimes provide a comparable or even a clearer insight into specific details of the signal structure. The oscillograms of Fig. 2.4-8 were produced with a Mingograph direct-writing ink-recorder covering the frequency range of 0-800 c/s. They thus show the waveform of F_1 plus extra formants and the residue of the vocal cord periodic air pulsations. The most apparent contrast is that between [n,a] and [z,a]. The oscillation constituting the formant waveform is more heavily damped in the [a] following the [n,] than in the [a] following the [z,]. If the formant ripple is extracted, there remains the residue from the voice source pulsations which is most clearly seen in the [n,a] but also apparent in the [na]. The reduction of the first formant intensity by means of an anti-resonance is a process identical in principle with that

utilized for extracting the waveform of the voice source by means of inverse filtering discussed in *Section A.21*.

The conclusions reached by Delattre (1954) and House and Stevens (1956), that the intensity reduction of the first formant is the major perceptive cue for nasalization, have been verified by experiments with the formant-synthesis speaking machine *OVE*. A reduction of L_1 by means of an increase in B_1 creates a very apparent nasal quality. Experiments with the line analog *LEA* show that the addition of the nasal cavities also causes a quality change of the type expected as the result of a shift up in frequency of F_1 . This is due to the tendency of an actual rise in F_1 or rather to a rise of the mean position of F_1 and the first nasal resonance. Because of the rather open nasal passages and nasal outlet assumed for our subject, both the first nasal zero and the associated nasal formant were generally found above F_1 . One theoretical implication of this finding is that nasalization may compensate for too neutral a position of the tongue when the vowel [a] is to be articulated; that is, a widening of the pharynx or an increase of the volume behind the point of minimum cross-sectional area may be compensated for by nasalization. Without this compensation the vowel quality comes too close to the schwa. Systematic reconstructions of speech spectra from X-ray data and *LEA* synthesis support in part this observation.³ The calculated F_1 of back vowels always comes out too low; compare the [a] of Fig. 2.3-3. On the other hand, the opposite effect is theoretically possible if the nasal passages are narrow. A small positive shift in F_1 due to the pharynx wall vibrations, should also be taken into account as discussed in *Section 2.33*, but this effect appears to be of the order of only 3 per cent at $F_1 = 600$ c/s. A similar shunting effect of the glottis inductance is theoretically possible but cannot be accurately estimated owing to the lack of physiological data and flow data.

The subjective evaluation of the degree of nasalization is influenced by the amount of pitch variation given to the synthesized vowels. As found by van den Berg and Fant in informal listening tests, the effect of an added vibrato is to neutralize the nasal quality of a vowel synthesized with the electrical line analog *LEA* arranged to incorporate the nasal tract.

* Unpublished data.

2.5 THE LIQUIDS

The acoustic theory of the production of the liquids [l] and [r] is simple, provided turbulent noise generated at narrow passages can be disregarded. This is normally the case as judged from spectrographic data and synthesis experiments. Voiced [r] can be treated by the same acoustic theory as vowels since there are no shunting side chambers introducing anti-resonances. In the study of laterals there enters the complication of a zero function, similar, but not of the same importance as the zero function of nasal sounds.

The X-ray tracings and the area functions of non-palatalized [l] and [r] and the palatalized [l_p] and [r_p] can be studied from *Fig. 2.5-1*. The point of articulation is more advanced for [l] than for [r] and more advanced for [l_p] than for [r_p], these differences being mainly attributable to the dental versus the alveolar point of apical contact. The typical articulatory difference between the palatalized and the non-palatalized tongue configuration is apparent when [r_p] is compared with [r] and [l_p] with [l]. It can be described simply by a reference to the relation of [l] to [u]. In spite of the fairly fixed position of the apex, the back of the tongue is free to approach the upper part of the pharynx in [l] and [r], thus dividing the cavities behind the point of articulation into a configuration resembling a double Helmholtz resonator. The non-palatalized sounds thus have a secondary place of articulation at or slightly below the uvula. The palatalized tongue position, on the other hand, determines a back cavity configuration of a wide unobstructed pharynx and a gradual narrowing of the mouth cavity towards the region of articulatory constriction. This obstruction is complete in the laterals but periodically variable with the trill frequency of [r] and [r_p] as produced by our subject.

The lateral outlet combining the interior cavities of [l] and [l_p] with the cavities in front of the tongue is estimated to extend from a position close to the posterior molar teeth. Evidently, the cross-sectional areas specified for this connection, which may be unilateral as well as bilateral, involve a good deal of guesswork. However, the length dimensions are of primary significance for the calculations.

The results of the spectrum calculations may be studied in *Fig. 2.5-2* and *Fig. 2.5-3*,

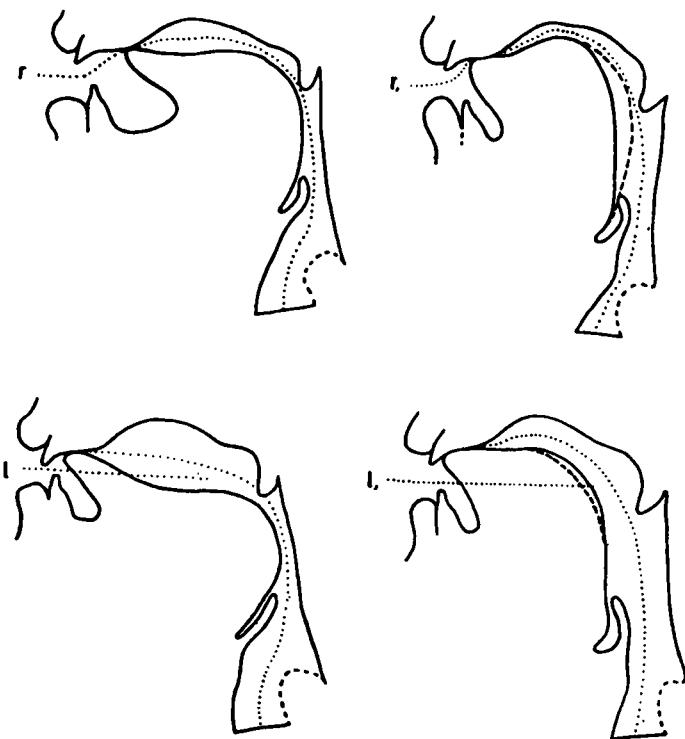


Fig 2.5-1. a) X-ray traces.

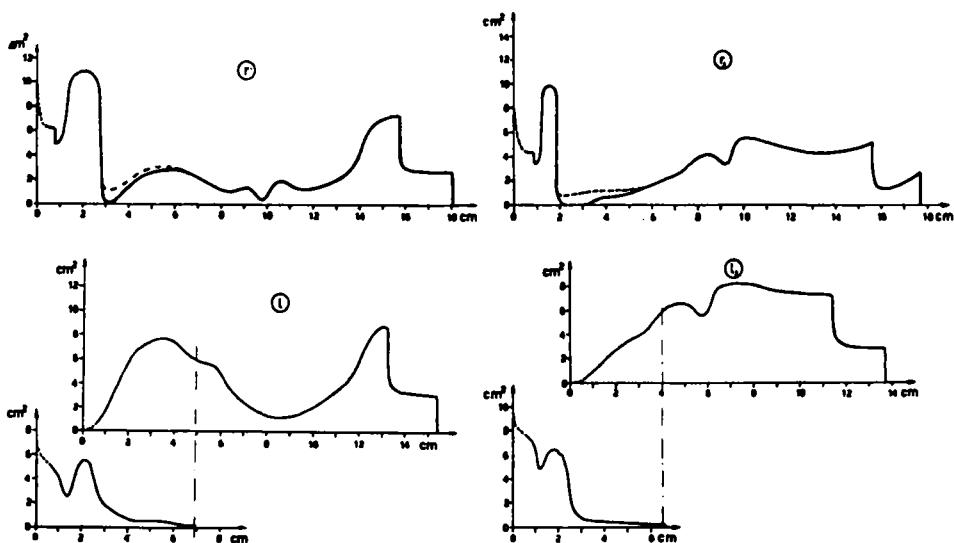


Fig. 2.5-1. b) Area functions of the non-palatalized [l] and [r] and the palatalized [i,] and [ɛ].

where solid lines and dotted lines pertain to spectra calculated with the aid of the line analog *LEA*. Broken lines refer to measured samples from the mono-syllabic test words uttered by the subject and submitted to a detailed-time-frequency-intensity analysis, as shown in the *Appendix*, *Fig. A.13-3* and *A.13-4*.

Curve 1 of *Fig. 2.5-2* provides a comparison of spoken data of [l] with the data calculated from the area functions of *Fig. 2.5-1*. There is a fair amount of agreement, for instance in the low position of F_2 , and thus the rather great separation of F_2 versus higher formants, as could be expected from the secondary articulation characteristics in the form of the divided cavity system behind the primary point of articulation. Judging by the higher $F_1 = 350 \text{ c/s}$ of the spoken sample as compared with $F_1 = 220 \text{ c/s}$ of the calculated curve, the width of the lateral passage was underestimated. A second calculation was undertaken with an increase of the minimum cross-sectional area of the lateral passage from 0.3 cm^2 to 0.65 cm^2 , and a narrowing of the pharyngeal pass from 1.3 cm^2 to 0.65 cm^2 . The effect of the latter change alone is to shift down F_2 from 850 c/s to 620 c/s , which is lower than for the spoken sample.

The effect of the increase of the lateral outlet was a shift up of F_1 from 220 c/s to 290 c/s . The pharyngeal narrowing did not appreciably influence the spectrum above F_2 . The topology of this higher part of the spectrum is essentially the same in the calculated and spoken sample except for about 5 to 10 per cent higher frequency values in the calculated data. The spoken sample thus shows an anti-resonance at 1800 c/s followed by a very weak formant at 2000 c/s . The next peak at 2700 c/s constitutes the fourth formant, and there are three more peaks below 5000 c/s . The corresponding calculated data include an anti-resonance at 2100 c/s and a plateau at 2300 c/s indicating the third formant. The fourth formant is found at 2900 c/s and corresponds to the 2700 c/s peak of the spoken sample, and there are three additional peaks below 5000 c/s to be seen.

This agreement, especially with regard to the anti-resonance of the spoken sample, provides a support for the estimate of the length of the lateral passages and specifically of the point where they start in the posterior part of the mouth. The anti-resonance is evidently due to the shunting effect of the mouth cavity behind the tongue blade which can approximately be described as a tube closed at its far end. The frequency of anti-resonance, i.e., the zero frequency, is evidently $c/4l_s$, where l_s is the length of the tube and c the velocity of sound. A zero frequency of 2000 c/s thus corresponds to an effective length of the shunting system of $l_s = 4.4 \text{ cm}$.

The calculated effect of breaking the apical barrier and closing off the lateral passages is illustrated in diagram 3 of *Fig. 2.5-2* which contains the calculated curve from diagram 2 as the basis of comparison. As seen from the dotted curve, the anti-resonance has disappeared and the weak third formant at 2250 c/s of the true lateral sound has shifted up to a position close to the fourth formant at 3000 c/s . The fifth formant remains unchanged, and the sixth and seventh formants have been shifted up 10 per cent in frequency and are weakened. The third formant is dependent on the anterior mouth cavity, including the lateral passages before the break. The fourth

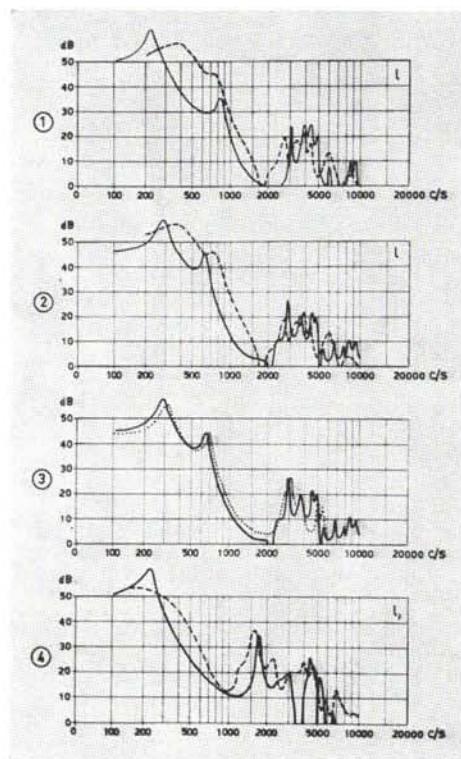


Fig. 2.5-2.

- (1) Spectrum of non-palatalized [l] calculated from the area function of Fig. 2.5-1 (solid line curve) and the corresponding data measured from connected speech (broken line curve);
- (2) Same as (1) except that the calculations have been based on an increase of the minimum cross-sectional area within the lateral passage by a factor of 2 and a decrease of the back tongue-pharyngeal pass by the same factor;
- (3) Solid line curve as that of (2). The dotted line curve shows the effect of breaking the apical obstruction and closing the lateral shunt;
- (4) Calculated and measured spectrum of palatalized [l].

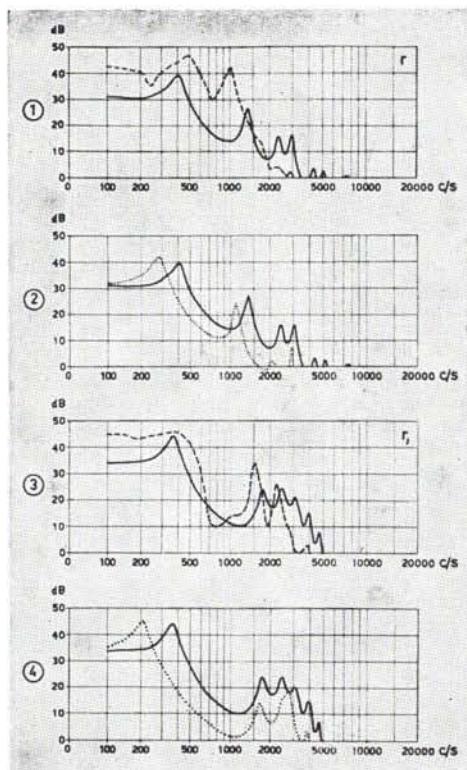


Fig. 2.5-3.

- (1) Spectrum of the open phase of a non-palatalized [r] calculated from the area function of Fig. 2.5-1 (solid line) and the corresponding measured data (broken line curve);
- (2) Calculated spectrum of the open phase of [r] as in (1) and a relatively closed phase (dotted line);
- (3) Calculated and measured spectrum of palatalized [r;];
- (4) Calculated open and closed phase of [r;].

formant at 2900 c/s and 3000 c/s respectively is mainly dependent on a half-wavelength resonance in the mouth cavity from the uvula to the apical closure. The fifth, sixth, and seventh formants are dependent on the larynx tube resonance and on various standing wave effects in the double resonator system above the larynx.

This clustering of no less than five formants in the frequency region between 2250 c/s and 5000 c/s, which here represents an average spacing of 700 c/s, is often seen in spectrograms and may in part be responsible for the "ringing" quality of the stationary part of a lateral. This close formant spacing is evidently related to the total length of the system, the mouth shunt included, which is $l_{tot} = 22 \text{ cm}$, as seen from the area function. The average spacing to be expected is thus $c/2l_{tot} = 800 \text{ c/s}$.

The effect of the anti-resonance, in addition to the suppression of the close-lying third formant, is to provide a high frequency emphasis starting at a frequency 40 per cent above the anti-resonance frequency, as determined by the shape of the unit zero or pole spectrum level curve; see *Section 1.32*. Alternatively, the third formant and the anti-resonance may be regarded as a pole-zero pair that can be removed from the spectral description because of the mutual neutralization, which may be regarded as complete except for the small constant level shift of $40 \log_{10}(F_3/F_{z1}) = 1.6 \text{ dB}$ above F_3 , as required by *Eq. 1.3-13*.

The result of this neutralization is that F_4 takes over the role of F_3 , and F_5 the role of F_4 , and so on. It has been verified from synthesis that a perfectly good [l] can be produced without a zero and that a shift up of F_3 towards a frequency of 2900 c/s close to F_4 adds to the naturalness of the sound. The identification relies on the sudden shift up of F_1 from the lateral to the adjacent vowel.

F_1 and F_2 of [l] are approximately equally dependent on the cavities in front of and behind the pharyngeal constriction. The dependency of F_1 on the lateral constriction, and of F_2 on the pharyngeal constriction, is definite and similar to the conditions for the production of [u] as stated in *Section 2.32*. For the palatalized [l,] the F_1 -and F_2 -cavity dependencies are the same as for the vowel [i]. It is thus obvious that F_2 represents a half-wavelength standing wave of the combined mouth pharynx system behind the point of articulatory closure. Since this point is more advanced for [l,] than for [i] the cavity length is larger and F_2 thus lower.

The calculated formant frequencies of [l,], $F_1 = 210 \text{ c/s}$, $F_2 = 1700 \text{ c/s}$, $F_3 = 2500 \text{ c/s}$, $F_4 = 3050 \text{ c/s}$, compare well with the measured data $F_1 = 230 \text{ c/s}$, $F_2 = 1600 \text{ c/s}$, $F_3 = 2300 \text{ c/s}$, and $F_4 = 3100 \text{ c/s}$. The third formant is associated with the external mouth cavity as for [l] and the fourth formant corresponds to a whole-wavelength resonance of the internal cavity system. F_1 is the fundamental mode of the complete system, as usual. The zero is calculated to fall at 3600 c/s. The higher value compared with the zero of [l] depends largely on the horn shape of the shunt cavity. In the spoken data this zero is less clearly seen, but it appears to coincide with the 2600 c/s minimum. The length of the shunting cavity thus seems to be underestimated in the calculations.

The production of [r] and [r,] involves several elements that are similar to the

production of [l] and [l̄]. The more retracted point of articulation of the [r]-sounds reduces the maximum attainable cross-sectional area values at the internal cavity system. The secondary place of articulation dividing this back cavity system of [r] into two parts was found at the uvular region and thus higher up than for [l]. As a result, F_2 will be more dependent on the anterior of the two internal cavities and reaches a somewhat higher position. F_3 remains affiliated to the mouth cavity in front of the tongue.

The spectrum of the spoken [r] shows a rather high F_1 ($= 500$ c/s), as seen from the broken line curve of *Fig. 2.5-3*, diagram 1. This is due to the time location of the sample within an open interval of the trill. The spoken sample displays a higher F_1 and lower F_2 than the calculated sample, and the spectrum envelope slopes off faster above F_2 . This is in part due to the high location of F_2 of the calculated curve. The effect of reducing the cross-sectional area at the apical passage, as in the transition from an open to a closed interval, can be studied in diagram 2. Besides the expected decrease of the level of the spectrum above F_1 , following the shift down in frequency of F_1 , there is an apparent lowering of F_3 towards a 2000 c/s position. This is due to the removal of the inductance from the apical passage, shunting the anterior mouth cavity in the open phase of the [r].

The agreement between the calculated F_1 , F_2 , and F_3 of [r̄] and those measured from the spoken sample is fairly good, as seen from diagram 3 of *Fig. 2.5-3*. The calculated $F_2 = 1700$ c/s and $F_3 = 2400$ c/s are 200 c/s greater than for the spoken data. The latter curve lacks the fourth and higher formants seen in the calculated curve. The relations of the first three formants of the [r̄]-spectrum to the vocal tract configuration are essentially the same as for the [l̄]. Thus F_2 is the half-wavelength resonance of the interior cavities, and F_3 is to a substantial degree dependent on the anterior mouth cavity.

The effect of a reduction of the cross-sectional area at the apical passage may be studied in diagram 4. There is a very small decrease of F_2 and a lowering of F_4 towards F_3 . The reduction of the spectrum level above F_1 , following the shift down of F_1 as the apical passage is closed, is evidently the main characteristic of an intensity minimum within a trill period. As seen from the spectrogram, *Fig. 1-4*, the duration of the speaker's trill interval was 40 msec and the trill frequency thus 25 c/s. Two voice fundamental periods of the pitch $F_0 = 100$ c/s are seen within the more open interval and there are traces of two within the closed interval.

2.6 FRICATIVES, AFFRICATES, AND STOPS

2.61 *Fricatives and Affricates*

The previous sections have been devoted to the study of the vocal tract response to a voice source. Whispered liquids and vowels should theoretically differ from voiced variants mainly by the random fine structure and more high frequency emphasized spectrum envelope of the source, owing to the glottis noise source. However, the production of these *vocalic* sounds may also involve noise produced at a supraglottal constriction. Such noise is not a necessary attribute but may add noticeably to the sound spectrum in, for instance, a stressed [i], [l], or [r]. The first part of a liquid following an unvoiced consonant may assimilate the lack of voicing and thus becomes aspirated¹ or even fricative, with reference to noise generation at a narrow supraglottal passage, as discussed in Sections 1.11 and A.22.

Acoustically, the common denominator of all sounds produced from a resonator system of a prescribed configuration is the particular set of formant frequencies of the vocal tract, i.e., the F-pattern. The differences in location of the source and the spectrum envelope of the source will only influence the relative intensity levels of the formants. It is to be expected that the cavities in front of and in the vicinity of a source will be of major importance for the spectral shape of any sound.

One complication that must be taken into account is the influence of the glottis-opening on formant frequencies and formant damping. The first formant is raised in frequency and highly damped when the glottis area is large. These general relations are common to all sounds that are produced in part or solely from a supraglottal source, thus also to stops, affricates, and fricatives.

From an acoustical point of view, when stationary sound spectra only are considered, there is no essential difference between an unvoiced fricative [r] and a fricative consonant [š] of the same articulation or between the corresponding voiced sounds [r]

¹ The term aspiration is here contrasted to frication primarily by a greater coupling between various parts of the vocal tract, but may also involve the participation of a glottal noise source alone or together with other simultaneous sources. In short, aspiration is equivalent to the sound [h].

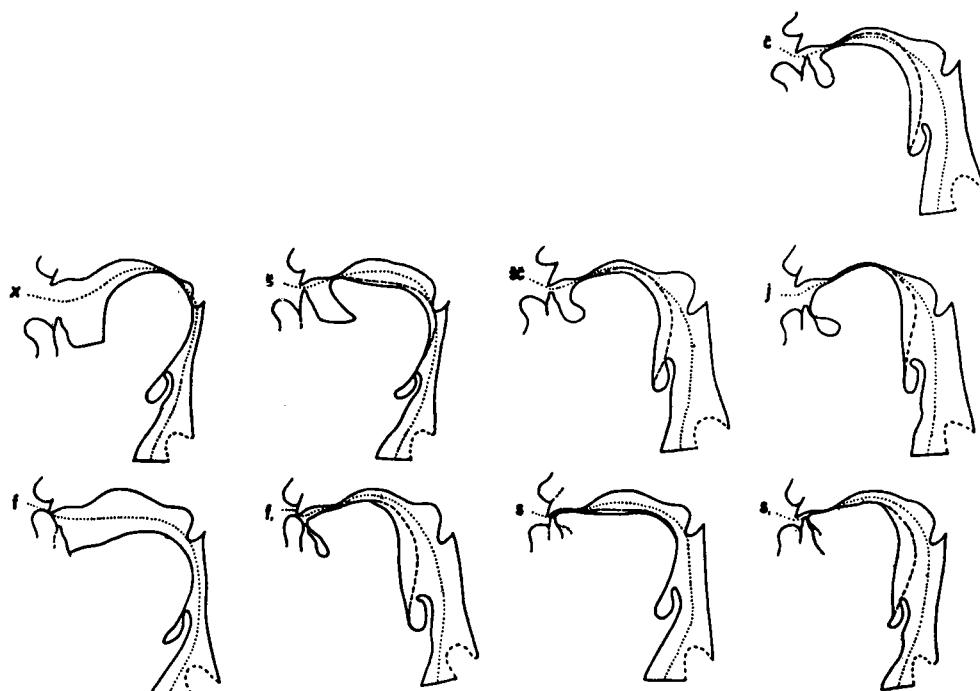


Fig. 2.6-1. X-ray tracings of the affricate [χ] and the fricatives [x], [s], [ʃχ], [j], [f], [f.], [s], and [s.].

and [z]. The difference lies mainly in the temporal modulation of the intensity which for [r] is maintained by a single or several successive flaps of the apex.

The role of the noise components of voiced fricatives, affricates, and stops should not be overestimated. The secondary role of the noise in the perception of the phoneme /j/ has motivated its classification as a glide; Halle (1954). Similarly, the voice source appears to be of primary importance for the identification of the phoneme /v/, as indicated by synthesis experiments with the Swedish speaking machine *OVE II*. A removal of the noise component impairs the naturalness more than the intelligibility. The structural ordering demands, however, the treatment of /f/-/v/ as a minimal pair.

The X-ray pictures of the subject's [v], [v.], [z], [z.], and [z] demonstrated sufficient similarities to the corresponding voiceless sounds [f], [f.], [s], [s.], and [s] for motivating calculations on the basis of the same vocal tract area function for both voiced and voiceless sounds. In addition to the voiceless sounds discussed above, Fig. 2.6-1 contains the X-ray tracing of the velar fricative [x], the glide [j], the palatal affricative [χ], and the palatalized variant of [s] which, following Halle (1954), is transcribed [ʃχ], indicating that the palatalization is conditioned by a following phoneme /χ/ which in some dialects appears as a mere lengthening of the palatalized fricative. The X-ray

tracing of the dental affricative [č] was not clear and has thus been omitted, but judging from the acoustic data there seems to be no doubt that its articulation was very similar to that of [s].

The area functions are shown in *Fig. 2.6-2*. It can be noted that the size of the cavity in front of the point of articulation is less for the labio-dentals [f] and [f̄] than for the dentals [s] and [s̄] and that this anterior cavity is larger for all other sounds. When taking length measures from these diagrams, it should be noted that the plane of radiation is located at a coordinate about 0.5 cm posterior to the origin, as indicated by the dotted first part of each area curve at the lips.

The actual width of a very narrow passage at the point, or rather region, of articulation cannot always be estimated with any accuracy from sagittal X-ray pictures, and the perpendicular dimensions are generally inaccessible to measurements. Supplementary palatographic and tomographic studies of the subject would have been of value. The missing dimensions have in part been replaced by data available from the literature. The minimum available cross-sectional area that could be simulated by means of the electric analog *LEA* was 0.16 cm^2 , and this was the standard value utilized for very narrow constrictions of both fricatives, affricates, and stops. Errors in the estimation of these dimensions do not severely affect the calculations since the length measure of a narrow passage is more crucial than the cross-sectional area. It is mainly the frequency of the fundamental resonance of the system, F_1 , that will be affected by *width errors*. However, even during complete articulatory closure, there is sound propagation through the walls of the vocal resonators to an extent that limits the minimum F_1 to the order of 100 c/s , as discussed in *Section 2.33*.

The configuration of the cavities behind the articulatory constriction, when *hard* and *soft*, i.e., *non-palatalized* and *palatalized* consonants such as [f]-[f̄], [s]-[s̄], [š]-[š̄] are compared, can simply be described by the presence versus absence of a tendency towards a narrowing in the region of the uvula, and in the upper part of the pharynx. As previously described in connection with the liquids, the result of this velarization, or rather the pharyngealization dividing the internal cavity system into two parts, is to lower the frequency of the second formant. The palatalization, on the other hand, provides optimal conditions for a high F_2 because of the single cavity structure and the gradual narrowing towards the closed end in the mouth.

Spectra of the voice source component of [z], [z̄], [v], [v̄], [ž], and [j̄] obtained from the *LEA* calculations and the corresponding measured data from the consonant+[a] syllables spoken by the subject can be studied in *Fig. 2.6-3*. There is a general tendency of 10-20 per cent higher formant frequencies in the calculated data than in the measured data but otherwise the agreement is good. The effect of palatalization is, as expected, largest for the labiodentals, comparing the calculated $F_2 = 850\text{ c/s}$ for [v] and the calculated $F_2 = 1900\text{ c/s}$ for [v̄].

The general fit between calculated and measured data of these sounds suggests that the noise component of the spoken sample was not very large. This seems to be the case for [j̄] also. The spoken samples of [z] and [z̄], on the other hand, show a high

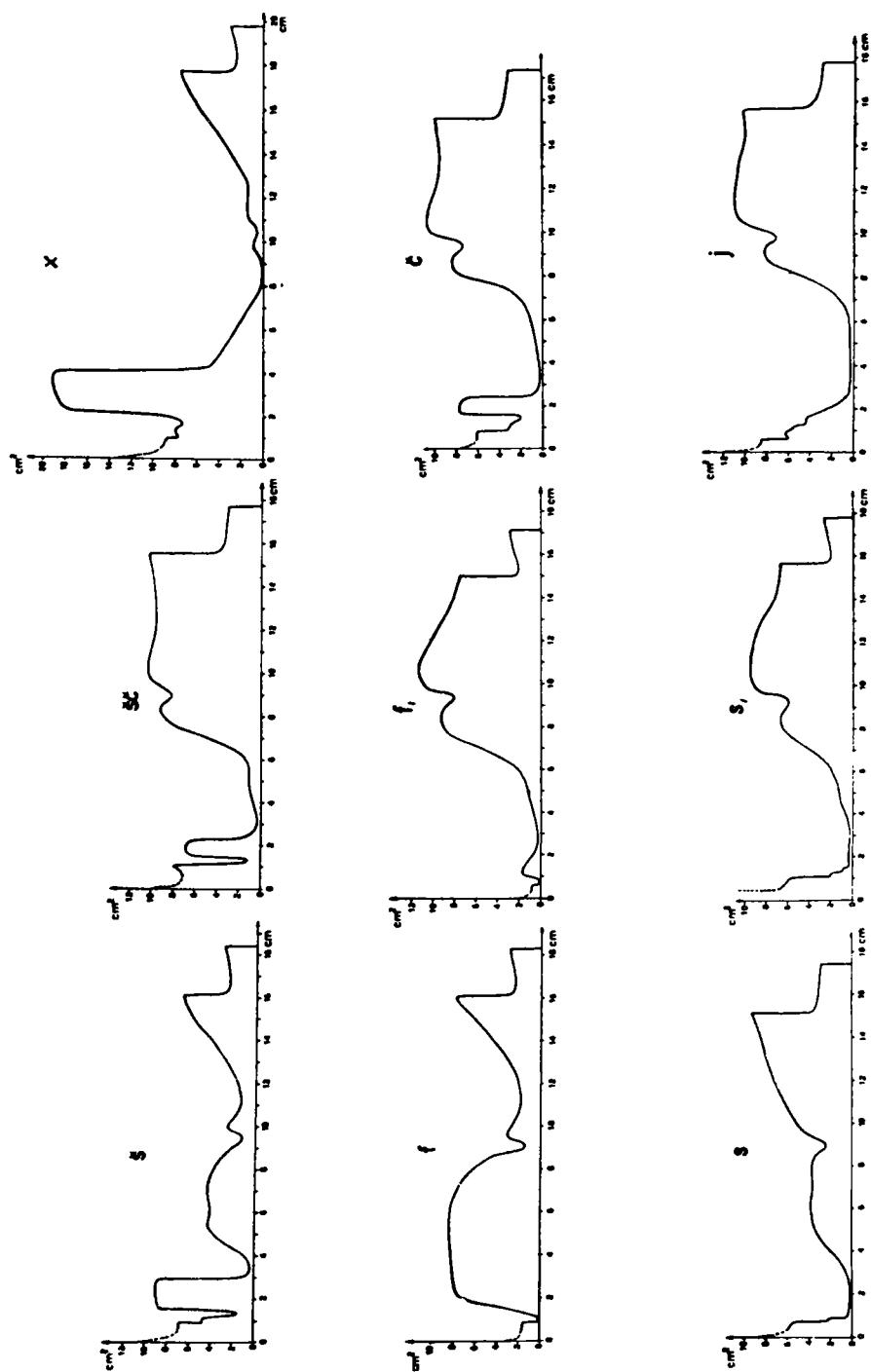


Fig. 2.6-2. Area functions of the sounds of Fig. 2.6-1.

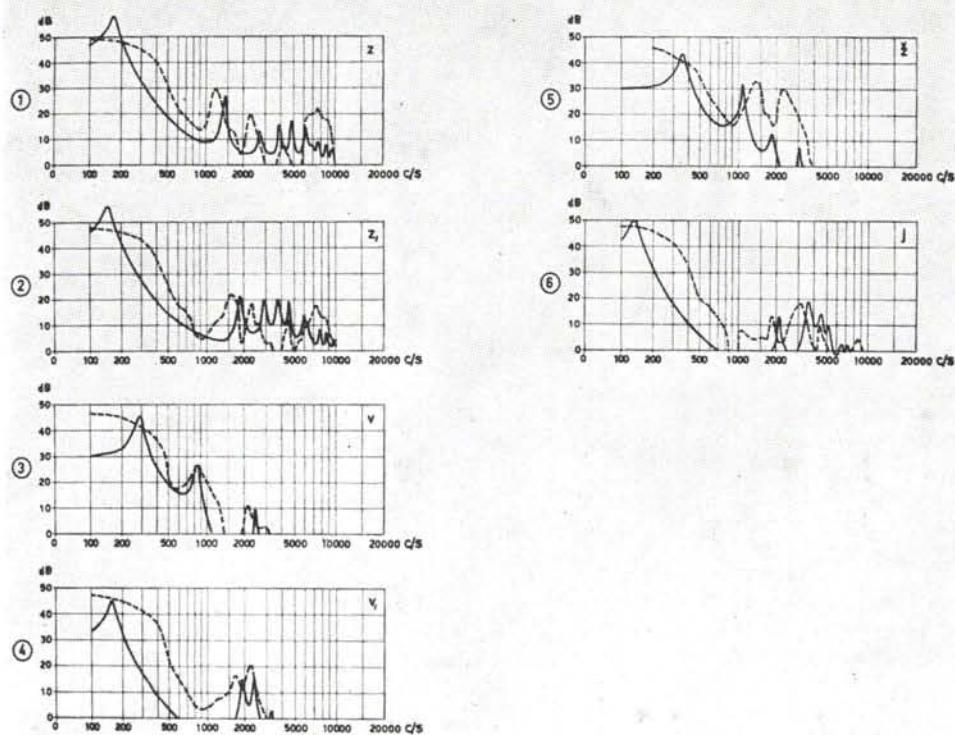


Fig. 2.6-3. Spectra of [z], [z̥], [v], [v̥], [ʒ], and [j] calculated on the basis of a standard voice source (solid line curves) and spectra of these sounds from connected speech (broken line curves).

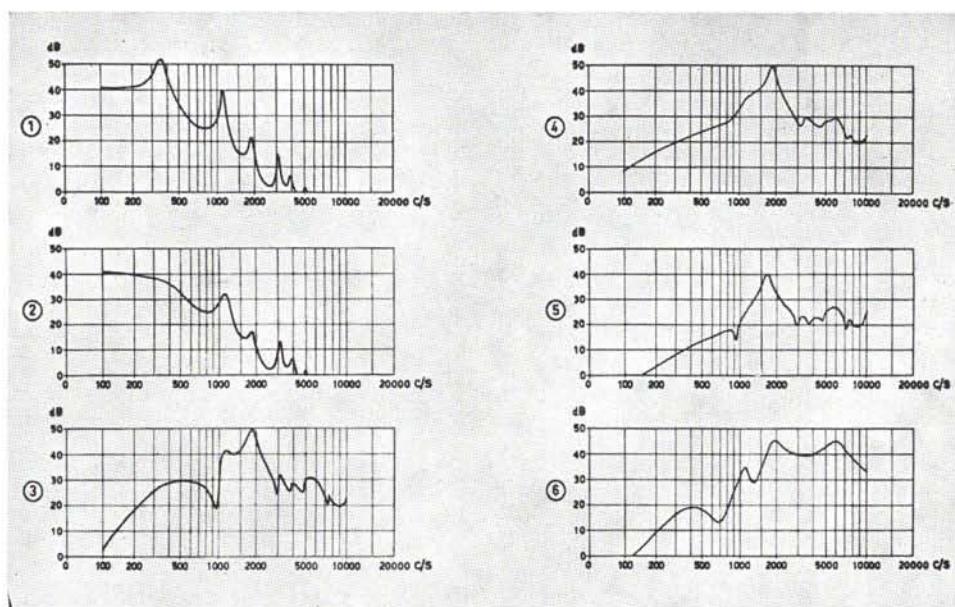


Fig. 2.6-4. Variation of source location and damping elements within the vocal tract configuration of [š].

- (1) Standard voice source -12 dB/octave , resistance 5 q.c. located at the glottis;
- (2) Same as (1) with the addition of a 0.25 q.c. resistance in series with the apical tongue pass;
- (3) Constant spectrum level source located at the tongue pass, otherwise as (2);
- (4) Glottis resistance decreased to 0.33 q.c. , otherwise the same as (3);
- (5) Glottis resistance restored to the value 5 q.c. and tongue pass resistor increased to 2 q.c. ;
- (6) Source located at the teeth in series with a $3/40 \text{ q.c.}$ resistor, in other respects same as (3).

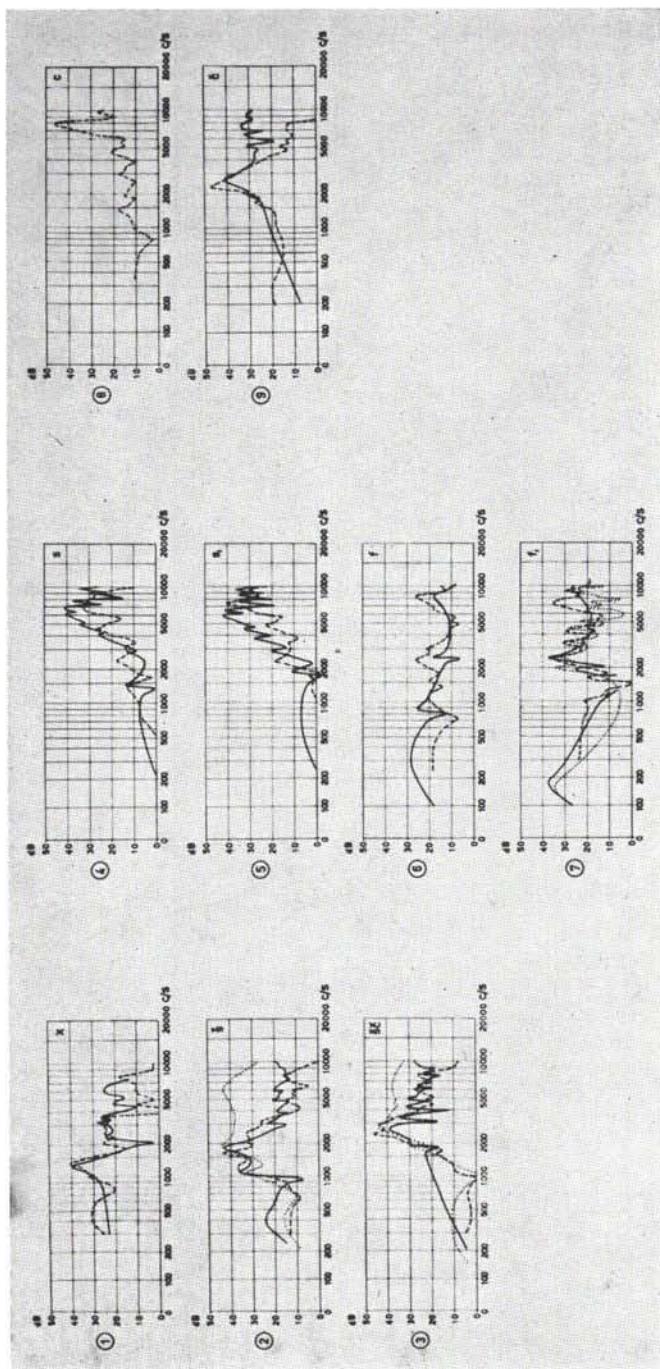


Fig. 2.6-5. Calculated spectra (solid and dotted lines) and spectra from the subject's connected speech (broken lines) of fricatives and affricates. Glottal termination $5 \text{ } \rho\text{c}$. The source is located in the middle of the articulatory constriction if not otherwise specified.

- (1) [x]—source of -6 dB/octave slope and resistance $0.25 \text{ } \rho\text{c}$;
- (2) [ʃ]—solid curve: tongue pass source of constant spectrum level and resistance $0.25 \text{ } \rho\text{c}$, tongue pass resistance $0.25 \text{ } \rho\text{c}$;
- (3) [ʃ̩]—same as (2) except for a $2.5 \text{ } \rho\text{c}$ source resistance;
- (4) [s]—teeth source of constant spectrum level and $0.025 \text{ } \rho\text{c}$, tongue pass resistance $2 \text{ } \rho\text{c}$;
- (5) [s̩]—same as (4);
- (6) [f]—labiodental source of -6 dB/octave slope and resistance $0.25 \text{ } \rho\text{c}$;
- (7) [t̩]—solid curve: labiodental source of -6 dB/octave slope and $0.25 \text{ } \rho\text{c}$ resistance. Dotted curve: glottis source of -6 dB/octave slope, resistance of labiodental pass retained;
- (8) Measured spectrum of the dental affricate [s];
- (9) [ʃ]—tongue pass source of constant spectrum level and resistance $2.5 \text{ } \rho\text{c}$.

frequency formant region extending from 5 to 10 kc which must be due to the dental noise source. From the spectrographic pictures of the spoken syllables, shown in the *Appendix*, it is further apparent that the third formant of [z] has a random fine structure, signaling noise, and because of the higher F_2 of this sample it can be concluded that it has been articulated more frontedly than during the X-ray session.

The unnaturally narrow bandwidths of some of the formants of the calculated data are due to the neglected constriction damping. The glottis resistance was 5 ρc as in the vowel calculations. It should also be kept in mind that the measured spectral sections were obtained from analysis with a 150 c/s wide filter, which causes a broadening effect of this order of magnitude.

The effect of varying the resistance elements, the position of the source, and the source characteristics at a constant vocal tract configuration is exemplified in Fig. 2.6-4, referring to calculations on the basis of the X-ray data for [š]. Curve 1, which is identical with that of [z] of Fig. 2.6-3, represents the response to the standard voice source of -12 dB/octave spectrum envelope and a 5 ρc glottis resistance. The F-pattern displayed by this curve resembles that of [r] in terms of the low-positioned F_3 typical of sounds of retroflex articulation. As seen from curve 2, the introduction of a 0.25 ρc resistance in series with the apical constriction causes a 13 dB reduction of the level of the first formant at 350 c/s, which, as a result, is damped out almost completely. The second formant at 1050 c/s is damped 8 dB, and the third formant at 1900 c/s is attenuated 4 dB only. The fourth and fifth formants are not much affected.

Curve 3 shows the effect of shifting the source to the apical constriction, retaining its resistance value, and changing the source to a constant spectrum level function. The effect of the source shift is to introduce anti-resonances, i.e., zeros, at 0 c/s, 950 c/s, 3000 c/s, 3900 c/s, 4900 c/s, and at some higher frequencies. These are the frequencies of infinite impedance, as seen from the apical passage towards the glottis. The first zero causes a high degree of attenuation of the spectrum below 200 c/s. The second zero, found just below F_2 , reduces the level of the second formant, and F_3 constitutes the main peak, which is to be expected because of the F_3 -dependency of the cavity system anterior to the source. Formants of higher number are considerably attenuated owing to the presence of associated zeros.

Curve 4 shows the damping effect when the glottis resistance is reduced to 0.33 ρc . This value would correspond to a resistive matched termination of the trachea at the entrance to the bronchial system and the lungs, assuming a trachea cross-sectional area of 3 cm^2 and a wide-open glottis. As a result of this high damping, the residual traces of F_1 and F_2 are eliminated, and the zero just below F_2 is barely detectable. Probably this degree of damping is larger than what occurs in natural speech at least at frequencies above 1000 c/s. As seen from curve 5, a similar effect occurs if the constriction resistance is increased by a factor of 10 retaining the 5 ρc glottis load.

Finally, it may be seen from curve 6 that a shift of the source to a position at the teeth, retaining the conditions underlying curve 3, except for a series resistance of

$0.067 \text{ } \mu\text{c}$ at the teeth passage, introduces zeros in-between all first three poles. Corresponding anti-resonances and formants are appreciably damped by the source resistance. The spectrum level at frequencies above the main 1900 c/s peak is rather flat up to the 6000 c/s peak. This latter peak is the fundamental resonance of the cavity anterior to the teeth, i.e., between the lips.

The emphasis of the frequency region above 2000 c/s , when the source has a more advanced position, is due to the presence of the additional conjugate complex zero at lower frequencies, that is, one at 700 c/s and one at 1300 c/s instead of the one at 950 c/s associated with a source location at the tongue constriction. It can be shown from the equivalent circuit representation that the effect of shifting the location of a source from the posterior to the anterior constriction terminating a fairly isolated cavity is to introduce a small shift of zero frequencies and to introduce an additional zero at a frequency tuned by the cavity volume and the inductance of the constricted passage behind the cavity. In the example above, this is the 700 c/s zero. The zero enters the filter function of vocal transmission via an additional factor $[1 - (f^2 - jfB_z)/F_z^2]$, where F_z is the zero frequency and B_z its bandwidth; compare Eq. 1.3-12. At a frequency f of the spectrum envelope 40 per cent above F_z this factor is almost equal to unity and at higher frequencies it approaches a $+12 \text{ dB/octave}$ rise. It behaves quite inversely to the unit pole function discussed in Section 1.32.

From diagram 2 of Fig. 2.6-5, it can be seen that the measured curve for the subject's spoken [ʃ] (broken line) shows a closer overall correspondence to the curve calculated for the tongue constriction source (solid line) than for the teeth passage source (dotted line). In both calculations the source was assumed to have a constant spectrum level. The curve representing the spoken data, sampled from the monosyllabic test words, see the Appendix, follows the curve calculated for the constriction source within 12 dB in the frequency range from 300 c/s to 9000 c/s , and lies constantly $8\text{-}10 \text{ dB}$ lower than the calculated curve at frequencies below 800 c/s .

The fit in the low frequency region is better for the teeth source, but at frequencies above the main peak the teeth source should be attenuated 12 dB/octave in order to provide a good fit with the measured curve. There is no doubt that the main peak at 1650 c/s in the measured curve is identical with F_3 and thus with the 1900 c/s peak of the calculated data. The calculated $F_2 = 1050 \text{ c/s}$ compares well with the $F_2 = 1150 \text{ c/s}$ of the measured data. The extremely sharp fall of the spectrum curve from 1050 c/s to 950 c/s is found in both curves. It is possible that the spoken [ʃ] has been produced with both a dental source and a tongue constriction source of comparatively low spectral level below 1000 c/s . It is also conceivable that the measured 750 c/s zero originates from a large coupling to the trachea. It is of some interest to compare our calculated [ʃ]-spectrum with that from the word *shack* of the article by Hughes and Halle (1956). The similarities are appreciable.

The calculations of the spectrum of [x] represent the most successful of all attempts made here to reconstruct the spectral characteristics of sounds of connected speech from the same subject's sustained X-ray pictures. Except for the small dip at 750 c/s

and the minor plateau at 450 c/s, which might be explained from a finite trachea coupling causing a rise in F_1 and introducing a zero, there is a substantial agreement of the spectral topography, both with regard to peaks and valleys. The main peak at 1300 c/s represents the resonance of the cavity in front of the articulatory constriction and is identical with F_2 . The calculated zeros at 2000 c/s and 4200 c/s correspond to measured zeros of 1800 c/s and 4050 c/s. This conformity is a support of the assumption that the major source is located in the region of the constriction.

The calculations were based on a source of integrated white noise, i.e., a source spectrum envelope falling 6 dB/octave. An additional attenuation of 6 dB/octave above 4000 c/s would provide an even better fit. The source resistance has damped out the F_1 peak and compensates for the rather too high, and thus less loading, standard glottis resistance.

The spectrogram of the test word [xa] of the *Appendix* is worth some comments. There are a few temporally well-defined striations occurring at 30-50 msec intervals, which must be related to either some spurious vocal cord vibrations or to a similar vibrational mechanism at the velar constriction. The latter explanation seems more probable in view of the high F_1 . The mechanism would thus involve an accidental complete closure followed by a break-through and a restoring *Bernoulli* force causing an irregular modulation of the expirated air. The repetition of this process would cause a vibrational force on the posterior part of the velum and the uvula, in an extreme case exciting a uvular trill.

On the other hand, this or similar phenomena of spurious energy quanta can often be seen in the noise interval of velar, palatal,² and even labial stops. These energy quanta are probably due to the impact of suddenly released air bubbles more than to the associated turbulence. The source spectrum envelope of -6 dB/octave inferred from the calculation conforms with the step function spectrum of an air stream that is suddenly switched on. From the intensity versus time records of the word [xa], it is apparent that it is mainly the spectral level in the region below 3000 c/s that is modulated by these spurious quanta. The intermediate sound intervals carry noise of a lower intensity originating from a source with less steeply falling spectrum envelope. There is also the theoretical possibility that the sound quanta are due to a modulation of the air stream by a few spurious vocal cord vibrations. These phenomena are not satisfactorily explained yet.

The calculations for the sound [šč], see diagram 3 of *Fig. 2.6-5*, show the same general tendencies as discussed for the sound [š]. The participation of a dental source seems more probable than for [š], in view of the conformity with regard to the zero at 1000 c/s, the close agreement between the curves below the main peak at 2700 c/s, and the lesser high frequency attenuation of the source needed to fit the spoken curve and the calculated dental source curve in the region above the main peak. A part of the remaining level differences at higher frequencies may also be accounted for by

² Double explosions in [k] have been observed by Fischer-Jørgensen (1954).

the lower frequency location of the main peak in the spoken data which is supported by F_3 at 2400 c/s and F_4 at 3000 c/s, as compared with $F_3 = 2700$ c/s and $F_4 = 3300$ c/s of the calculated curve.

The main acoustical difference between [šč] and [š] lies in the higher frequency location of the main spectral peak of [šč] attributable to the smaller dimensions of the front cavity. This relation was also found when [š] and [x] were compared. In the latter sound, F_2 dominated the consonant spectrum; in [š], F_3 was slightly more apparent than F_2 ; and in [šč], F_3 was the leading formant with F_4 second in intensity, only 3 dB weaker.

The spectrum of the affricate [č], measured in the middle of the fricative noise segment, is dominated by F_3 , but also here F_4 is next in intensity, 10 dB below the F_3 -peak, as seen more clearly in the spectrographic presentation. A dental source seems less probable for [č]. The residual F_2 of the noise spectrum of [č] and [šč] is of an insignificantly low level, but the higher F_2 -location in comparison with [š] and [x] is apparent from the more positive transitions seen in the spectrograms.

The calculations for the non-palatalized [f] were carried out on the assumption of a labiodental source of -6 dB/octave slope. The overall level of both the calculated and the measured sound spectrum is essentially flat, and there is a peak at 8500 c/s in both data. The residual formant structure in the measured sample is obscured by an extra formant at 1500 c/s which probably is due to a large coupling to the trachea system. The measured zero at 750 c/s appears to be identical with the first conjugate zero at 800 c/s of the calculated curve. The measured curve originates from section sample *II* of the spectral display in the *Appendix*. It should be observed that F_2 of sample *III* immediately before the onset of the vowel is located higher than the starting point of the F_2 -transition within the vowel. This is typically the effect of the removal of the trachea coupling at the onset of the phonation and conforms with the observation of a 200 c/s lower F_2 in the speaker's [v] than in his [f]. Those formants that are sufficiently dependent on the pharynx cavity are slightly higher in frequency when the glottis stands open. Here the combination of lip-rounding and pharyngealization conditions the necessary amount of F_2 -back-cavity dependency; see Sections 1.43 and 2.32.

Diagram 7 of Fig. 2.6-5 contains a calculated curve for [f] derived from a labiodental source of -6 dB/octave slope, the measured spectrum, and further a calculated curve derived from a larynx source of the same slope as the constriction source. The main shape features are the same in the curve from the spoken data and the curve derived from the labiodental source. The level of the main peak coinciding with F_3 at 2400 c/s is 14 dB above the level of F_2 in the measured curve and 13 dB in the calculated data. There is a prominent anti-resonance effect in the region between F_1 and F_2 which is caused by the first conjugate zero located at 1500 c/s in the calculated spectrum and at 1300 c/s in the measured spectrum. The effect of the next zero found just above F_2 is small owing to its close proximity to F_2 with a resulting mutual neutralization of the zero and the F_2 -pole.

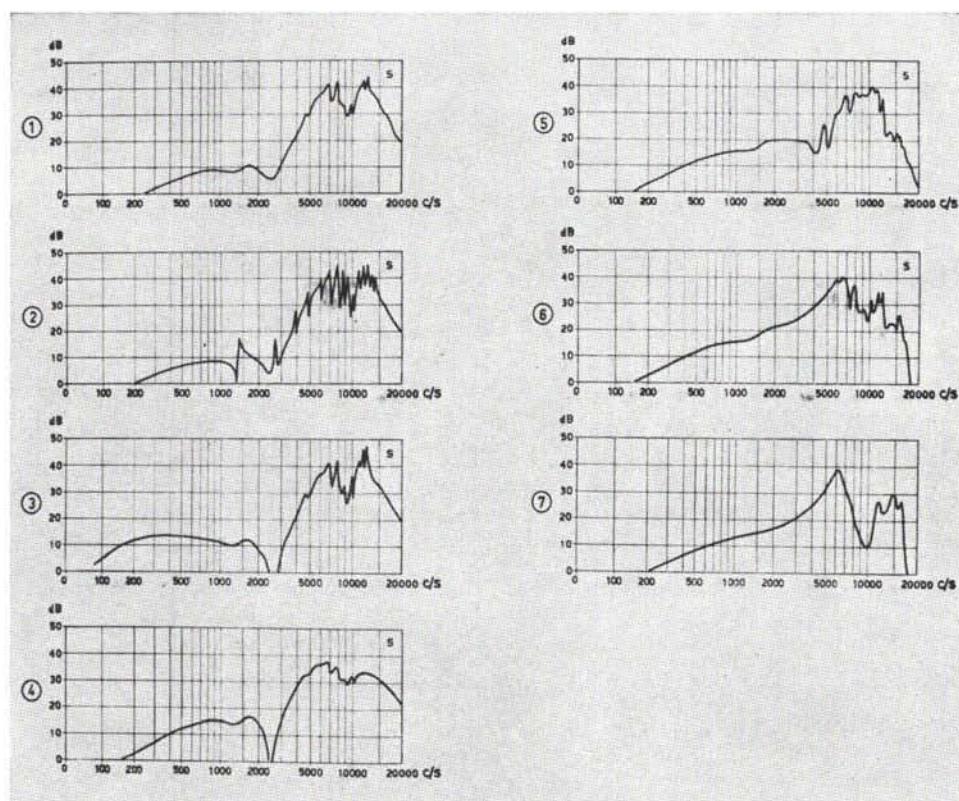


Fig. 2.6-6. The effects of glottis damping, source resistance, the position of the source, and of the cavities behind the source on the calculated spectrum of [s]. A constant spectrum level source was assumed. The following area settings at successive 0.5 cm sections along the electrical vocal tract *LEA* from the radiating surface at the lips were utilized as an approximation of the front part of the area function:

Section nr	1	2	3	4	5	6
Area cm^2	6.5	0.65	0.16	0.16	0.16	0.32

The remaining sections of the vocal tract were chosen in close approximation to the area function of [s] from Fig. 2.6-2.

- (1) Source resistance $0.025 \text{ } \Omega\text{c}$, inserted in front of section 2, i.e., at a location corresponding to the teeth. A $2 \text{ } \Omega\text{c}$ series resistance was placed between sections 4 and 5 and the glottis resistance was $0.33 \text{ } \Omega\text{c}$. These were the conditions utilized for the calculation of [s], diagram 4 of Fig. 2.6-5;
- (2) The glottis resistance removed, otherwise as (1);
- (3) Glottis resistance same as in (1) (the $2 \text{ } \Omega\text{c}$ series resistance at section 4 removed);
- (4) Source resistance at the teeth increased to $2 \text{ } \Omega\text{c}$, otherwise the same as (3);
- (5) The location of the source in front of section 3, otherwise the same as (4);
- (6) The location of the source in front of section 5, otherwise the same as (4);
- (7) The cavities behind section 6 replaced by a short-circuit, otherwise the same as (6).

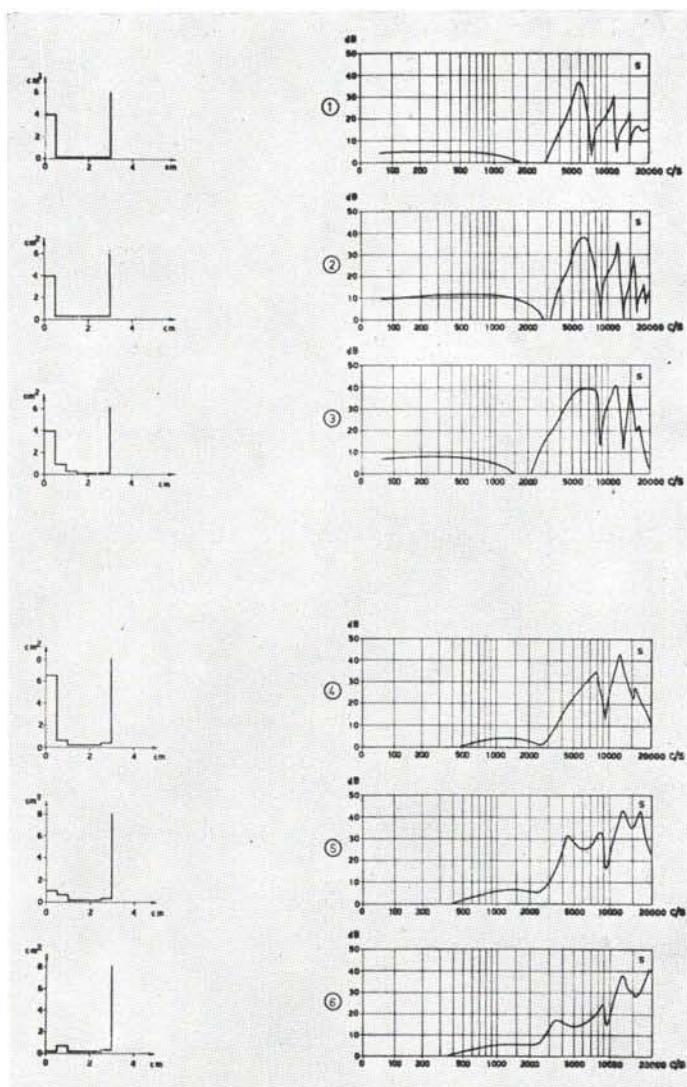


Fig. 2.6-7. Highly simplified front cavity systems for the production of [s] and the corresponding spectra. Cavities behind the first 3 cm of the system are represented by a short-circuit. The minimum cross-sectional area is 0.32 cm^2 for model (2) and 0.16 cm^2 for the others. A source of constant level spectrum and of 0.025 pc resistance was inserted at $x = 0.5 \text{ cm}$ from the front end.

The pole-zero cancellation is most apparent in those pairs of the calculated spectrum found above F_3 . The measured curve, on the other hand, shows a clear and well-developed structure of $F4$ and $F5$, indicating that the additional source must be active and that the position of this source must be such as to minimize the pole-zero overlap at these formants. A secondary source at the palatal narrowing can occur in strongly palatalized labials, providing the lips are opened at a faster rate than the tongue is lowered. As judged from the spectrogram, the [f]-section referred to here was sampled at a point where the aspiration sets in in addition to the frication, as is typical of the latter part of an unvoiced consonant in initial position. The existence of an additional glottal source is plausible in view of the good agreement between $F3$, $F4$, and $F5$ of the spoken data and those derived for the glottis source. The main fricative formant of the [f] is found at 7500 c/s and is related to the small cavity in front of the upper front teeth in combination with the radiation impedance. In the calculated curve it appears at 9500 c/s. The 8500 c/s peak in the [f]-spectrum is of the same origin.

The spectrum envelope curves calculated for [s] and [s,], see diagrams 4 and 5 respectively, were derived on the basis of a constant spectrum level source situated at a coordinate corresponding to the upper front teeth. Typical of both calculated and measured curves are their high-pass characteristics with a cutoff frequency at 2500 c/s and about a slope of 15-25 dB/octave. The steepness of the filter slope is greater in the calculated curves than in the measured curves. The maxima of the spoken curves are found in the frequency region of 6000-9000 c/s and at about 20 per cent lower frequencies in the calculated data.

As will be described in more detail in the following discussion, the resonance mechanism underlying the typical spectral shape features of [s] is the half-wavelength resonance of the tongue-teethridge channel regarded as a tube of about $l = 2.5 \text{ cm}$ effective length. The main [s]-formants would thus include the $c/2l = 7000 \text{ c/s}$ resonance of this tube, the next higher mode of $2c/2l = 14000 \text{ c/s}$, and in addition the resonance of the lip cavity in front of the teeth. The high-pass characteristics of the spectrum are further accentuated by a zero at about 3000 c/s originating from the quarter-wavelength conditions of the constriction channel, as seen from the source.

The details of the formant structure superimposed on the main spectrum envelope are not very well reproduced in the calculations, except for the higher F_2 - and F_3 -positions in the palatalized [s,]. As judged from the much suppressed F_2 and fairly prominent F_3 in the spectrum of the spoken [s,], there could be active an additional source at the prepalatal inlet to the alveolar channel. There is no essential difference between the main spectral shapes of [s] and [s,].

The main spectrum peak of the affricate, [c] is also located at 7000-8000 c/s, but it is narrower than for [s], which would imply a single-resonance structure. No calculations were performed on [c].

A few systematic variations of the synthesis conditions for the calculated [s] were made. The cavity configuration within the first 3 cm of the system, i.e., of the first

six filter sections of the analog device *LEA*, was as follows:

Section No.	1	2	3	4	5	6
Area cm ²	6.5	0.65	0.16	0.16	0.16	0.32

The posterior parts of the vocal tract were those determined from the X-ray tracings.

As seen from *Fig. 2.6-6*, the superimposed formant structure may be damped out either by a large series resistance within the constriction or by the assumption of a small glottis resistance. The effect of introducing a high series resistance at the teeth is to decrease the level difference between the parts above and below the cutoff frequency 2500 c/s.

The same leveling effect (but to a greater extent) is found by shifting the source from a dental location to a location at the internal inlet to the tongue channel. An intermediate source position³ provides a lesser degree of reduction of the typically rapid rise of the [s]-spectrum envelope from 3000-6000 c/s, as shown by a comparison of curve 5 with curves 4 and 6.

The substitution of the cavities behind the tongue channel for a short-circuit does not change the spectrum much except for the introduction of a zero at 10000 c/s and the complete elimination of the formant ripple on the spectral curve.

A study of a few systematic changes of the configuration of the effective resonator system of [s] is carried out in *Fig. 2.6-7*. The total length of the model is 3 cm, and the source was located at $x = 0.5$ cm from the radiation end. A source resistance of 1 acoustical ohm was utilized. The posterior end of the system was short-circuited. Comparing curves 1 and 2, it may be seen that a widening of the channel cross-sectional area from 0.16 cm² to 0.32 cm² causes a shift of the first zero from 2500 c/s to 3000 c/s and a shift down in frequency of one of the poles within the 4000-8000 c/s formant. A widening of the front part of the channel causes a further shift down in frequency of one of these poles and further a shift down in frequency of the first zero.

From diagrams 4, 5, and 6, it may be seen that a narrowing of the first 0.5 cm section at the lips causes no appreciable shift in the zero frequency and essentially a shift down in frequency of the pole representing the cavity anterior to the narrowing. It is of interest to see that the change of lip-opening from 6.5 cm² to 0.16 cm², does not change the frequency of this pole more than from 8000 c/s to 3500 c/s. The main spectral characteristics are apparently retained during lip-rounding except for this ascending labiodental resonance. The calculations for the frequency range above 10000 c/s, however, cannot be given very great importance owing to the restricted applicability of the one-dimensional wave theory.

Palatograms of [s]-sounds (Dieth, 1950, p. 187; Koneczna and Zawadowski, 1956) reveal the large variability in width and shape of the tongue-teethridge channel as between various speakers and different languages. As an alternative to the single tube approximation extensively made use of here, there are [s]-articulations which

³ According to Meyer-Eppler (1953) the effective source position is at the place of maximum constriction and not necessarily at the edge of the upper incisors. This statement has not been disproved by the present investigation but the edge effect interpretation appears to be more likely.

are better described by the existence of an additional constriction somewhere in the middle of this tube.⁴ The effect of narrowing the center part of a tube of length l open at both ends is not very great. The fundamental resonance at $f = c/2l$ is not changed much, and the next higher resonance $f = 2c/2l$ of the original tube is shifted down close to the $c/2l$ position. If the main constriction has a length of the order of 1.5 cm, the result is approximately that of diagram 4 or 5 of Fig. 2.6-7.

2.62 Stops

The experimental control data of spectral sections for fricatives and affricates were taken at the interval of maximum intensity, which occurred near the middle or towards the end of the sound. The sections for stops were also taken at the interval of maximum intensity which generally fell at the beginning of the unvoiced sound interval identified as the stop burst. Palatalized dentals have a strong tendency towards affrication which advances the instant of maximum intensity.

The position of spectral sections was classified by the system of $I = \text{voiced occlusion}$, $I = \text{initial transient of the explosion}$, $II = \text{fricative sound intervals}$, $III = \text{aspirated or mixed aspirated and fricative sound intervals}$, referring to the terminology introduced in Section 1.11. All calculations and measurements (see the Appendix) reported here pertain to either the interval I or II or to their combination, the latter in the case of sounds of short duration and in general when intervals I and II cannot be held apart.

Calculations were carried out on the unvoiced stops only. As seen from the X-ray traces of Fig. 2.6-8 and the area functions of Fig. 2.6-9 compared with those of Fig. 2.6-1 and Fig. 2.6-2, there is a substantial similarity between [k] and [x], [k₁] and [j], [p] and [f], [p₁] and [f₁], [t] and [s], and [t₁] and [s₁].

All X-ray photographs pertain to the state of complete closure before the explosion, but the calculations were carried out on the assumption of a minimum cross-sectional area of 0.16 cm². The particular value chosen is not very critical for the main shape of the derived spectral characteristics.

A source spectrum envelope of -6 dB/octave and a source resistance of 0.25 ρc were adopted for the stop sound calculations, except for [t₁], which was treated as the fricatives, i.e., on the basis of a source resistance of 2 ρc and a flat source spectrum. Because of the lack of data on the glottis impedance, no attempt was made to simulate the open glottis conditions, and the standard value of 5 ρc representing the semi-closed conditions during phonation was thus retained in all calculations. This resistance provides too low formant damping at low frequencies and is thus more representative of voiced stops. On the other hand, the resistive, matched impedance termination of the larynx tube of 0.33 ρc gave too large damping effects in the frequency range above the first formant. The actual impedance of an open glottis probably contains appreciable reactive components. As long as the vocal

⁴ According to Svend Smith (personal communication) the place of maximum constriction is generally at the "Papilla incisiva palati".

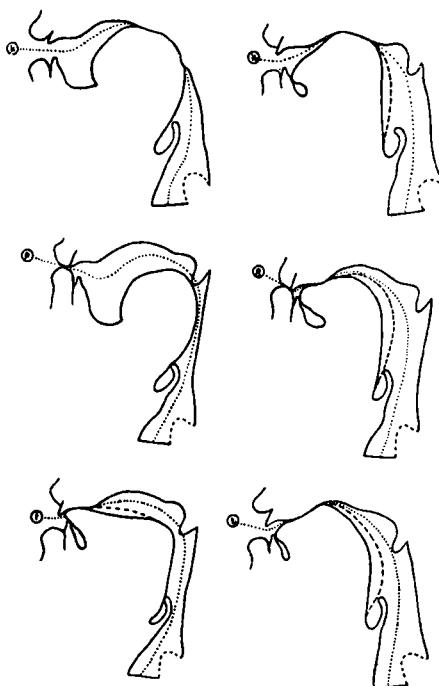


Fig. 2.6-8. X-ray tracings of [k], [p], [t], and [k̪], [p̪], [t̪].

cords are not maximally apart, the impedance of the passage in-between will be mostly that of an inductance, causing a small increase of formant frequencies as discussed above for fricatives. Resistive elements in series with the glottis inductance have to be transformed to a parallel resistance before their effect can be evaluated, as described in *Section 2.33*.

The high source resistance anticipated for fricatives, affricates, and for affricated stops represents a probable maximum value, as can be derived from *Eq. A.2-10* and *14* on the assumption of an airflow of high velocity, $v = 6000 \text{ cm/sec}$, through a narrow passage, $A = 0.08 \text{ cm}^2$. For stop sounds the source resistance is expected to vary over great ranges as determined by the time-variable flow conditions and dimensions. The values chosen for the calculation have provided fairly representative spectral curves, and they are not very critical for the essentials of the spectrum shapes; see further the discussion in *Section 2.64*.

Calculations of the spectra of [p] and [p̪] were performed on the basis of a source location at a coordinate of 0.5 cm from the front end of the vocal tract analog model. The source was located at a coordinate of $x = 1 \text{ cm}$ for [t] and [t̪] and in the middle of the tongue constriction for [k] and [k̪]. The small difference in source coordinate as between dentals and labials is not significant.

Calculated and measured spectra of the unvoiced stops can be studied from *Fig.*

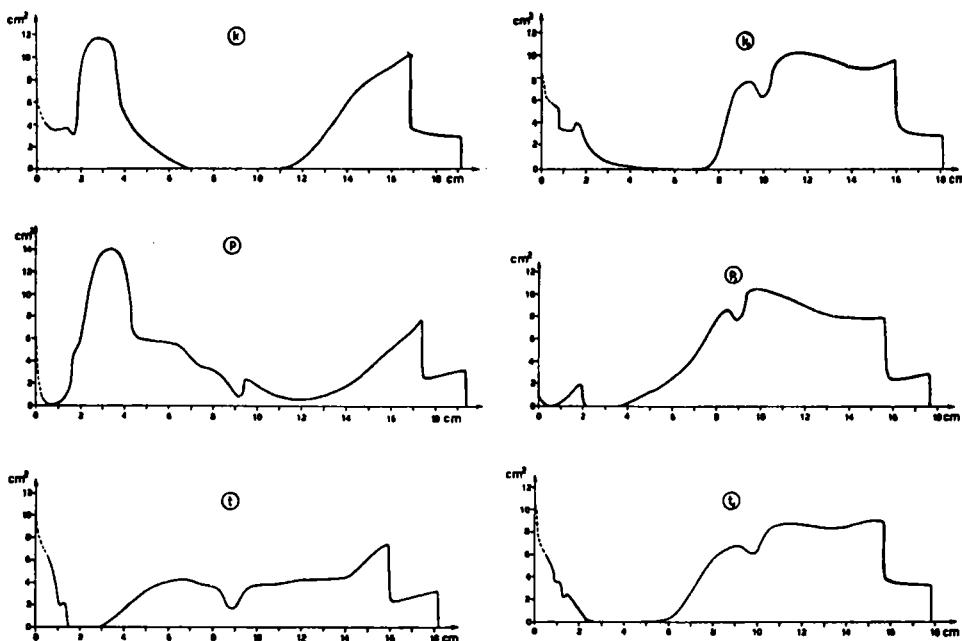


Fig. 2.6-9. Area functions of [k], [p], [t], and [k.], [p.], [t.].

2.6-10, which also contains some spectra of the voiced stops. From the general agreement in terms of the main spectrum shape and the average spectral levels at low, middle, and high frequencies, it can be concluded that representative values of the source spectrum envelope have been utilized.

The conformity between the spoken and calculated [k]-spectra of diagram 1 is so good that each peak of the spoken sample may be identified with a corresponding peak in the calculated curve. The latter shows about 20 per cent higher frequency values, in the frequency range below the 5000 c/s zero. It thus appears that the spoken sample was articulated with a somewhat more advanced tongue position. Referring to the calculated curve, the main peak at 1200 c/s is the fundamental mode of the mouth cavity and is identical with F_2 .

The next higher mode of resonance in the mouth cavity enters the F-pattern as F_3 . There is a zero at 3100 c/s representing the first frequency above $f = 0$ where the impedance in the direction of the glottis, as seen from the source, is infinite. This zero and the 4900 c/s zero reduce the levels of the fourth and fifth formants originating from the pharynx cavity and the larynx tube. The higher formant region at 7 to 9 kc is supported by the third and fourth resonance of the mouth cavity. In the spoken sample of [k], the first conjugate zero occurs at a frequency close to F_3 , and it is thus F_4 that constitutes the second resonance of the mouth cavity.

The main peak of the palatalized sound [k.] comprises F_3 plus F_4 . The second

formant is eliminated by a zero that is not very well seen in the calculated curve. The next higher zero occurs at 4900 c/s.

Except for the greater traces of a residual formant structure in the measured spectra of [p] and [p̄], there is a conformity in terms of the distribution of the main spectral energy. The 1300 c/s zero of [p̄] constitutes a well-developed anti-resonance dividing the sound spectrum into a lower and a higher part in a manner similar to the anti-resonance effect in the [f̄]-spectrum. One main cause of the very weak trace of formant structure superimposed on the calculated [p̄]-spectrum is the minimal area assumed for the secondary constriction at the palate bringing the poles and zeros of the pharynx cavity close together. Similarly the absence of a formant structure above F_2 in the calculated [p]-spectrum may be due to both too low a constriction area and too large a source resistance.

The double peak of $F2+F3$ of palatalized labials often shows up in spectrograms as an apparent central energy concentration, which might make the labial interval resemble that of palatals. Such examples have been demonstrated by Halle, Hughes, and Radley (1957). However, the spectrum is distinct from that of palatals owing to the zero just below F_2 and the more prominent low frequency region. The two peaks are generally less close than in palatals, and their overall energy is lower. In addition there are, of course, the apparent transitional cues.

The calculated [t]-spectrum does not compare very well with the measured spectrum except for the overall distribution of the sound energy. The first zero at 850 c/s and the weak formant at 1100 c/s in the measured curve below the $F2$ of 1600 c/s must belong to an interfering resonator system, either to the nasal system or more probably to the trachea, as discussed in connection with the fricatives. This spectral detail is not incidental. This has been observed in an earlier survey of Swedish consonants (Fant, 1949). The first attempt at calculating the [t̄]-spectrum was made on the basis of the area function of Fig. 2.6-9, resulting in the curve of diagram 6 of Fig. 2.6-10. However, if the posterior part of the palatal constriction is widened to the double area, with the minimum area retained in the anterior 2 cm of the constriction, the fit becomes much better, as seen from diagram 7. This is probably a more natural articulatory configuration.

As seen from Fig. A.13-16 of the Appendix, three spectral sections of [t̄] were taken—one at the explosion, one at a fricative interval 20 msec later, and one at a mixed fricative aspirated interval at the end of the noise part of the sound. The second spectral sample was utilized as the basis of comparison with the calculated data. All three samples show a high spectral level in the region of 6000-9000 c/s. The average slope of the first interval is flat, i.e., a constant spectrum level, and there are very weak traces of the F-pattern, except for $F2$ at 1900 c/s and the extra formant at about 1300 c/s. In the fricative interval the average spectral slope rises about 6 dB/octave up to the main peak at 7500 c/s. Finally, in the third interval the average slope is again flat owing to a rise of the spectrum level below 5000 c/s and an attenuation at higher frequencies. In this interval, the superimposed formant structure is very

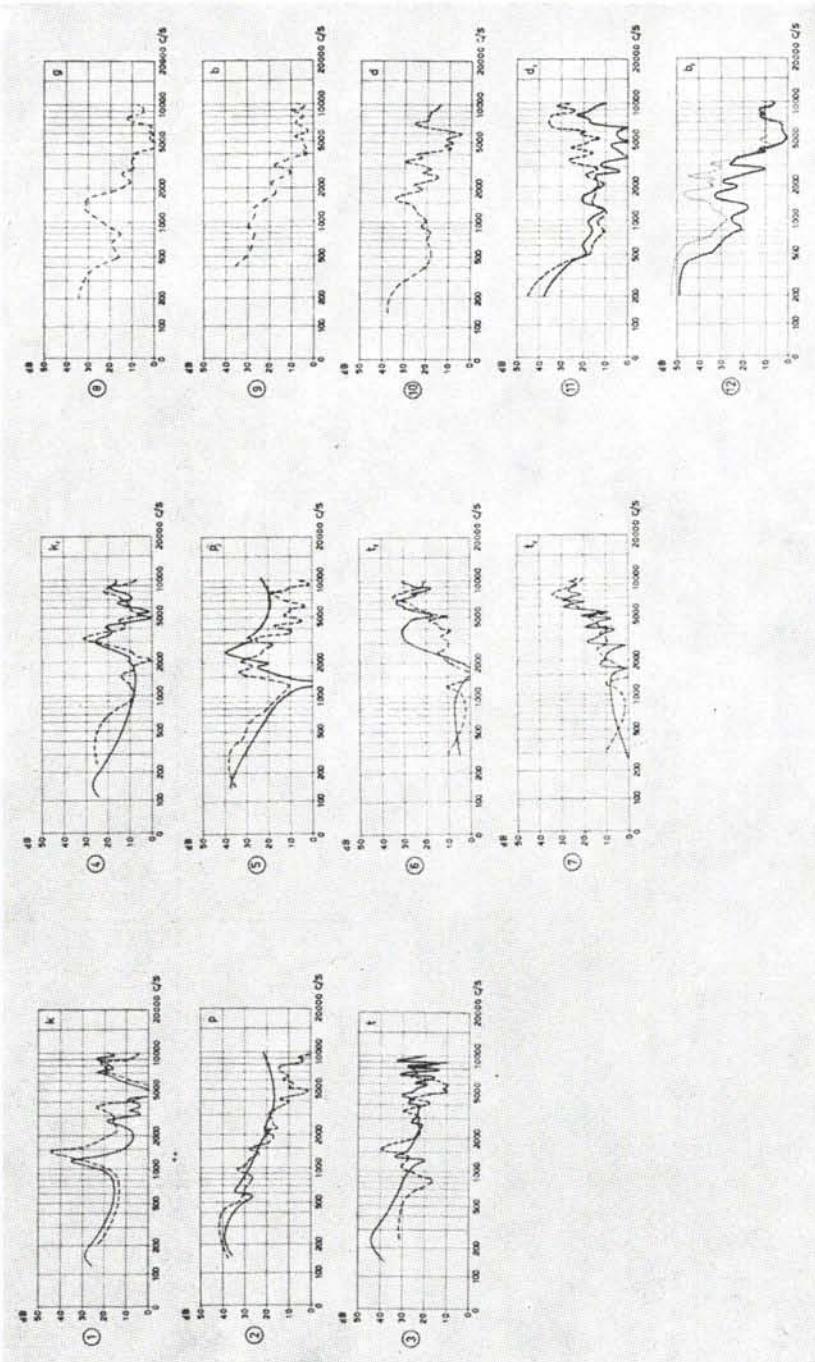


Fig. 2.6-10. Calculated and measured spectra of [k], [p], [t], [k.], [p.], [t.], and measured spectra of [g], [b], [d], and [d.], [b.]. All calculations were based on a $-6 \text{ dB/octave}, 0.25 \text{ } \mu\text{c}$ source except for [t.] where a source of constant spectrum level and resistance $2 \text{ } \mu\text{c}$ was adopted, as for fricatives. The time location of the spectral samples is the first 10 msec after the explosion except for [t.], diagrams 6 and 7, and [d.], diagram 11 where a fricative sound interval 20 msec after the explosion was measured. The dotted line curve for [b.], diagram 12 pertains to a sample 10 msec after the explosion which coincides with the first voiced period of the following vowel.

clear, and there is on the average one peak per 1000 c/s all the way up to 8000 c/s. The explanation of these aspiratory sound characteristics is, as previously mentioned, the more open articulation resulting in a more intimate coupling between the various parts of the vocal tract and possibly the development of one or more additional noise sources within the vocal tract, e.g., at the vocal cords.

The rise of the overall intensity from the explosion to a maximum value in the middle of the fricative interval is typical of strong and tensely articulated dental stops, but it is seldom seen in amplitude displays of palatal tense stops. It is probably due to a relatively slow movement of the tongue from closure to the optimal degree of opening. The same intensity level was noted for the subject's explosion and fricative [k]-intervals.

One important question raised by Halle et al. (1957) regards the effect of a rapid formant frequency variation. It is evident that if too rapid variations occur there will result a scattering of spectral energy in the form of a broadening, attenuation, and even split of spectral peaks, the effects increasing with the duration of the sample, i.e., decreasing with a broadening of the analyzing filter. This effect may account for some of the spectral leveling in the very first noise interval of labial and dental stops. In the latter category, the constriction resistance probably accounts for a larger part of the suppression of the formant structure. In palatals and velars the rate of change of the frequency of the main resonance originating from the mouth cavity will be small if the constriction area changes are moderate. As seen from the [k]-spectrum, the response of the vocal tract to the initial transient is characterized by a spectrum dominated by a single sharply defined peak which remains throughout the larger part of the [k]-noise.

2.63 *Idealized Models of Fricatives and Stops*

From the previous section it is clear that there exists a reasonable acoustic predictability of the spectral composition of fricatives, affricates, and stops, given the necessary physiological data concerning their production. This fairly detailed investigation will next be supplemented by a simplified presentation of the role of the vocal resonators for the establishment of the main characteristics of each of the three following classes of speech sounds originating from a noise source located at or near the articulatory constriction:

- A. Labials and labiodentals;
- B. Dentals;
- C. All other sounds produced with a more retracted place of articulation.

The three groups will be represented by the phonetic symbols [p t k] for stops and [f s ʃ] for fricatives. The symbol [ʃ] of these figures thus stands for any fricative that is opposed to both the labials and the dentals in terms of a greater front cavity volume. The possible distinctive role of a shift of source location at constant cavity configuration will also be commented on. An acoustic problem of specific interest is the

effect of palatalization on the spectral composition of the noise interval of a speech sound.

A maximal simplification of the cavity structure has been attempted in the idealized presentation of *Fig. 2.6-11*. Cavities posterior to the articulatory constriction have been omitted here but will be added in the later discussion. It should be observed that the acoustically pertinent boundary between the posterior and the anterior cavity system does not lie in the center of the constriction but towards its posterior end, ideally at a coordinate of maximal rate of cross-sectional area change. If the front cavity system is to include all elements of major importance for the sound characteristics, it is necessary to include in the calculations the total impedance of the constriction and thus its effective length.

The physical mechanism underlying the production of turbulent sounds, either in simplified mechanical models or in speech, is not sufficiently well understood for a rational prediction of the exact location and extent of a source along the length coordinate of the cavity system; see further *Section A.22*. As judged from the results of the calculations in the previous section it is, however, reasonable to associate the source with eddies, i.e., turbulent circulation effects, at places of maximal rate of area change preferably in narrow channels. In a model profile rectangularly shaped there may occur sources both at the inlet and at the outlet.

The exact location of a source within a narrow channel that can be considered short compared with wavelengths of interest, for instance within the labial constriction, is of no importance. Also, the location of the source within a narrow, not necessarily short channel posterior to a cavity of appreciable size, as in palatals or velars, is not very critical since the main shape of the sound spectrum will be conditioned by the fundamental resonance of the front cavity. Neither of these two conditions applies to the dental sounds. The representability of a calculated [s]- or [t]-spectrum is appreciably, though not altogether, dependent on the particular source location, the dental location providing the most distinct quality. For simplicity, the shallow cavity in front of the teeth has been neglected in the simplified cavity model. The effect of this and other configurative details affecting the [s]-calculations was exemplified in the previous section.

The equivalent network of the three models and one-dimensional pole-zero mode spectra specifying their filter functions, losses neglected, are included in *Fig. 2.6-11*. The specific properties of each of the three models are as follows:

A. *Labials*.—No resonator of any significance in front of or behind the major constriction. The source is close to the center of the constriction. The resulting mode spectrum of the model lacks any pole or zero, i.e., resonance or anti-resonance, within the range of frequencies 0-10000 c/s.

Labials and labiodentals should thus be characterized by a flat spectrum envelope, i.e., an even distribution of energy within the spectrum. This statement represents, however, an oversimplification for labiodentals. The fundamental resonance of the articulatory constriction plus the shallow cavity formed by the lips in front of the

upper incisors may be found as low as 6000-7000 c/s. The lip cavity resonance is also neglected in the following discussion of dentals.

B. *Dentals*.—No resonator of any appreciable size in front of the major constriction. A significant part of the constriction lies behind the source. The constriction channel provides a half-wavelength resonance at a frequency of the order of 5500 c/s. The high-pass characteristics in the form of an abrupt rise of the spectral level above 4000 c/s are accentuated by a zero at about 3500 c/s associated with the quarter-wavelength frequency of the narrow channel behind the source.

The main spectral peak of the dental sound at 4000-7000 c/s will, under actual speaking conditions, also be supported by the lip cavity resonance as previously discussed.

C. *Velars, palatals, etc.*—A resonator of comparatively large size in front of the main constriction. The fundamental resonance of this cavity is the main spectral determinant and occurs at a lower frequency than the main peak of dentals, even if the cavity length is the same as that of the resonating channel for dentals because of the approximate closure at the posterior end. In dentals, on the other hand, this end is openly terminated. In terms of the behavior of the [k]- and [t]-cavity models of Fig. 2.6-11, this is a difference between quarter-wavelength and half-wavelength resonance.

Another difference in the spectral determinants between a dental and a prepalatal sound is that the main peak of the latter generally comes close to the frequency of the main zero of the former.

In extreme simplification these resonator characteristics can be stated as:

- [f] No or very high resonance frequency;
- [s] High resonance frequency or rather high cutoff frequency of a high-pass filter;
- [š] Single resonance frequency lower than for [f] or [s].

The corresponding articulatory features are:

- [s] and [f] Absence of a large resonator in front of the place of articulation;
- [s] Articulatory constriction forming a short tube open at both ends;
- [š] Resonator of fairly large volume in front of the place of articulation.

A classification in terms of the source position, i.e., [s] and [f] as more fronted than [š], may also be used but not unambiguously, since cavity size differences alone are capable of maintaining the distinctions. A [š]-sound may be produced with a dental source at least in model experiments, and, as mentioned in Section 2.63, it is not impossible that the [s]-source may be positioned at the place of maximum narrowing in the tongue-teeth-ridge channel.

The effect on the simplified pole-zero mode spectrum of adding the back cavities is illustrated in Fig. 2.6-12. The total length from the bottom of the back cavity to the plane of radiation at the front has been limited to 15 cm. Both a neutral single tube back cavity and a twin-tube back cavity configuration simulating palatalization have been incorporated. For the study of palatalization within the labial models there are included two different widths of the front tube, representing the mouth cavity.

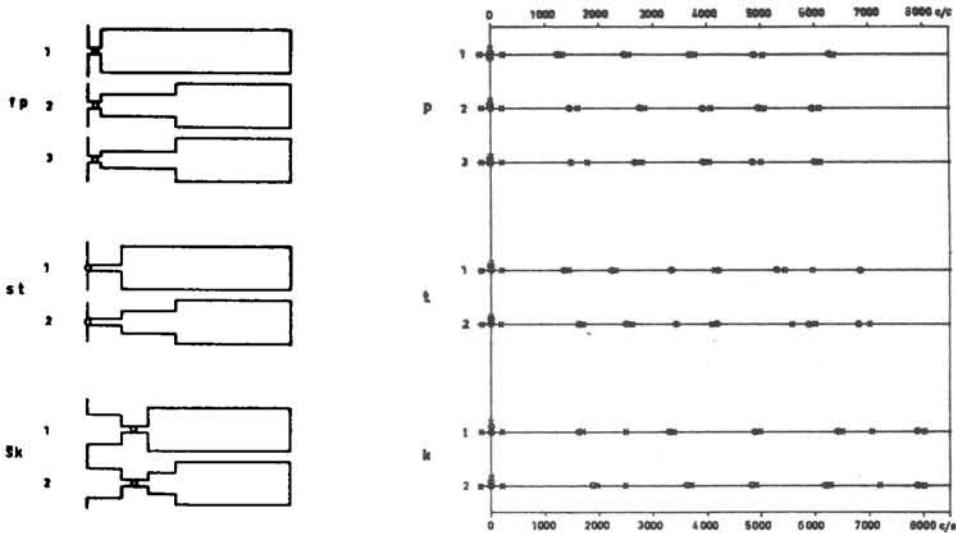


Fig. 2.6-12. The effect of various back cavity configurations on the pole-zero mode spectra of the labial, dental, and palatal models of Fig. 2.6-11. The back cavities are labeled 1, 2, 3, in increasing degree of similarity to palatalization.

The particular pole-zero mode spectra pertain to stops. Accordingly, the pole indicated by the cross just above the origin represents the -6 dB/octave step function source. The two circles at the origin represent the zero due to radiation and the first frequency of infinite impedance as seen backwards away from the source, that is, the first zero of the function $H_z(s)$ of Eq. A.34-4 and 6. The two crosses on both sides of the origin indicate the conjugate pair of poles from the fundamental resonance of the entire vocal tract, i.e., that responsible for the first formant, F_1 . Since F_1 of unvoiced sounds is normally more than critically damped, it is apparent that these two points will actually lie on the negative real axis, i.e., on the vertical axis of the simplified diagram. This is of no concern here. The net effect of this clustering of crosses and circles near the origin is 3 poles minus 2 zeros; there is a reduction to a single pole which conditions the -6 dB/octave average slope of the explosion interval of a stop sound. As a corollary of this description in terms of poles and zeros, it could be said that the very narrow but loss-less articulatory constriction behaves like a large inductance source impedance, since one of the two conjugate poles associated with F_1 is traded for the zero of $H_z(s)$ at the origin. The effect of constriction area changes on the intensity level of higher formants can, accordingly, be related either directly to the varying source impedance, Eq. 1.23-9, or to the frequency F_1 , whichever happens to be more convenient for the reasoning.

The mode spectra No. 1 of the [p]-, [t]-, and [k]-models contain approximately the same free poles and zeros as the corresponding mode spectra of these models without back cavities, shown in Fig. 2.6-11. In addition, there are pole-zero pairs representing

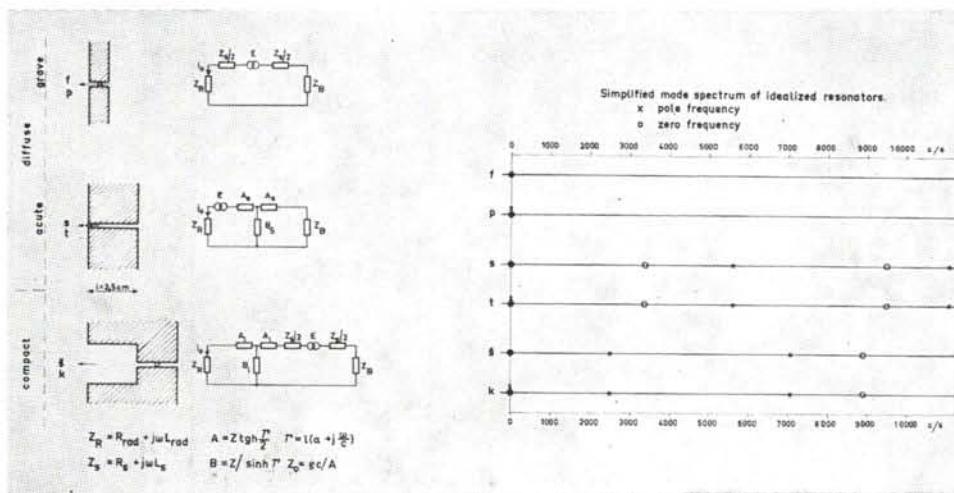


Fig. 2.6-11. Highly simplified cavity models, equivalent networks, and pole-zero mode spectra of labial, dental, and palatal stops or fricatives. Palatals have a more retracted place of articulation and a larger cavity in front of this place than dentals and labials. The source of labials and dentals is located at or near the front end, differences being insignificant. Dentals have a resonating channel of appreciable length behind the source as compared with labials.

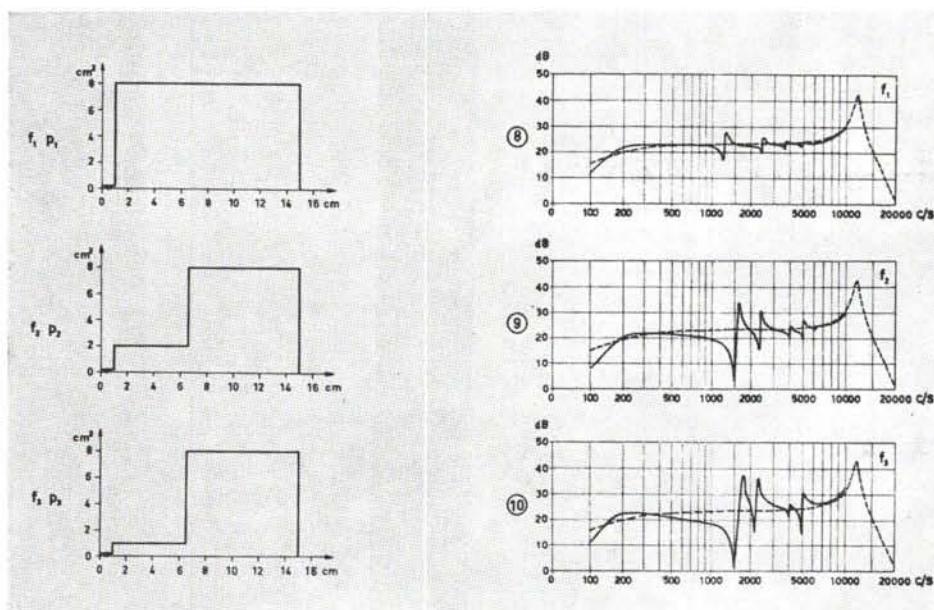


Fig. 2.6-13. Area functions and sound spectra of the labial models of Fig. 2.6-11 when excited by a constant spectrum level source, resistance $0.25 \text{ } \rho c$, located at 0.5 cm from the radiating end. Glottis resistance $5 \text{ } \rho c$. The emphasis of F_2 and F_3 and the zero just below F_2 are apparent at increasing degrees of palatalization.

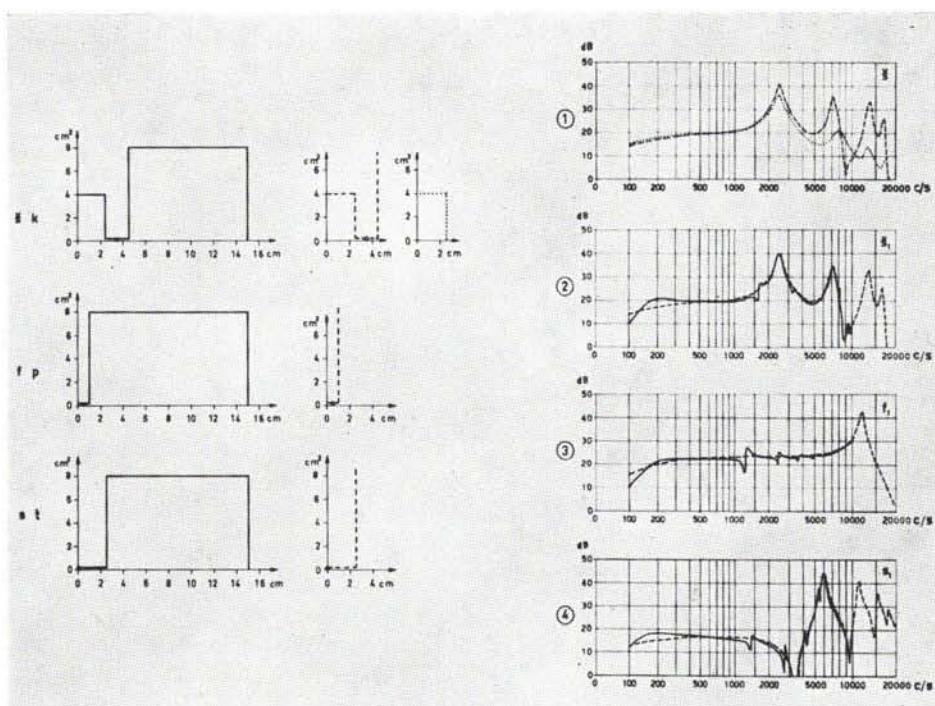


Fig. 2.6-14. The effects of adding a simple tube as a back cavity, supplementing the front cavity configurations of Fig. 2.6-11. For the palatal model the effect of removing the filtering contribution of the constriction channel is also shown.

Source data as in Fig. 2.6-13.

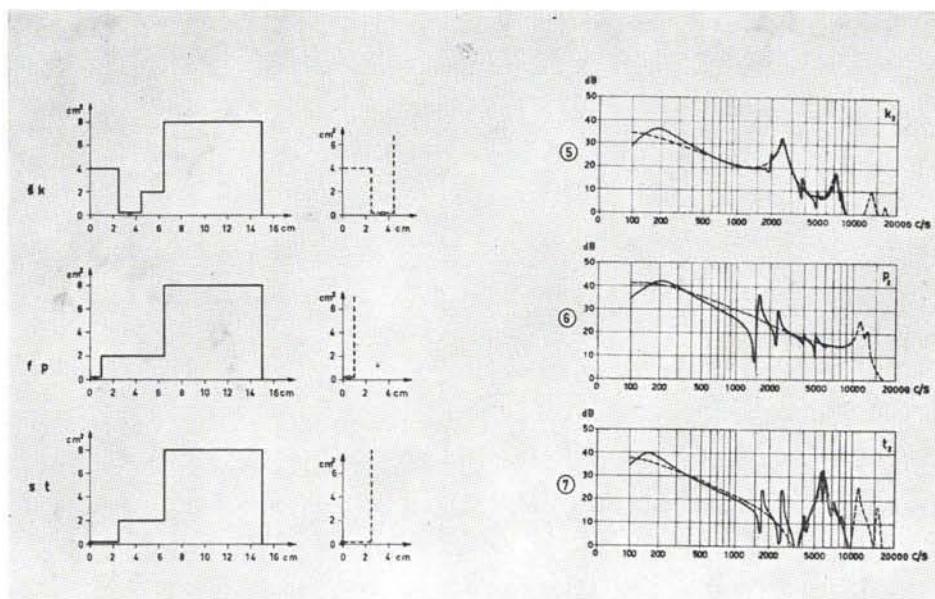


Fig. 2.6-15. The effects of adding a *palatalized* back cavity to the palatal, labial, and dental models of Fig. 2.6-11. Source spectrum -6 dB/octave .

the largely neutralized resonance structure of the back cavities. The effect of palatalization is invariably a rise in F_2 , as can be seen from a comparison of mode spectrum 2 with 1, or even more so, spectrum 3 with 1. Within the mode spectra of the labials there is also a successive separation of F_2 from its associated zero as the mouth tube is narrowed. This is also true for F_3 of the labials, and F_2 and F_3 of the dentals, though to a lesser extent. In general, this emphasis of F_2 and F_3 is largest when the impedance of the palatal mouth section becomes comparable with that of the source constriction, which implies that the latter section will be relatively short and wide, or the former long and narrow. The requirements for the formant structure to appear are thus similar to those for aspirative noise in general, as described in *Section A.22*.

The complete spectrum envelopes of continuant sounds generated from the three labial models are shown in *Fig. 2.6-13* together with the area function diagrams specifying the resonator dimensions. The broken line curves in the spectrum diagrams indicate the spectrum envelope of the sound with the effects of the back cavities disregarded. It is apparent that the residual formant pattern is superimposed fairly symmetrically around this basic envelope curve. The spectrum level is constant, on the average, owing to the flat spectrum source chosen for the continuant fricatives. It can be seen that F_2 rises in frequency from 1300 c/s in diagram 8 to 1700 c/s in diagram 10, representing the maximum degree of palatalization. The intensity level of the second formant, and also the third, rises 10 dB as a result of the associated pole-zero separation. The zero just below $F_2 = 1500$ c/s was found at 1300-1500 c/s in the measured and calculated spectra of our subject's [f], *Fig. 2.6-5*, and [p], *Fig. 2.6-10*, as previously discussed in detail. The detailed structure of the spectrum derived from the simplified model of *Fig. 2.6-12* actually compares better with that of the spoken sample, *Fig. 2.6-5*, than the associated calculation from the X-ray data. The explanation is obviously that the subject pressed his tongue too firmly against the palate during the X-ray exposure thus overemphasizing the narrowing of the palatal channel.

Fig. 2.6-14 contains the area functions and the corresponding spectra of the simplified fricatives with neutral, i.e., single tube, back cavity, and also the spectra of the same models exclusive of back cavity. In addition, diagram 1 provides a comparison of the effect of neglecting the finite filtering contribution from the source constriction of the palatal model. The effect on the main peak at 2500 c/s is negligible, but the next peak at 7000 c/s is shifted to 8000 c/s which is the three-quarter-wavelength resonance of the front cavity regarded as a tube closed at the back end and open at the front end. The shift is also associated with a decrease of the level of the second peak, associated with the stronger dependency of this resonance on the front cavity and thus the larger damping effect due to the large radiation resistance at these higher frequencies.

The additional effects of the back tube on the palatal, labial, and dental models, as illustrated by diagrams 2, 3, and 4, are basically the same as the ones previously discussed. The main peak of the palatal sound is supported by F_3 , and a very weak

remainder of F_2 may be seen at 1700 c/s. Because of the fairly low constriction impedance of 0.25 ρc chosen for all these calculations, it is possible to see the main zero of the dental sound at 3500 c/s. The superimposed formant ripple is of very small amplitude compared with the level of the main peak of the dental at 7000 c/s.

The effect of adding a palatalized back cavity to each of the three stops is shown in Fig. 2.6-15, where the spectrum curves have been constructed on the basis of a -6 dB/octave source slope. As expected, F_2 and F_3 of the labial and dental stops are emphasized as compared with the non-palatalized examples of Fig. 2.6-14. The effect of palatalization on the labial model was shown earlier in connection with Fig. 2.6-13. These results indicate that the relative prominence of F_2 and F_3 within the consonant noise in fricatives and stops might be a secondary cue for palatalization along with the influence by the palatalized sound on the formant transitions in adjacent vowels, as determined by the high F_2 - and F_3 -loci.

The effect on the palatal model of this idealized palatalization is negligible. In actual speech, however, the palatalization of [k] is associated with a decrease of the size of the front resonator so that the main resonance is shifted from an identity with F_2 in a velar [k] to an identity with F_3 or F_4 of a prepalatal [k]. Since the effective *mean* pitch of F_2 and higher formants of a back vowel is close to F_2 , but for an extreme front vowel close to F_3 or F_4 , it is apparent that the central position occupied by the main resonance of a velar or palatal stop or fricative is linked closer to the effective higher formant pitch of the following vowel, especially its transitional interval, than to its F_2 . The physiological invariant of this pitch is simply the cavity in front of the articulatory constriction, the resonance frequency of which is determined by the cavity volume and degree of lip-rounding. The latter variable accounts for an appreciable part of the F_2 -locus shift of [k] from [ka] to [ko].

Typical transitions of the F-patterns from the closed to a completely open interval of palatal, labial, and dental stops are illustrated by the models and associated stylized spectrograms of Fig. 2.6-16, showing formant frequencies F_1 , F_2 , and F_3 . The models are similar to those previously dealt with except for the addition of a larynx tube and the more retracted place of articulation for the palatals. All three models are designed so as to reach the same front vowel configuration at the end of the successive area increase at the place of articulation. Stepwise area variations are indicated by broken lines on the area functions of the models. The horizontal axis of the stylized spectrograms is calibrated in cross-sectional area of the articulatory constriction as a time dependent parameter. The F-pattern can thus be studied as a function of the degree of articulatory opening.

At complete articulatory closure, F_2 and F_3 of the palatal model meet at 1900 c/s. This is due to the coincidence of the fundamental resonance frequency of the front cavity and the half-wavelength resonance of the back cavity. In the labial model both F_2 and F_3 start low, F_2 at 1400 c/s and F_3 at 2200 c/s. The closed state F-pattern of the dental model is close to that of the adjacent vowel, $F_2 = 1700$ c/s and $F_3 =$

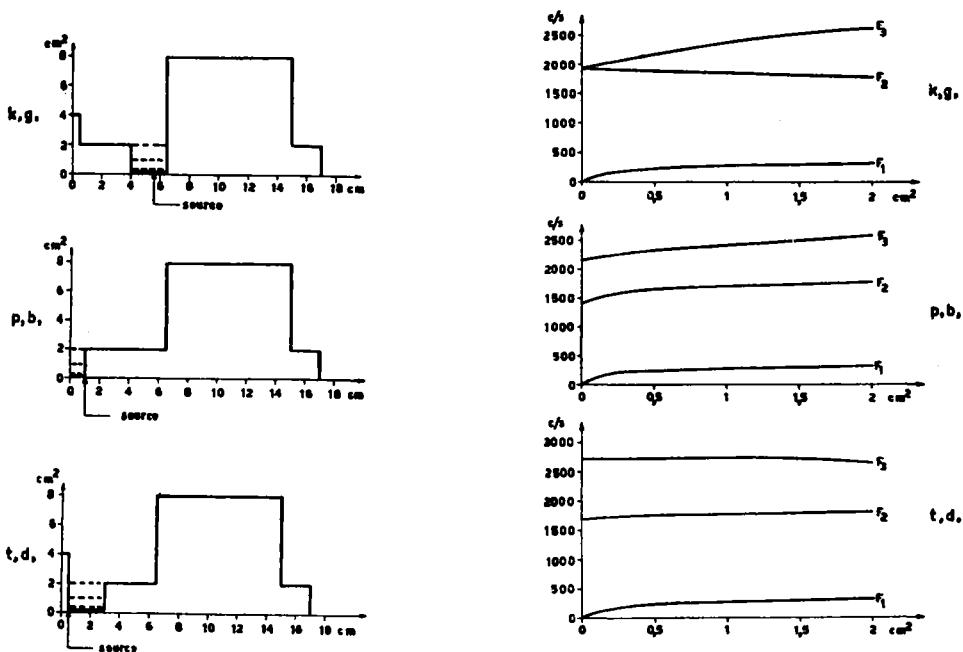


Fig. 2.6-16. F-patterns of a palatal, a labial, and a dental model as a function of the cross-sectional area at the place of articulation. The three vocal tract models differ in terms of the constricted section only.

2700 c/s. The limiting F-pattern $F_1 = 400$ c/s, $F_2 = 1750$ c/s, and $F_3 = 2600$ c/s of the open state is close to that of a slightly rounded, half-open front vowel.

From comparisons with similar transitional patterns in real speech as studied from spectrograms, it may be concluded that the rate of cross-sectional area variation following the explosion is of the order of 5 to 20 cm^2/sec , which implies that the horizontal scale of the stylized spectrograms is expanded 2 to 8 times. A major part of the F_2 -transition of the labial stop is completed when the lip-opening has reached 0.2 cm^2 area. In the following process of opening, the lip inductance is no longer large compared with that of the palatal tongue passage and has no longer any considerable influence on the tuning of the pharynx half-wavelength resonance which is largely responsible for F_2 . The progressing lip-opening will, however, continue to cause a substantial shift up in frequency of the mouth-section resonance and thus of F_3 . The F_2F_3 -divergency following the [k]-, [g]-release continues all the way to the following vowel. These examples show that the rules governing the relations between degrees of articulatory opening and the extent to which a transition has progressed are not simple. The speed of the articulatory movements will also differ. A labial opening generally shows a greater acceleration than a lowering of the tongue from a palatal or a velar closure. It is thus often seen that the major part of a labial transition is completed within about 15 msec.

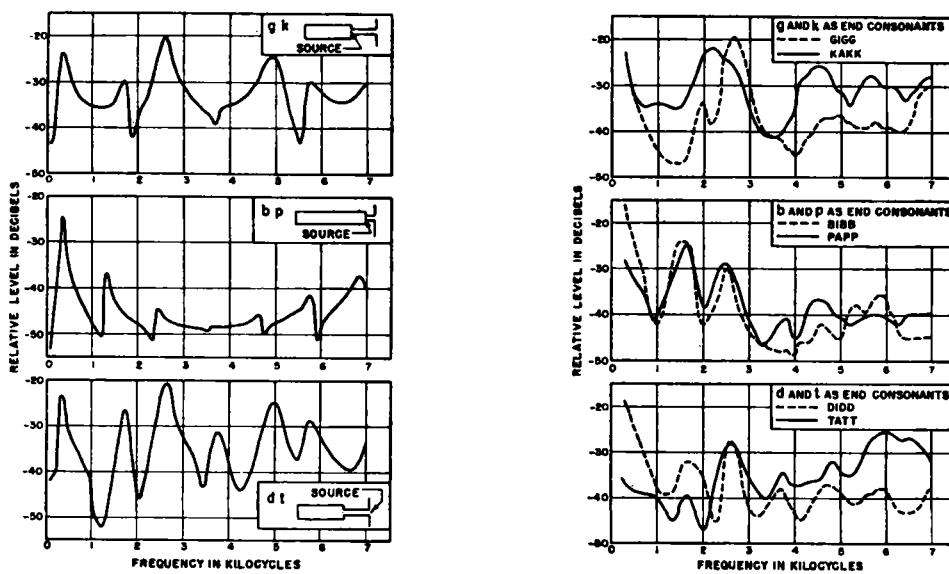


Fig. 2.6-17. Numerically calculated spectra of stops from twin-tube models; see Section 1.54. The palatal and the dental model are identical with regard to configuration, the difference being the source location. Source spectrum -6 dB/octave and no source resistance. Distributed losses in the twin-tube. Measured spectra of Swedish stops are included for comparison.

A shift of the location of a source within the vocal tract model, the cavity configuration held constant, will obviously not change any resonance frequencies as long as the constriction impedance remains the same. The resistive parts of a constriction impedance as well as possible noise sources, are basically conditioned by the direct durrent airflow, and the change of source location thus remains a theoretical operation performed to study the function of the model under various assumptions. It may be questioned whether the changes in formant amplitudes caused by the change in the zero function, accompanying a shift of source at constant cavity configuration, may cause a change in spectrum shape large enough to be judged phonemically distinctive. Experiments on the perception of synthetic speech are needed for a conclusive answer, but the general conditions appear to be favorable.

One example of such a minimal change in the process of synthesizing stop sounds is included in the calculations on twin-tube models, Fig. 2.6-17. These spectral curves were calculated numerically according to the formulas described in Section 1.54 and have been published earlier, Fant (1950b). No other damping than that inherent in the transmission line network of each section and that due to radiation was included. Additional source resistances were thus neglected. The [g]-, [k]- and the [d]-, [t]-models are identical in configuration and represent a rather high degree of articulatory opening. The zeros of the [g]-, [k]-model occur at the frequencies $nc/2l_2 = n \cdot 1850 \text{ c/s}$,

where $l_2 = 9.5 \text{ cm}$ is the back cavity length. The first of these occurs at zero frequency and the next, at 1850 c/s, has the usual neutralizing effect on the second formant. The third formant at 2500 c/s, constituting the major peak, is a half-wavelength resonance of the front section, and the 5000 c/s peak represents a full-wavelength resonance of this section. These formants are also fairly prominent in the [d]-, [t]-spectrum but are no longer dominating owing to the intermediate formants which alternate with zeros. Because of the extreme front location of the source in the dental model, the average spacing of poles and zeros must be the same. This statistical feature, as well as several of the specific dental and palatal spectrum shape features, may also be seen in the respective measured spectra to the right in the figure. These pertain to a mixed fricative-aspirative phase of Swedish stops in final position. The labial model is the simplest possible but retains some of the features of the spoken data, such as the successively decreasing amplitudes of the first three formants and the higher formant region above 5000 c/s.

2.64 Conclusions Regarding Source Characteristics of Fricatives and Stops

The calculations in the preceding sections have shown that in spite of the incomplete physical basis for describing turbulent sound generation, the equivalent circuit representation provides an acceptable phonetic basis for the theory of fricatives and stops. The essentials of the sound production may thus be specified in terms of source and filter characteristics just as for vowels, and in some instances the predictability of the spectral characteristics has been equally good. There remain some uncertainties concerning the location and spectral distribution of turbulent sound sources.

The especially good fit between the calculated and spoken [x]-spectra may be due to a fairly accurate localization of the source at the constriction center during spurious closure and opening movements of the articulators caused by the Bernouilli force of the outflowing air. The -6 dB/octave source found to be representative for [x] up to 3000 c/s may thus originate from the impact of these spurious air quanta.

There is probably a dental source besides the tongue constriction source of [š], and it is possible that this is the case also for [šč]. The location of the source for labiodentals seems to be close to the upper front teeth, but if there is a strong superimposed palatalization and a less contracted lip-opening, there may occur a competing palatal source. A secondary palatal or glottal source is not necessary for the appearance of the F-pattern within the spectrum, since its determinant is the degree of articulatory opening, but it is hard to state offhand from a spectrogram whether one of these additional sources is present. The glottal source as well as a labial source would emphasize F_1 , and a palatal source would tend to emphasize F_3 or F_4 and to neutralize F_2 unless distributed over a considerable length of the palatal section. However, what is stated above for palatal sources also holds, but to a lesser extent, for the combination of a labial source with a palatal constriction. In this case there is clear positive indication of the labial source position from the zero at approximately 1300 c/s.

The source impedance is strictly a property of the filter function, but its resistive part is dependent on the direct current airflow which, under turbulent conditions, is proportional to the flow velocity, as shown in *Section A.22*. The resistance to inductance ratio R_s/L_s of the constriction will apparently increase at increasing flow. The primary reference for the source is the source voltage $E_s(\omega)$ of the equivalent circuit, but it is apparent that the volume velocity current supplied by the source

$$U_s(\omega) = E_s(\omega)/(R_s + j\omega L_s) \quad (2.6-1)$$

and thus the pressure in the radiated wave, are also dependent on the constriction impedance. As shown in *Section A.36*, A.3 the cutoff frequency of the constriction impedance, $F_c = R_s/2\pi L_s$, will under turbulent conditions be approximately independent of the constriction area,

$$F_c = v/2\pi l, \quad (2.6-2)$$

where v is the air velocity and l the constriction length. Under normal conditions for the production of dental or palatal fricatives, $l = 2.5$ cm and $v = 3000$ cm/s, the cutoff frequency is of the order of 200 c/s but closer to 500 c/s for the shorter labial passages. As long as this frequency is higher than F_1 , the vocal tract is more than critically damped with regard to the first formant as presumed in this analysis.

When deriving the source spectrum as the difference between the spectral level curve of the spoken sound and that of the vocal tract filter function, there enters at low frequencies the uncertainty regarding the representability of the source resistance chosen for the vocal tract model. If this resistance has been underestimated or the inductance overestimated in the calculations, there will result an underestimation of the spectral level of the source voltage at low frequencies. The particular R/L -values adopted in *Sections 2.62-2.63* are correct only in the order of magnitude and may be in error by as much as a factor of 2. It is therefore not possible to extend the derivation of source spectra in this work to frequencies below 300-500 c/s. Above 5000 c/s, the vocal tract dimensions perpendicular to the direction of the airflow are not always small enough that spectrum level losses through transverse resonant modes might be disregarded. This is less of a problem with dentals than with retroflex palatals. These effects will be ignored in the following discussion.

Instead of stating a separate source spectrum curve for each of the sounds, it is considered more appropriate to summarize the results as follows:

Spectrum envelope of the source

Applicable to:

pressure (voltage): — 6 dB/octave,
i.e., integrated white noise for
continuant sources and a step
function for transient sources.

- | | |
|-------------------|--|
| [x] | 300-4000 c/s |
| [f] | 800-10000 c/s. A better fit is, however, obtained by a source spectrum envelope of
—3 dB/octave; |
| [f _s] | 400-8000 c/s. The explosion phase as well as
the noise interval of maximum intensity within
[k k, p p] 300-8000 c/s; |

<i>0 dB/octave,</i> i.e., constant spectrum level ("flat" spectrum)	[§]	300-6000 c/s assuming an apical source, except for the uncertainty in the region 300-800 c/s;
	[§]	300-2000 c/s assuming a dental source. This is less probable and must be combined with a -12 dB/octave source spectrum fall above 2000 c/s;
	[§§]	300-3000 c/s followed by a -6 dB/octave fall above 3000 c/s if a dental source is assumed;
	[§§]	1500-8000 c/s assuming a constriction source;
	[s]	800-4000 c/s followed by a -6 dB/octave drop above 4000 c/s;
	[s.]	1000-8000 c/s;
	[t] and [t.]	500-8000 c/s for the fricative interval.

It should be observed that these data pertain to specific samples of spoken utterances only; see the spectrographic presentation in the *Appendix*. The tolerances of the estimated spectral slopes are rather great and at the most 3 dB/octave, the greatest source of error being unavoidable differences in the location of spectral peaks, owing to differences between the natural speech and the X-ray session articulation and to inaccuracies in the estimation of the vocal tract dimensions and the resistive elements of the equivalent networks.

There do not seem to be very great differences between the average spectral slopes of the explosion segments of voiced stops and those of the corresponding unvoiced stops. Halle, Hughes, and Radley (1957) found less steeply falling spectrum envelopes of the tense English stops as compared with the lax stops, and this tendency is also apparent from the data on Swedish stops, *Fig. 2.6-17*.

The data for palatal and dental continuant fricatives compare reasonably well with that found by Heinz (1957a) for the source spectrum of mechanical models of 1 cm in length and 0.03 cm² in constriction area. He obtained an essentially constant spectrum envelope between 1500-4000 c/s and approximately 12 dB/octave fall above 4000 c/s and below 1500 c/s.⁵

The question as to what are the optimal constriction dimensions for the generation of the various turbulent sounds has not yet been given a satisfactory answer. An area of 0.1 cm² appears to be of the correct order of magnitude according to preliminary investigations, and it seems probable that any constriction impedance behaves mainly as a large inductance at the frequency of the main spectral peak. Thus it can be anticipated that the area giving maximum source pressure, i.e., voltage in the analog network, will be smaller than that providing maximum sound output.

The typically much lower overall intensity of an [f] as compared with an [s] or [§] can, to an appreciable extent, be explained by the lack of front cavity resonance for the [f] in the frequency range below 8000 c/s or, if this resonance is present, by the lower Q combined with the more negative slope of the source spectrum. The former of these two effects may be seen from *Fig. 2.6-14*. The latter is one aspect of the less efficient noise generation associated with a laterally spread but very narrow opening,

⁵ More recent data from M.I.T. suggest an even more flat source spectrum (personal communication from K. N. Stevens).

as shown in *Section A.22*. It also seems probable that the obstacle effect of the teeth is an important factor making dentals and labiodentals more intense and more high frequency emphasized than labials. Further quantitative studies of noise levels, noise spectra, flow conditions, differential resistances, and the vocal tract configuration during human production of turbulent sounds are needed in addition to model experiments.

PART III

SUMMARY

3.1 SEGMENTATION AND SPECIFICATION

The work reported in this book has mainly been concerned with a study of the theory of speech production with an emphasis on the analytical relations between articulation and speech wave and the acoustic theory of resonator systems approximating the vocal tract. In addition, some space is devoted in the *Appendix* to instruments and methods for analysis and description of speech waves, mainly for the purpose of illustrating the techniques and for supplementing the calculations with data from human speech.

This survey is general in nature, and references to Russian speech sounds are made only in connection with the calculations from the X-ray data of a Russian subject. It is the purpose of the following section to summarize and discuss some of the results of specific interest to phonetics.

The problem of segmentation of the visible record of the speech wave has been touched upon in *Section 1.12*. It is evident that the acoustic boundaries found from a visible record of the speech wave belong to a purely phonetic transcription and may not be directly interpreted as phoneme boundaries.¹ These phonetic (or acoustic) boundaries are best observed from spectrograms and are apparent from temporally localized changes in the sound pattern. They divide the speech utterance into units of the dimension of speech sound or smaller. Since any speech sound may be regarded as the filtered product of a source, it is practical to associate the boundaries with a change of the type of source, or of the intensity of the source, or with a rapid change in the vocal tract filter function, or with simultaneous changes in both filter and source.

Examples of sound units, logically constituting parts of speech sounds, are the voiced or voiceless sound interval corresponding to articulatory closure and the following noise interval of a stop sound. When length measures of stops are made, the need for well-defined fundamental units will be especially apparent. In a pre-vocalic position, the duration of the noise interval may be defined genetically as the distance in time from the break of contact to the onset of the vocal cord vibration.

¹ See for instance Truby (1959).

If voice and burst coincide in time, and if further, there exists a tendency of frication as in the Russian [d,], the burst of noise must be defined from its visible duration in the spectrogram. The noise interval of the stop may be divided up into the successive and partially overlapping intervals of explosion, frication, and aspiration;² see *Sections 1.11* and *A.22*. It is, however, only rarely possible to state exact boundaries of these units. Other examples of a division into units smaller than speech sounds are the possible recognition of a sound boundary at the instance of voice onset within a liquid, the first part of which has assimilated the voicelessness of a preceding consonant. These smaller units are the *natural* building blocks of connected speech, and it is necessary to adopt certain conventions when assigning them to specific speech sounds. As argued by Fischer-Jørgensen (1954, 1956), the aspiratory interval of a fortis stop sound could be assigned to the following vowel in terms of its articulation (i.e., filter function), but conventionally it is a part of the stop burst, and this latter choice seems practical for describing the tense-lax distinction within stops—the tense stops possessing the longer consonantal burst, aspiration included; see further *Fig. A.2-2*, and the discussion at the end of this section.

A transcription of speech on the structural level as a sequence of phonemes is not upset by the fact that the speech wave stimuli contributing to the listener's identification of any phoneme may be spread out over several successive minimal units of the speech wave (Halle, 1956). The only trouble occurs if investigators insist on measuring or defining phoneme durations. Of course, it is possible to accept a convention whereby an initial phonetic transcription is simply transferred to the structural plane by the necessary broadening of the transcription. By this logical operation the phonemic overlap in time is avoided, but it still exists in the code of the physical speech signals in the sense that one interval may carry information relevant to two successive phonemes.

The specification of the speech wave in terms of quantitative data is not a simple matter. Any choice of a reference system should be preceded by an evaluation of the accuracy needed in order to achieve optimal compromise between the conflicting requirements of shortness and completeness. The requirements differ with the particular purpose. One extreme is the study of the signal data without reference to message units, for instance for a detailed study of speech waves and their relation to articulation. Complete spectrum envelopes at successive samples of the speech wave will then be needed. The other extreme is the minimum redundancy formulation of acoustic correlates to distinctive features, e.g., as attempted by Jakobson et al. (1952), and by Jakobson and Halle (1956). The description of the acoustic properties of tense phonemes in words such as *large spread of energy in the spectrum and in time* is rather vague and of no use for practical applications unless the concept of *energy spread* is defined quantitatively, e.g., in terms of formulas of the type developed by Halle (1954); see also Hughes and Halle (1956).

² In some authors' terminology, e.g., Schatz (1954), the term *burst* signifies explosion + frication. Others utilize the term burst for the total noise interval of the stop.

3.2 THE RELATIONS BETWEEN THE F-PATTERN AND ARTICULATION

When collecting and comparing primary data from speech wave records, it is recommended to supplement quantitative data on the spectral distribution of the energy with a statement of the F-pattern; see *Section 1.13*. Dealing with vowels, the F-pattern is simply the formant frequencies with specific restrictions needed for nasal coupling. More generally, the F-pattern is the set of resonance frequencies of the vocal tract. It conditions the essence of a vowel spectrum and also some aspects of a consonant spectrum and serves as a good correlate to articulatory positions. The translation from speech wave back to articulation is to some extent restricted by the existence of compensatory forms of articulation. Only some general aspects of such compensatory forms have been discussed here; see for instance *Sections 1.41-1.43*. A deeper insight into the potentialities of this aspect of the physiological interpretation of spectrograms must rely on extensive correlative work involving X-ray moving film and spectrography.

The F-pattern, specified continuously along the speech utterance, serves as a generalized *locus* or *hub* concept since it includes not only F_2 but also F_1 and F_3 and those higher resonance frequencies which may be needed in any special case. The formant transitions in vocalic intervals adjacent to a consonant are conditioned by the F-pattern at the interval of maximum closure of the consonant together with the F-patterns of the associated vowels.

The accumulated data on speech available in the phonetic literature are mainly from the area of speech production and are generally presented in the form of articulatory positions and only to a minor extent supplemented by X-ray photographs and palatograms. For an analytical prediction of speech wave data from data on the production, there is needed, however, a complete three-dimensional mapping of the vocal cavities. There exists a correlation between these dimensions and the articulatory positions, but it should not be anticipated that a place of articulation in the traditional phonetic aspects always coincides with a pertinent cavity boundary within the vocal tract; see *Section 2.32*.

An approximate specification of the cavity data by means of a three-parameter

system as developed by Stevens and House (1955, 1956) and further made use of here in an alternative form, *Section 1.43*, is, from an acoustical point of view, more rational than the traditional highest-point-of-the-tongue reference for classifying vowels. It is also useful for relating consonant F-patterns to vocal tract data. The three parameters are: (1) the place and (2) the cross-sectional area of the tongue constriction and (3) the degree of lip-opening. These parameters may be utilized in various ways. When dealing with labials, the two parameters related to the tongue section may be utilized for a specification of the secondary, i.e., internal, place of articulation. Several diagrams relating the F-pattern to the three parameters of such models are given in *Section 1.43*. Interpolation can be used for specific estimates, and if the overall length of a specific speaker's vocal tract does not fit that of the model, it will be sufficient to apply a linear scale factor, all formant frequencies varying in inverse proportion to the overall length.

One implication of the three-parameter description of articulatory data is that the vowel [i] has the same degree of opening at the effective tongue pass as [ɑ] and that sounds may be ordered in terms of the location of this pass, e.g., [æ], [ɑ], [o], [u], [i], [e], [ɪ]. If a sound is very openly articulated with regard to the effective tongue pass, the F-pattern is quite close to that of the neutral vowel, $F_1 = 500 \text{ c/s}$, $F_2 = 1500 \text{ c/s}$, $F_3 = 2500 \text{ c/s}$, etc., irrespective of the location of the tongue pass, and a specification of the latter then loses its significance. If there exists a definite tongue pass, the optimal place of the pass for a high F_1 is in the back part of the model, i.e., in the pharynx, and the optimal place for a high F_2 is at the middle of the front half of the model, i.e., in the palatal region. F_1 decreases and F_2 increases if a tongue pass located in the front half of the model is narrowed, and the effect is the reverse within certain limits when the tongue pass is located in the posterior half of the model. There is an optimum tongue pass area providing maximum F_1 , but F_2 increases consistently with increasing tongue pass opening. Narrowing or lengthening of the lip passage decreases all formant frequencies of interest.

In general, all parts of the vocal tract cavity system contribute somewhat to the tuning of all resonance frequencies. Some of these dependencies may be directly observed from an inspection of the nomograms of the three-parameter models. If a shift of the tongue pass in a front direction, *ceteris paribus*, causes a formant frequency to rise, it may be concluded that it is more influenced by the part in front of the tongue pass than by the back cavity. Formants thus change their cavity affiliations at the maxima and minima of these curves, as pointed out by Stevens and House (1956). A maximum of one of the first two formants is located at a tongue pass coordinate close to a minimum of the next formant.

A special investigation has been devoted to the applicability of the double Helmholtz resonator model. It may be used for a prediction of F_1 and F_2 of back vowels and the vowel [i]. The single Helmholtz resonator may be applied to the whole of the vocal tract for calculating F_1 of front vowels. The front cavity of [i] conditions F_3 much more than F_2 , and F_2 of the Russian [i], analyzed here, is definitely a half-wavelength resonance of the back cavity. A detailed study of the coefficients by which

each formant frequency is related to the front and back cavities and to the lip and tongue passes has been undertaken in *Section 2.32*, and the results are compared with those predictable from the Helmholtz resonator theory. It is found, using the electrical line analog *LEA*, that in back vowels the front and back cavities both contribute appreciably to F_1 and F_2 with some emphasis on the generally postulated affiliation of F_1 to back cavity and F_2 to front cavity. On the other hand, for the vowel [u] the lip orifice will be definitely more affiliated with F_1 than with F_2 , and the reverse is true of the tongue pass. If in addition the cavity volumes are taken into account, it may be concluded that F_1 of [u] is mainly influenced by the back cavity and the lip orifice, and F_2 by the front cavity and the tongue pass. However, the sum of the front cavity and front orifice contribution to F_1 of [u] is larger than the sum of their contributions to F_2 , and the sum of the back cavity plus back orifice contribution to F_2 of [u] is larger than their contribution to F_1 . In this respect, F_2 of [u] is a back resonator formant and F_1 a front resonator formant.

As regards the analyzed vowels, the classical affiliations: F_1 -back-cavity and F_2 -front-cavity, were found to apply well only to the vowel [i]. In the absence of any marked tongue constriction within the vocal tract, each formant is equally dependent on each part of the vocal tract, and for half-open front vowels there will be a considerable influence from both front and back cavities to F_2 and F_3 , the F_2 -front-cavity affiliation being greater at tongue locations posterior to the palatal coordinate providing maximum F_2 . The effect of an increased lip-rounding on the cavity formant affiliations is similar to that of a retraction of the place of the tongue pass. That is, at tongue pass locations anterior to the palatal F_2 -maximum, an added lip-rounding may shift the front cavity affiliation from F_3 to F_2 . At a pharyngeal place of articulation, the added lip-rounding will shift the front cavity influence so that the F_2 -front-cavity affiliation will be decreased and the F_1 -front-cavity affiliation will be increased.

3.3 SOME ASPECTS OF THE THEORY OF DISTINCTIVE FEATURES

Some aspects of the theory of distinctive features will next be taken up for discussion. Ideally, all features that are independently commutable should be independent in terms of their acoustic correlates. This ideal, which is by no means necessary for the linguistic analysis, cannot be reached in a vowel system based on three or more non-identical features since vowels can be fairly well synthesized by two formants only, the higher formant, of frequency F_{2e} , serving as a perceptive substitute for the second and higher formants of naturally produced front vowels and corresponding to F_2 alone in back vowels. Vowel quality is thus basically related to the two variables F_1 and F_{2e} .

Of the three main vowel features, defined by Jakobson et al. (1952), it is feasible to associate increasing compactness with F_1 and decreasing gravity with the effective upper formant frequency F_{2e} or to F_2 alone. Decreasing flattening (lip-rounding) may be associated with $F_1 + F_{2e}$. This choice makes the compactness feature acoustically independent of the gravity feature. A closing or prolongation of the lips causes, ceteris paribus, a decrease of both F_1 and F_{2e} , which is a shift towards a less compact and more grave vowel in the light of the definitions above. On the structural level these interdependencies may be ignored, but they upset the orthogonality principle on the level of the sound substance. This type of complication is common in speech analysis, but it need not be very disturbing once the interrelations are known.

The theory of distinctive features is not only a powerful tool in the hands of the structural linguist, whose main interest is the construction of a system, free of redundancy, but it is also of basic importance in the statistical measurement of phonemic information and in technical applications, e.g., as a useful principle of designing mechanical speech writers. The minimum redundancy principle is, however, not of the same importance for the engineer or the phonetician. First of all, engineers need some redundancy in order to increase the reliability of the automatic recognition process. Secondly, it is not always possible by means of some mechanical operation to remove redundant aspects of the speech wave without losing at the same time a part of the information about the significant aspects.

Alternative solutions or formulations of the acoustic counterpart of a feature or even alternative choices of features are possible so that the system chosen by the structural linguist may not be the same as that adopted by the engineer or phonetician. When phonemes are compared in minimal pairs, it is sufficient to use F_2 or F_{2e} as the criterion of gravity. However, if the whole series of back vowels within a language shall be opposed to all front vowels by means of a single acoustic criterion, it is more convenient to utilize $F_2 - F_1$ or $F_{2e} - F_1$ as a measure,¹ and a corresponding dividing line may be drawn in the F_1 -versus F_2 - or F_{2e} -diagram. Similarly, a critical $F_1 + F_{2e}$ value may be stated for separating all rounded front vowels from all unrounded front vowels.

In view of the potentialities of the $F_2 - F_1$ - and $F_{2e} + F_1$ -parameters it may be considered sufficient to adopt these two for the basic reference frame, the $F_2 - F_1$ -parameter in addition being utilized for separating the various unrounded front vowels, and the $F_2 + F_1$ -parameter being utilized also for separating the various back vowels.

Two binary features of *spectral spread*, defined from the $F_2 - F_1$ -parameter, could thus replace gravity and compactness, and include aspects of both features. It may be observed that $F_2 - F_1$ has a more simple relation to the three-parameter vocal tract model than F_1 or F_2 alone. The two polar extremes of unrounded vowels are the mid-palatal [i] and the pharyngeal [ɑ], both produced with narrow tongue constriction. The two middle terms are either half-open and front vowels or non-closed front vowels and non-optimal back vowels, the latter being more open with regard to the width of the pharyngeal pass, possibly in combination with a laryngeal narrowing ([æ] versus [ɑ]) or in combination with a more fronted tongue pass location (compare [ü] and [u], [ö] and [o]). It may be observed that the step from narrow to wide back tongue articulation is mainly associated with an increase in F_2 in conformity with the established grave—acute dimension.

On the perceptual level, $F_2 - F_1$ is a criterion of the two- versus one-formant structure of vowels. Similarly, the $F_{2e} + F_1$ -parameter ought to reflect the *mean* timbre pitch of vowels.² The main articulatory correlate to $F_{2e} + F_1$ is lip-rounding but tongue-backing and mouth volume expansion contribute also.

The simplest form of the mechanized analysis is thus one where series of phonemes are opposed to each other whenever possible. This is more economical than designing the machine so that it readjusts its identification criteria on the basis of every successive decision made in the analysis of a sound interval.

The theory of distinctive features is sometimes opposed to the concept of the multiplicity of cues that may enter the acoustic description of a feature. A subdivision of this kind may be useful when there exists a variety of contexts where the feature

¹ ($F_2 - F_1$, e.g., signifies F_2 minus F_1 , etc.).—This classification scheme—see the formant data of Fant (1952, 1957)—is introduced here merely for the sake of provoking a discussion. Its linguistic merits need to be evaluated further.

² Playing a tape back at half-speed changes [a] to [o] and [e] to [ö], etc.

may occur, including the effects of speaker variability, sequential constraints, and the particular set of features coexisting within the sound interval to be studied. The pertinent question is whether the feature need be acknowledged at all as the cues have to be made use of in practical phonetic applications anyhow and these vary with the context.

There are two typical cases to consider, depending on whether the separate cues are found always, or nearly always, to coexist or are on the whole mutually exclusive. It does not seem fruitful to follow the common approach of singling out all but one of coexisting cues on the motivation that this cue has been observed to be capable of functioning alone and thus represents minimum necessary conditions. Listening tests on systematically varied synthetic speech are of course highly valuable for determining the relative importance of the various cues, but it is desirable, whatever the results of such tests, to attempt a formulation of the acoustic manifestation of the feature that all the cues are logically contained in a single concise statement which ideally should not be longer than the formulation of the cue. An example close at hand is stress. In some languages duration has proved to be an important stress correlate; see Fry (1955). In such cases duration and stress may be combined by means of their product, which is of the dimension energy and a more realistic stress correlate than intensity alone and equally simple; see further Fant (1957). In certain cases a parallel reasoning can be applied to the tenseness feature.

The difficulties involved in the formulation of the fundamental aspects of features such as compactness, gravity, tenseness, flattening, etc., which operate within both vowels and consonants, are obvious, and some of these have been pointed out by Fischer-Jørgensen (1957). The treatment of mutually exclusive features as being identical is, however, not only a means of gaining coding economy. The phonetic similarities do exist, though perhaps not always to the extent that a fundamental formulation covering all contexts may be expected to be self-explanatory, or optimal for all contexts. This is natural since the degree of similarity varies. How accurately and succinctly a distinctive feature is described may also depend on how well the phonetic facts are known. Acoustic speech analysis is still in a relatively early stage of development.

3.4 COMMENTS ON THE ACOUSTICAL NATURE OF DISTINCTIVE FEATURES

The formulations given by Jakobson and Halle (1956) are in several instances better founded than those proposed by Jakobson et al. (1952), but they generally lack supplementary explanations for facilitating a phonetic interpretation of the statements. A few comments on the distinctive features will next be made in the light of the results gained in the present work and on the basis of the formulations of Jakobson and Halle (1956), which are quoted here.

- (1) The feature *vocalic/non-vocalic* has been formulated as follows: *acoustically—presence (vs. absence) of a sharply defined formant structure; genetically—primary or only excitation at the glottis, together with a free passage through the vocal tract.*

Sharply defined formant structure may be explained as the physical materialization of the F-pattern, i.e., of F_1 , F_2 , and F_3 , and the dominance of this part of the formant structure versus other formants or formant regions such as the superimposed noise of voiced fricatives and stops and further the nasality characteristics of nasal consonants. It may be observed that the genetic description above applies closely to the relation of aspiration to friction in stop bursts, as described in *Section 1.11* and *Chapter 2.6*. However, in relation to a vowel of the same F-pattern, an [h]-sound is less vocalic since the open glottis executes a large damping of F_1 of acute sounds and of F_1 and F_2 of grave [h]-sounds. These effects are analogous to nasalization.

- (2) The feature *consonantal/non-consonantal* is formulated as follows: *acoustically—low (vs. high) total energy; genetically—presence (vs. absence) of an obstruction in the vocal tract.*

This feature is closely related to the classical division of speech sounds into vowels and consonants, but it is not intended to have only this function. The higher energy (or better, intensity, since the length of unstressed vowels may be extremely short) is acoustically conditioned by the degree of opening—the smaller the opening the larger the low-pass filter effect, which may also be described in terms of a lower F_1 and thus a lower overall intensity, as described in *Sections 1.11* and *1.32*.

The remarks of Halle et al. (1957) that consonants show a spectrographic orientation in the vertical direction and vowels in the horizontal direction apply to the general difference between transients and continuants, especially to sudden shifts of the spectrographic pattern due to a breaking of an articulatory closure or to a rapid movement of the articulators towards closure. However, spectrographic pattern discontinuities generally signal boundaries between speech sounds. The discontinuity may serve as a boundary between a continuant vowel and a continuant consonant or may by itself constitute the consonantal sound interval.

The coding of vowels as vocalic and non-consonantal, liquids as vocalic and consonantal, and remaining consonants as non-vocalic and consonantal seems well founded. The F-pattern of liquids is more apparent than that of other consonants. Liquids share the clear F-pattern formant structure of vowels and they also share the closure element of consonants apparent from the transient pattern variations at the sound boundaries and the lower sound intensity.

Glides are more problematic. The [h]-sound of English or Swedish has a more damped first formant than the liquids and is thus non-vocalic, and it may be regarded as non-consonantal on the basis of the relatively large articulatory opening above the glottis. On the speech wave level of specification, however, there enters the difficulty that [h] is not opposed to any other sound as being more intense. A shift of the tongue or lips from an [h]-position to a partial closure will generally be followed by a more intense *fricative* source developed at the constriction. The only non-consonantal cue of [h] is that it lacks the boundary interval discontinuities, since its F-pattern approximates that of the following vowel. The discontinuity in source from the aspiration of the [h] to the normal voicing of the vowel is not very apparent especially not in intervocalic positions and in whispered speech.

A weak non-fricative [j] connected smoothly to a following vowel may be labeled a glide and can be treated as the /h/-phoneme above (Halle, 1954). However, the Russian [j]-sounds are mostly voiced palatal fricatives, the friction being especially apparent in terminal positions (Boyanus, 1944). The relative amount of friction is about the same for [j] as for [v], and these sounds could thus be treated similarly. Retaining the intensity definition of the consonantal feature, the phoneme /j/ materialized as a voiced fricative, would be classified non-vocalic and consonantal and opposed to /ž/ as palatalized (sharp). If the friction is considered unimportant, the /j/-phoneme would be classified as vocalic and consonantal. Both /v/ and /j/ would then have to be opposed to /l/ and /r/ as less compact, /v/ being grave versus the acute /j/. This latter solution causes an undesirable split of /v/ from /f/ but has some support from the patterning of Russian phonemes (Jakobson, 1956) and from experiments with the Swedish speech synthesizer *OVE*. The phonemes /r/ and /l/ and, in initial position, /v/ and /j/ can be simulated on a pure F-pattern basis, and with voice source only, better than can any other consonants.

- (3) The feature *compact/diffuse* has been formulated as follows: *acoustically—higher (vs. lower) concentration of energy in a relatively narrow, central region of the spectrum, accompanied by an increase (vs. decrease) of the total amount of energy; genetically—forward-flanged (vs. backward-flanged). The difference lies in the relation between the volume of the resonance chamber in front of the narrowest stricture and behind this stricture.*

Table 2.33-1 provides control material with regard to the dimensions of the Russian vowels analyzed here. The volume ratio criterion is found to be applicable to the relation of [a] to [o], [o] to [u], and [e] to [i], but the same parameter is also capable of distinguishing between acute and grave vowels. Within the back (grave) vowels, the front cavity volume changes less than the back cavity volume, and the reverse is true of front (acute) vowels where the various degrees of compactness are primarily associated with front cavity volume and tongue pass area changes. The terms “inward-flanged” and “outward-flanged” horn are directly applicable to the idealized vocal tract models shown in *Fig. 1.4-4*. These models give the acoustic correlates to a differential shift in F_1 .

It should be observed, that, if the compactness feature is defined on the articulatory level, independently of the lip-rounding, there results a corresponding restriction on the level of the speech wave, so that F_1 -variations due to lip-rounding must be considered as insignificant for compactness. Under actual conditions, an increasing compactness in this form and a decreasing flattening (lip-rounding) are typical of the changes within the series of back vowels [u], [o], [a]. Which of the two is considered the primary variable depends on the particular language and the coding efficiency criteria laid by the structural linguist. On the other hand, if the primary definition of compactness in vowels is attached to the speech wave level, and F_1 is the criterion, it will be necessary to include the lip-rounding as a part of the articulatory means of reaching this effect. This latter solution does not restrict the practical use of the flattening feature, since all formants enter into the acoustic definition of flatness.

Compactness in consonants, i.e., the acoustic and articulatory features found to be typical of palatals and velars, is, as far as fricatives are concerned, simply related to the size of the cavity in front of the articulatory constriction, as described in detail in *Sections 2.62 and 2.63*. When this cavity is negligibly small, it will coincide with the constriction as in labials and dentals, and the mouth cavity configuration is then closer to that of an inward-flanged horn.

A high F_2 -locus is typical of palatals only. The F_2 of velars is lower owing to the larger and longer front cavity associated with the coarticulation with a following back vowel. A larger part of the shift down in F_2 -locus within the series [ga] [go] [gu] is, however, to be ascribed to the increasing degree of lip-rounding. The common denominator of the acoustic cues for [g] or [k] plus vowel can, as earlier stated, be described as a concentration of energy in the region of the F_{2e} of the following vowel. The articulatory invariant is the mouth cavity in front of the constriction, and

the energy is that of the burst, if present, plus the first part of the transition.¹ Within stops and fricatives the degree of spectral concentration is the main characteristic of compactness, and an appreciable variation of the location of the main formant is allowed. This freedom conforms better with the F_{2e} - F_1 -parameter than with the F_1 -parameter of vowels; see further the discussion in *Chapter 3.3*. The front cavity to the back cavity volume ratio is equally well applicable to these two parameters.

Of some importance for the distinction between [k] and [p] is the higher energy of the [k] which is related to the existence of a front resonator and to the comparatively slow variation of resonance frequencies in the first interval after a [k]-explosion. One further cue is the single formant of the [k] versus the several formants of the [p] due to the different source locations and thus the different zero functions.

(4) The feature *grave/acute* has been formulated as follows: *Acoustically—concentration of energy in the lower (vs. upper) frequencies of the spectrum; genetically—peripheral (vs. medial): peripheral phonemes (velar and labial) have an ampler and less compartmented resonator than the corresponding medial phonemes (palatal and dental).*

The term *resonator* above apparently applies to the mouth cavity. The articulatory description of the maximally acute vowel as medial (in the middle of the mouth and thus medio-palatal) applies to the [i]-position of the tongue and conforms with the nomograms of the three-parameter models²; see *Section 1.43*.

In the case of consonants, the relations are more complex.

When discussing the consonantal interval of stops and fricatives, it should be noted that the optimal point of articulation providing the highest pitch of the noise is that of dentals and not of palatals. A small front resonator gives a more effective high-pitched noise than *no* front resonator, but an increase of the front resonator length following a retraction of the point of articulation will lower the pitch of the fricative noise. The front cavity size is thus a common element of compactness, gravity, and flattening. If two front cavities plus associated constrictions are of the same length,

¹ Experiments performed at Haskins Laboratories (Delattre et al., 1955) on the perception of synthetic consonant-vowel syllables composed of two-formant vowels with initial formant transitions simulating voiced stops, have demonstrated a large shift down in the F_2 -locus from [ga] to [gɔ] dividing the group of [i] [e] [a] from the group of [ə] [o] [u]. These loci, discussed by Liberman (1957), are closely related to but not identical with the loci in the articulatory sense referred to here. Since the synthetic stimuli employed in these tests lack the third formant and a consonant burst, it can be expected that the F_2 -transitions will have to be exaggerated or in other respects somewhat different from those encountered in human speech. The speech wave characteristics do not change discontinuously from [ga] to [gɔ]. A well-known fact is, however, that the F_2F_3 -loci proximity of [g] is lost in the step from front to back vowel combination. As pointed out by Fischer-Jørgensen (1957) the observed changes are a natural effect of the coarticulation and these are discussed in some detail by Stevens and House (1956). The conclusion of Liberman (1957) that the /g/ of /ga/ and /gɔ/ are perceived as the same phoneme not because of acoustic similarities but via the similarities in articulation, does not seem well founded. Articulation and sound waves *never go separate ways*.

² The mouth cavity configuration is not necessarily of a main importance for the relation of [a] to [æ]. In terms of pharynx dimensions, however, the statement above holds in a converse sense, i.e., a sound is optimally grave (minimum F_2 — F_1 of the three-parameter model) when the back of the tongue fills up the pharynx cavity and the tongue pass is located in the center of the pharynx; cf., the common aspects of gravity and compactness.

the one with the smaller difference in cross-sectional area between cavity and constriction will give the less compact and more acute spectrum; see further the presentation in *Section 2.63*.

The F-pattern variations associated with the *grave/acute* opposition in consonants are those of a lower/higher F_2 - and/or F_3 -locus. The latter cue is essential for the distinction between palatalized [m,] and palatalized [n,], as can be seen from *Fig. 3.4-1*. It is of interest to note that the F_2 -locus of [m,] is higher than that of [n]. On a relational basis and physically in the form of the direction of the formant transition, the *grave/acute* opposition is still retained.

- (5) The feature *flat/plain* is defined by the same authors as follows: *Acoustically—flat phonemes in contradistinction to the corresponding plain ones are characterized by a downward shift or weakening of some of their upper frequency components; genetically—the former (narrowed slit) phonemes in contradistinction to the latter (wider slit) phonemes are produced with a decreased back or front orifice of the mouth resonator, and a concomitant velarization expanding the mouth resonator.*

To this formulation it may be added that a lengthening or decrease of the lip-opening will invariably cause a shift down in frequency also of F_1 and not only of the higher formants.

As stated by Jakobson et al. (1952), the feature of flatness is genetically related to *lip-rounding, pharyngealization, or retroflexion*. The effect of the two former articulatory variables on the F-pattern of a vowel may be studied from the three-parameter nomograms of *Section 1.43*. It should be observed that a narrowing of the tongue passage causes a shift down of F_2 only when it is located in the posterior half of the model, i.e., in the pharynx. At the uvula the effect is nil, and it is the opposite in the mouth cavity. The effect of retroflexion cannot be studied from these curves, but it is well known in the form of the low F_3 typical of retroflex [r]-sounds and retroflex-modified vowels of American English. If the retroflexion is applied to a very advanced place, such as the alveolars, it may happen that F_4 or even F_5 takes over the role of F_3 , especially in the speech of males with long vocal tracts.

The effect of pharyngealization on the F-pattern of consonants is qualitatively the same as that discussed above for vowels, the pharynx representing a secondary place of articulation. This modification will, however, cause only a minor change of the spectrum of the consonantal noise interval of a fricative or stop, since its characteristics are generally determined by the vocal tract configuration at and in front of the primary place of articulation. Labialization and retroflexion, on the other hand, modify the effective front cavity where the noise is filtered, and thus cause an appreciable influence on the noise spectrum as well as on the F-pattern, as can be seen by comparing *Fig. 7* and *8* of the earlier publication (Jakobson et al., 1952).

- (6) The feature *sharp/plain* has been defined as follows: *Acoustically—sharp phonemes in contradistinction to the corresponding plain ones are characterized by an upward*

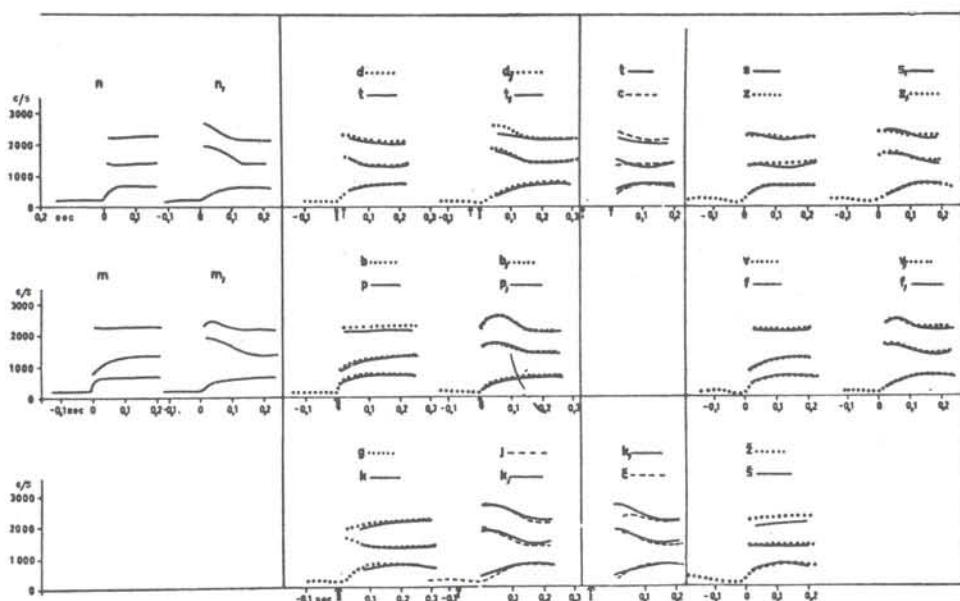


Fig. 3.4-1. Formant transitions within the vowel [a] following each of the consonants indicated. These F-pattern variations were traced from the spectrograms of the Appendix. Voiced and voiceless sounds (not always phonemically minimal pairs) have been paired and plotted in synchrony with regard to their articulation, as judged from the temporal variations of their F-patterns. The instance in time when the articulators have been judged to have started moving away from the closed phase (in the case of stops, already at the onset of the explosion) has been marked by arrows under the time scale. In order for [d,a] and [t,a] to be synchronized within the vowel, it can thus be seen that [t,_a] must start earlier. This is due to the longer fricative interval of [t,_a]. The opposite effect is found when [d] and [t] are compared. This is not typical, but it is a possible effect since the *voiced/voiceless* and not the *lenis/fortis* feature is phonemically significant for Russian stops. The [g] and [k] start at the same time, indicating equal speed of the articulatory movements during the first 70 msec after the explosion, which is unvoiced for [k].

shift of some of their upper frequency components; genetically—the sharp (widened slit) vs. plain (narrower slit) phonemes exhibit a dilated pharyngeal pass, i.e. a widened back orifice of the mouth resonator; a concomitant palatalization restricts and compartments the mouth cavity.

The *sharp/plain* feature in the formulation above is the opposite to the *flat/plain* feature in terms of the pharynx width. The latter feature contains an element in common with the *grave/acute* feature in the form of the large front cavity. Phonetically the *sharp/plain* feature is attached to palatalization, the typical effect being the rise of F_2 and also of F_3 as the tongue approaches the [i]-position. This can be seen from Fig. 3.4-1 pertaining to the F-patterns of Russian consonants calculated from the X-ray data analyzed in this work. The differences between the hard and soft Russian

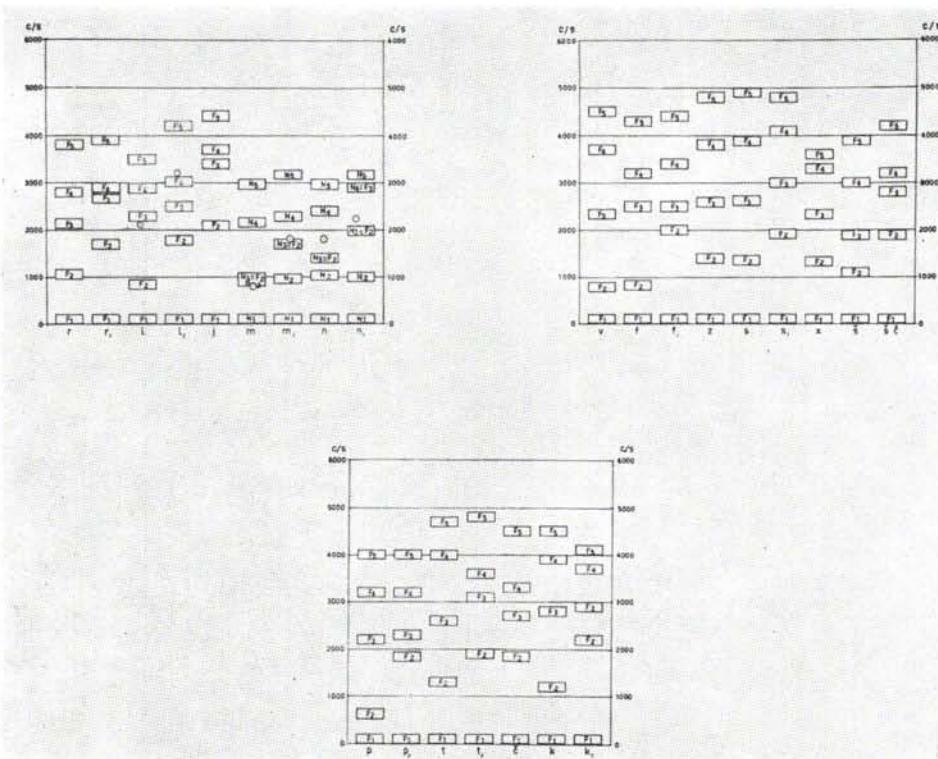


Fig. 3.4-2. Calculated F-patterns of the Russian consonants analyzed in this work. No differentiation with regard to F_1 has been undertaken. Zeros of nasals and liquids are indicated by circles.

consonants as they appear in this display are in some instances greater than those measured from the same subject's connected speech and demonstrated in *Fig. 3.4-1*. This figure contains F-pattern curves, traced from the spectrograms of the *Appendix*, for the entire interval of the vowel [a] preceded by each of the consonants. It should be noted that sounds articulated at the same place, e.g., all labials or all palatalized labials, show F-pattern transitions that are almost identical. From stops and fricatives the following data have been assembled:

Formant frequencies in c/s		
	F_1	F_2
	c/s	c/s
Plain labials		
Calculated	800	2250
Measured	920	2025
Sharp labials		
Calculated	1800	2400
Measured	1820	2400
Plain dentals		
Calculated	1370	2640
Measured	1440	(not accurate)
Sharp dentals		
Calculated	1900	3050
Measured	1700	(not accurate)
Plain [k]		
Calculated	1200	2800
Measured	1425	2600
Sharp [k.]		
Calculated	2200	2900
Measured	1800	2700

The contrast effect is greatest for the labials, the average F_2 -shift from plain to sharp phonemes being close to 1000 c/s. In this instance the plain member achieves an especially low F_2 -locus owing to the freedom of the back of the tongue to approach the pharynx, causing an extra flattening of the plain member. This flattening is also typical for the plain members of the *sharp/plain* distinctions within liquids.

Acoustically, the *sharp/plain* distinction in Russian is essentially a matter of F-pattern variation, since the cavity structure at and in front of the primary point of articulation does not vary much. The extent of such primary articulatory changes under actual speaking conditions may not be concluded from the present work, but it has been shown from the calculations of *Section 2.63* that a shift of the back cavity configuration towards a palatalized state, *ceteris paribus*, not only causes F_2 and F_3 to rise but will also be accompanied by a definite intensity increase of F_2 and F_3 . These formants, typical of the transitional characteristics, will also be apparent in the noise interval of the consonant. This effect is prominent for labials, as verified by the spoken data. In the case of dentals the effect is rather small on a purely fricative sound interval but more apparent in an aspirative off-glide.

- (7) The feature *tense/lax* has been defined by Jakobson and Halle as follows:

acoustically—higher (vs. lower) total amount of energy in conjunction with a greater (vs. smaller) spread of the energy in the spectrum and in time; genetically—greater (vs. smaller) deformation of the vocal tract away from its rest position. The role of muscular strain affecting the tongue, the walls of the vocal tract, and the glottis requires further examination.

Spread of the energy in time refers simply to the duration. Spread of energy in frequency should be interpreted as deviation from the neutral F-pattern in the case of vowels and as the greater frequency extent and intensity of the fricative noise of tense stops contributing to the upper frequency range of the spectrogram. This analogy is not very well founded.

The simple and fundamental cue of duration deserves greater attention than is conventionally paid to it. When discussing stops, it is the duration of the burst and not the duration of the occlusion that has been referred to above. If the distinction p/b, t/d, or k/g is of the *fortis/lenis* type, there is invariably found a greater length of the burst interval of tense stops from the first instance of explosion to the onset of the voicing. This appears to be primarily a matter of the status of the vocal cords at the onset of the explosion or immediately after. As illustrated by calculations and supplementary measurements on Swedish stops (*Section A.22*), the decay time for pressure equalization, and thus the duration of the airflow, is primarily dependent on the volume of the air reservoir involved. When the vocal cords close to start their vibration, this volume is that of the mouth cavity only, and the decay time is of a smaller order of magnitude than when the lung volume is also involved. The static pressure of the air to be released need not be greater for the fortis than for the lenis stop, and it would be incorrect to assume that the delay of onset of the vocal cord vibrations was caused by a higher over-pressure prolonging the time needed for pressure equalization. An additional factor contributing to the greater energy of the fortis stop is the tendency of affrication, or rather frication, by a slower off-glide of the articulators from the closed state towards the more open state of a following sound. The occurrence of a frication due to slow opening of the constriction is not confined to unvoiced sounds only. It occurs as a redundant characteristic of the Russian palatalized dentals even in the voiced [d,], which is voiced during its entire extent, the occlusion included.

If the degree of opening at the constriction has reached a relatively large value before the onset of the voice, there results an aspirative termination of the burst. The [h]-sound characteristics of the aspirative sound interval differ from those of fricative sound intervals by the greater prominence of the F-pattern formants in comparison with those of a higher frequency region.

It is apparent from the discussion above that the greater length of the burst segment in aspirated tense stops compared with lax stops is a matter of latency in the time of onset of voicing with respect to the instant of explosion. The traditional phonetic term *voiced/voiceless* thus has some sense even when the distinction is reduced to that of *lax/tense* by the disappearance of the vocal cord vibrations in the

occlusion interval. This devoicing is generally not complete in finally positioned [b], [d], or [g], and here the shortness of the occlusion and the concomitant lengthening of the preceding vowel is a cue for the *voiced/voiceless* distinction; see Lisker (1957). However, the larger energy of the [k], [p], and [t] burst is still to be observed from the intensity curve or the spectrogram.

The effect of the muscular strain traditionally claimed to be associated with the tense stops is to prolong the fricative interval of semi-closure. In the case of the tense vowels, the muscular strain cannot be expected to affect the damping of the cavity walls and thus influence the formant bandwidths to any significant extent. The tenseness and the longer duration condition an articulation further away from the neutral position, and those formant bandwidths variations that do occur, are due to the varying degrees of opening.

To sum up, the possible factors involved in the production of a tense stop versus the corresponding lax stop are (1) the static over-pressure before explosion, (2) the instant of time when the vocal cords close for the production of the following voiced sound, if any, and (3) the speed of the articulatory movements, and specifically the time spent in a position optimal for the creation of fricative noise. The first of these factors will contribute to the energy level under all circumstances but is not necessary in initial positions. The second factor is of main importance in initial positions and primarily the determinant of the degree of aspiration. The third factor governs the duration of the fricative noise. In a prolonged form it is also the main genetic determinant of affrication and thus of the stridency feature for separating the affricate [χ] from the stop [k,] and the affricate [c] from the stop [t].

(8) The feature *strident/mellow* has been defined as follows: *acoustically—higher intensity noise vs. lower intensity noise; genetically—rough-edged vs. smooth-edged*.

In view of the discussion of tenseness above, the durational aspect of the noise distinguishing affricates from stops should be included in the formulation; see further Halle (1954). The relative strengths of various noise sources in speech require further investigation.

APPENDICES

A.1 SPEECH WAVE ANALYSIS

A.11 Intensity Measurements

All measurements of the speech wave, of an oscillographic as well as of a spectrographic nature, involve the specification of amplitude or intensity measures. The term amplitude refers to the instantaneous or time average value of sound pressure, volume velocity, or particle velocity at a particular point in the sound field, or to the corresponding voltages or currents as delivered by a microphone. Any particular numerical value of the amplified, filtered, and by other means processed version of this electrical copy of the acoustic wave is also referred to by the term amplitude.

Sound intensity is the energy per unit time transmitted through a unit area. In the cgs system, which is the most common reference system in the acoustic literature and adopted in this work, the unit of intensity is erg per second per square centimeter. One such unit is 10^{-7} watts/cm². In a plane or spherical free-progressive sound wave the intensity in the direction of the propagation is

$$W = P^2/\rho c \text{ erg-sec}^{-1}/\text{cm}^2, \quad (\text{A.1-1})$$

where P is the r.m.s. sound pressure in dynes/cm², ρ is the density of the medium in g/cm³, and c the velocity of propagation in cm/sec. The product ρc is the specific acoustical resistance of the medium which is 41.4 dynes sec/cm³ at 20 °C and 40.0 at 35 °C, the latter value appropriate for the wave propagation within the vocal cavities.

Speech intensity is almost never measured directly. Pressure-sensitive microphones are utilized and the intensity, if of any interest, is calculated by means of Eq. A.1-1. Amplitude and intensity have become synonymous in a general sense because of this one-to-one correspondence. Sound pressure or intensity data are generally specified on a logarithmic scale with the unit decibel (dB) relative to a fixed reference. The standardized sound pressure reference is $P_0 = 0.0002$ dynes/cm² which corresponds closely to a reference intensity of $W_0 = 10^{-16}$ watt /cm².

The sound intensity level is defined as

$$L = 10 \log_{10}(W/W_0) \text{ dB}, \quad (\text{A.1-2})$$

which is identical with the sound pressure level

$$L = 20 \log_{10}(P/P_0) \text{ dB}, \quad (\text{A.1-3})$$

provided P_0 is related to W_0 by means of Eq. A.1-1. A sound pressure reference of $P_0 = 1 \text{ dyne/cm}^2$ is sometimes utilized in acoustics, especially in connection with microphone calibrations. The average sound pressure level at 30 cm distance from a speaker is of this order of magnitude.

When evaluating the performance of an intensity meter it is necessary to consider:

- (a) The prefiltering if any;
- (b) The type of rectification;
 - 1. Half-wave or full-wave;
 - 2. Linear, square, or intermediate rectifier characteristics;
- (c) The integration time and the shape of the temporal weighting function;
- (d) The amplitude calibration, whether linear or logarithmic or of an intermediate degree of amplitude compression in the final presentation.

Auditory criteria can be applied to all of these design stages, but it should be remembered that the relations between the physical constants of complex time variable sounds and the loudness sensations they evoke, are too complicated to be successfully materialized in the physical design of a reliable loudness meter. In addition, the psycho-acoustic basis is not sufficiently well established with regard to speech-like stimuli. Technical comments on the standardization of sound level meters are given by Hardy and others (1957).

A prefiltering stage is often incorporated in an intensity meter. The function of this filter is to perform a weighting of the relative contributions from spectral components of different frequencies so that components of the intermediate frequency range 1000-4000 c/s are emphasized. The standard sound level meter for noise measurements, see e.g., Beranek (1949), has three weighting networks, so-called frequency correction or emphasis networks.¹

Human speech covers an intensity range of about 30 dB. Vowels carrying a main stress have a sound pressure level of approximately 65 dB at one meter's conversational distance (Dunn and White, 1940), and the unvoiced consonants are on the average 20 dB weaker (Fant, 1949). These data could motivate the use of an *A-curve* pre-emphasis for measuring unvoiced consonants and the *B-curve* for collecting data on vowels. However, as an alternative to the loudness criteria it might be desired to

¹ These are labeled *A*, *B*, and *C*. The *A*-curve is recommended for use in the sound level range of 20-55 dB. This filter has maximum transmission at 2500 c/s and a low frequency attenuation that reaches 25 dB at 100 c/s. The small high-frequency suppression specified for the *A*-filter can be neglected for phonetical applications. The *B*-filter is recommended for use in the sound level range of $L = 55-85$ dB. Its frequency characteristics are intermediate between those of the *A*-curve and the purely flat *C*-curve, but closer to the latter. Thus the low frequency suppression is only 6 dB relative to the level at 1000 c/s.

obtain intensity measurements that are less influenced by the particular phonetic quality of a vowel, as specified by the F-pattern, and more directly related to the speaker's voice effort. In this respect, the use of the *C-curve* is more advisable since the first formant will then determine the major part of the intensity. A more effective method is the use of an integration, i.e., a low frequency boost, and the ideal procedure would be to remove the formants entirely in order to regenerate the voice source. This possibility is being investigated.

The rectifier characteristics are less important. Full wave rectification is recommended since otherwise one-half of the speech wave, that above or below zero pressure amplitude, will be ignored. The difference is noticeable in the case of deep pitched male voices, especially at the boundaries between speech sounds and if the rectifier characteristics are non-linear. The difference between linear mean value and square-law rectifiers can be expressed in terms of the summation of spectral components, the latter providing the ideal summation of the squares of the amplitudes of the frequency components. A linear and a square-law instrument may be calibrated to give the same scale reading for a sine wave. When they are used for speech measurements the linear device shows on the average 0-3 dB too low values, the larger deviation applicable for sounds composed of several partials of not too different amplitudes. One simple test of the summation is to superimpose two sine waves of the same amplitude on the measuring device. The square-law instrument will show a 3 dB higher level than for one sine wave only, but the corresponding figure is only 2 dB for the linear instrument. The difference of 1 dB in the indications of the two instruments is also found in measurements on random noise (Beranek, 1949, p. 453); see also the comments of Snow (1957).

Unless combined with a root extracting device, the square-law rectification provides an intensity proportional measure. Square-law rectification probably has a greater auditory significance than linear-law rectification, but the difference is not very great and can generally be taken into account in the calibration procedure.² It should also be observed that the linearly operating device is phase dependent (Haase and Vilbig, 1956). Linear rectifier characteristics are preferable for most practical applications because of the simpler design and the lesser degree of compression needed in the final stage.

If the output from the rectification stage of an intensity meter is fed into an oscillograph directly and not via an integrating stage, there results merely an oscillogram of the instantaneous intensity. This has the form of an ordinary oscillogram in which all negative excursions of the curve have been folded up to the positive side. There is also an expansion of the amplitude scale provided the rectifiers have square-law characteristics. Obviously such a curve is not very practical to use because of the unnecessary detail structure that is retained.

The integrating or smoothing process inherent in the function of a low-pass filter

* The square-law system is not an established ideal from a perceptual point of view and peak measurements may be of the same significance as loudness correlates of wide-band complex sounds.

is analogous to calculating the area under the unsmoothed intensity curve at a definite interval, for instance, a voice fundamental period. A graphic-mathematical procedure of this type operates according to a rectangular weighting function; that is, the same importance is laid on the contributions from all parts within the predetermined interval. Auditory perception is not, however, so highly time selective and neither is any physical filter. A continuous weighting of the contributions from the past according to a specific memory function occurs in both instances. In addition, however, the memory function presumably varies with the type of stimulus and with the context.

Any particular point on a recorded intensity curve represents the accumulated intensity value from a smoothly bounded interval of the speech wave. The effective duration of this interval, called the *integration time* or *averaging time*, is the area under the memory curve divided by the ordinate value at its center of gravity (Laurent, 1953). In a large class of *LC*-low-pass filters this point on the memory function is close to the maximum value. The location of this effective center of the memory function relative to the instant of observation is called the *delay time* which is another characteristic time constant of the integrating device. In one class of smoothing filters, the simple *RC*-low-pass filter specified by a simple exponential memory function

$$h(t) = h_0 \cdot e^{-t/RC}, \quad (\text{A.1-4})$$

the peak of the memory, h_{max} , occurs at the instant of observation, but the delay time according to the precise definition of Laurent (1953) will be equal to $T_a = RC$, and the averaging time will be $T_a = e \cdot RC$.

In general, the averaging time is of the order of

$$T_a = 1/2B, \quad (\text{A.1-5})$$

where B is the cutoff frequency of the low-pass filter. The delay time is of no concern in the evaluation of speech records except when several differently processed derivates of the speech wave have been recorded on a multichannel oscillograph and a synchronous sampling of the data is desired. If, instead of the oscillographic display, the intensity values are presented as needle deflections on a meter, it is the mechanical inertia of the system which performs the integration.

The combination of a rectifier and a low-pass filter of *RC*-type with a large time constant can be designed to perform a peak detection. The instrument reading reaches a maximum value very quickly within a time of the order of a voice fundamental period and discharges slowly. The instrument does not respond to a peak amplitude unless this exceeds the value remaining from the previous charge. In a piece of connected speech of about 3 seconds' length, the highest instantaneous peak is of the order of 20 dB above the long time average value, the latter defined from the energy of the sample divided by its duration (Dunn and White, 1940).

The auditory inertia, called *smear* by Joos (1948), cannot very exactly be described by time constants of the type defined above, but it appears that both the delay and

the averaging times are of the order of 20-200 msec. Experiments on the perception of simple sine waves presented in pulses of different duration have shown that an increase of pulse duration beyond the order of 0.18-0.25 sec causes no increase in the loudness sensation. Longer pulses are merely sensed as a prolongation and shorter pulses are evaluated on the basis of both intensity and duration; see e.g., von Békésy (1929). The relative contribution from each of these parameters varies with the level at which the stimuli are presented (Munson, 1947) and with the individual subject of the test (Garner, 1949). Short bursts of random noise are perceived with an averaging time of the order of 50 msec (G. A. Miller, 1948).

Much is still unknown about auditory time constants. It appears that different aspects of the speech wave are perceived with different amounts of auditory inertia and that no single time constant is sufficient. Some information reaches the higher centers very quickly via short and direct nervous transmission paths and other data, processed more intricately, are delayed and smoothed out to a greater extent. Sound pulses of large amplitude travel faster than those of a lower amplitude.

A direct reading intensity meter of the standardized *VU*-type has an averaging time of the order of 250 msec, which is somewhat longer than the voiced part of an average syllable, the latter providing a meter deflection of 2 dB above the long time average (Boeryd, 1957). Because of the non-uniqueness of a single auditory time constant, it is not advisable to design an intensity meter with too large an integration time. If this time constant is chosen small enough, it is possible, by means of graphical integration, to simulate the effect of any particular smear.

On the other hand it should be observed that if the integration time is chosen smaller than the reciprocal value of the speaker's voice fundamental frequency F_0 , a ripple of this frequency will be found superimposed on the intensity curve. This ripple is harmless as long as it is small compared with the mean intensity, and it may be made use of for measuring F_0 . An integration time of 10 msec, according to Eq. A.1-5, corresponding to a low-pass smoothing filter of cutoff frequency 50 c/s, has been found to be satisfactory and conforms with the concept of *mean speech power*, as recommended by Fletcher (1929). A 20 msec integration time will effectively remove all voice frequency ripple even from very low-pitched segments but causes too large a smear for the study of the decay and onset characteristics of the burst interval of a stop sound.

The shape of the memory function should also be taken into account. If a passive *RC*-network is utilized for the smoothing, the memory function will have a considerably faster onset than decay. If a very short impulse from the speech wave, for instance, a very short stop burst, enters such a filter and the impulse duration is shorter than the averaging time, the resulting curve will merely reflect the memory function of the filter.³

³ The present practice developed at the Speech Transmission Laboratory is to use phase-compensated *LRC*-filters with 18 dB/octave attenuation above the cutoff frequency. The memory curve of our standard filters is fairly symmetrical and has a negligible overshoot. Design data will be given elsewhere (Fant, 1959).

The final amplitude calibration of the intensity record should be expressed by a table or curve relating the amplitude of the intensity curve in millimeters to the sound pressure level of the speech in dB relative to the reference pressure. The dB calibration is natural if the final stage contains a logarithmic transformation of the output from the integrating stage but can also be adopted in a purely linear display. A compromise between linear and logarithmic presentation can be obtained with a very simple compressing device which performs a logarithmic translation of large and medium size amplitudes but a linear translation of small amplitudes. The resulting amplitude scale, referred to as compressed, could be more representative of hearing than the purely logarithmic display insofar as the minimum perceptible intensity difference in dB is larger at the threshold of hearing than at higher sensation levels. These DL -values range from 0.5 to 5 dB . The DL for vowel intensity is close to 1 dB , as reported by Flanagan (1957a). According to S. S. Stevens (1956), an increase in stimulus level of 10 dB causes a doubling of the loudness sensation. These data enable an estimate of the accuracy needed in collecting intensity data.

As a physical correlate to stress it has been proposed to measure, from a linear amplitude recording, the area under the syllabic peaks, thus combining intensity and duration in a single measure (Fant, 1949, 1957).⁴ This is motivated by the dependency of loudness on duration and the experience from speech analysis and synthesis that a shortening alone can have the effect of changing a listener's stress response. This area measure, tentatively called *impulse index*, has the dimension of energy only if the rectifier characteristics are quadratic. The area measure obtained from linear rectification has a conceptual similarity to energy but is strictly of the dimension of sound pressure multiplied by time. It is practical to express the impulse index on a dB scale, with reference to area ratios. Assuming quadratic rectifiers, a 3 dB intensity increase of a syllable, i.e., a doubling of its amplitude, will accordingly be considered equivalent to a doubling of its duration. A linear rectification and recording will give less weight to the intensity, since a doubling of the amplitude of the intensity curve means a 6 dB increase. There is not yet enough experience accumulated for a decisive recommendation of any particular scale and form of presentation. Linear rectification seems to be the more practical solution, and a linear recording should be utilized.

A.12 Spectrum and Waveform Measurements

Present techniques of spectrographic and oscillographic speech analysis provide the following data:

1. Instantaneous amplitude;
2. Intensity averaged over a short time of the order of 10-20 msec;
3. Spectral composition in terms of one of the alternative forms:
 - A. Fourier series containing amplitude and phase of the fundamental and harmonics;

⁴ Other stress correlates are the increase of voice fundamental frequency F_0 and the increased vowel/consonant contrast in stressed syllables; see further Fant (1957).

B. Amplitude and phase of the outputs of a set of band-pass filters of sufficient number to cover the frequency range of interest.

The phase concept referred to here is that of instantaneous phase $\varphi_m(t)$ of the output $v_m(t)$ of a band-pass channel No. m of center frequency $F_m = \omega_m/2\pi$.

$$v_m(t) = V_m(t) \cos [\omega_m t + \varphi_m(t)], \quad (\text{A.1-6})$$

where $V_m(t)$ represents the amplitude information. Both $V_m(t)$ and $\varphi_m(t)$ are considered to vary only by small amounts within the period time $2\pi/\omega_m$. The oscillation of center frequency F_m is introduced as a carrier of the amplitude and phase information. After rectification and smoothing, only the amplitude information $V_m(t)$ remains. Any observed value $V_m(t)$ represents a sample of the effective duration $1/B_m$, where B_m is the bandwidth of the analyzing filter. This is due to the integrating function of the filter. Observe the similarity with the integrating function of a low-pass filter Eq. A.1-5 and the difference with regard to a factor of 2.

Phase information is of very little importance for the perception of speech and can generally not be recorded by our present analyzers. Phase information is generally predictable from the amplitude information. The reverse relation holds partially so that phase-frequency analysis can, to some extent, be substituted for amplitude-frequency analysis (Huggins, 1952). Phase is thus of a more theoretical interest. A large frequency-dependent phase shift in a transmission system or in a recording system results in a delay in the time of arrival of spectral components. If the phase shift is not linearly related to the frequency, there will be separate time delays for separate frequency intervals of the spectrum.

The intensity-versus-frequency distribution of very short clicks is the same as in white noise, i.e., uniform. In other words, the spectrum level per unit bandwidth is the same in every part of the spectrum. In the click, all frequency components are in phase. In random white noise, on the other hand, the phase is distributed at random throughout the spectrum. From this example it might appear important to measure phase. In view of the foregoing discussion, however, it is quite sufficient to restrict the specification to amplitude and time of arrival of separate spectral components of bandwidth B and duration $1/B$ distributed within the frequency-time plane of spectrographic representation. The response of the filter to an impulse of infinitely short duration is identical with its memory weighting function, the effective duration of which is $1/B$. Only if the duration of the noise is shorter than $1/B$ will the spectrograph fail to distinguish the click from the short burst of noise.

Graphical Fourier analysis, the so-called harmonic analysis, is nowadays seldom carried out, since it is so much easier to utilize a filtering method. Harmonic analysis has, however, been of great importance in an earlier period of experimental phonetics (Crandall, 1925; Steinberg, 1934; Lewis, 1936; Sovijärvi, 1938a; Tarnóczy, 1948). One unnatural aspect of the classical Fourier frequency analysis is that it is equivalent to the use of an analyzing filter of infinitely narrow bandwidth. The representation in terms of elementary frequency-time limited spectral components applied above

not only provides a useful theory for describing the function of a spectrograph, it also serves as a better signal representation than Fourier series and Fourier integrals when discussing auditory phenomena (Gabor, 1946, 1953). There is, however, need for more accurate data on the effective bandwidths and time constants for the perception of different spectral aspects as well as of the shape of the memory function; compare the discussion in *Section A.11*.

Joos's preference for narrow bandwidth in spectrographic analysis was motivated by the 50 msec smear time he suggested as representative of auditory integration. However, the *smear* effect in spectrographic analysis depends in part on the band-pass filters, in part on the integrating circuitry following the filters, and these may be designed to perform the essential part of the integration. It seems more probable that the auditory organs function like a broad-band analyzer, the multiple channel outputs of which are rectified and smoothed and intercorrelated in the nervous system (Licklider, 1952) with an integration time which is large as compared with the reciprocal of the bandwidth of the peripheral filters and of the order of 20-200 msec. There are also indications that the multipathway connections for transmission of spectral information are associated with different time constants (Gabor, 1946). The frequency analysis discussed so far is concerned with the *quality* or *timbre* aspects of speech sounds. The perception of the fundamental pitch and thus of intonation appears to be related, at least in part, to a separate auditory process. To make the analogy to broad-band spectrography complete, this might be thought of as a process of counting the number of energy maxima per unit time within any or all of the band-pass channels, i.e., a detection of the time function envelope periodicity. It must be stressed that this analogy is not intended to explain the auditory mechanism, but to discuss some aspects of its function.

In one class of spectrographs the amplitude information from a number of separate frequency bands is displayed almost simultaneously by a rapid scanning process, e.g., in the 48-channel spectrograph at the R.I.T. (Sund, 1957). In other analyzers requiring repeated playback for the analysis, e.g., the Kay Electric Sonagraph originating from the Bell Telephone Laboratories (1946) sound spectrograph, the mid-frequency of a single filter is shifted by a small amount for each completed loop of the playback.

The Sonagraph provides time-frequency-intensity records, so-called spectrograms, of a maximum length of 2.4 seconds. The time scale is horizontal, the frequency scale vertical, and the intensity of spectral components is represented by a continuous black-grey-white marking scale. Amplitude, or more precisely intensity level versus frequency sections, can be taken at various desired positions along the time scale. Both spectrograms and sections may alternatively be produced with a narrow analyzing filter, $B = 45 \text{ c/s}$, or with a broad-band filter $B = 300 \text{ c/s}$. A section constitutes a sample of a definite center position in time with an extent in time equal to the averaging time of the integrating circuitry, the inertia of the band-pass filter included. The latter is decisive for the sample duration of a narrow-band section which is of the order of 30 msec. In broad-band analysis the integrating circuitry following the filter determines a sample duration of the order of 5 msec.

Sonagraph analysis is exemplified in *Fig. A.1-1*. In narrow-band analysis, see *A* and *C*, in this figure the individual harmonics are resolved. They constitute a fine structure of fairly horizontal lines running through the formants of the frequency-intensity-time spectrograms. They appear as individual peaks within the sections. The width of one of these harmonic lines is equal to the bandwidth of the filter.

Formant frequency, formant bandwidth, and formant level can be defined from the envelope which can be drawn to enclose smoothly the harmonics within the spectral maximum (Fant, 1956). The frequency of a formant is generally measured as the frequency position of the envelope maximum, but this point cannot be utilized for an unambiguous definition. Difficulties are encountered in case of highly asymmetric formants and when the voice fundamental frequency F_0 is high. A center of gravity definition as proposed by Potter and Steinberg (1950) could have some bearing on perception but does not solve the problem. It is preferable to attempt an estimate of the corresponding resonance frequency of the vocal tract filter function for the benefit of the F-pattern specification; see *Sections 1.22* and *1.32*. The formant level is generally defined as the envelope level, i.e., the sound pressure level in *dB* of the envelope peak.⁵ It is also possible to define the formant level from the sum of the intensities of the individual harmonics within the formant. If the voice pitch F_0 is greater than the formant bandwidth, there is less than 1 *dB* difference between the intensity level and the envelope peak level. These relations will be discussed in more detail in a separate publication. Formant bandwidth is measured by drawing a line parallel to the frequency axis 3 *dB* below the envelope peak. The bandwidth is the distance between the two points where the envelope is intersected.

Provided the fundamental pitch is not too high, the broad-band spectrograms will fail to show the individual harmonics. This is because more than one harmonic at a time will be passed through the filter. A large bandwidth B means a small integration time $1/B$. Thus the intensity variations within a voice fundamental period will be detected. The vertical striations typical for a spectrogram of a low pitch male voice are indications of each vocal cord period. Each pulse of air from the glottis injected into the vocal cavities will give rise to a damped oscillation.

A formant in the frequency domain is mathematically identical with a damped oscillation in the time domain. If this had been quite clear 50 years ago, the classical Helmholtz theory and the Hermann-Willis theory of speech production would never have been opposed to each other as being different. This was pointed out already by Lord Raleigh (1896); see further Chiba and Kajiyama (1941); Trendelenburg (1950).

The apparent width of the formant bands in a broad-band spectrogram is the sum of the true formant bandwidth and the bandwidth of the analyzer, of which the latter is by far the larger—at least when dealing with the first three formants. If a speaker's average F_0 is high and of the same order of magnitude as the width of the broad

⁵ Peterson and Barney (1952) use the term *amplitude* for the *dB*-value of the formant peak. As long as the unit and procedure for measurements are defined, it does not matter what term is used.

analyzing filter, the spectrographic display will show the individual harmonics during sound intervals of high F_0 , i.e., at intonation peaks, but the formant structure may appear properly or in mixture with the harmonic fine structure at sound intervals of low F_0 . This may be rather confusing, and it is recommended to make supplementary spectrograms with the tape played back at half speed, thus reducing both F_0 and the formant frequencies by a factor of 2. The formant structure will then appear but at the expense of a reduced selectivity. Close-lying formants may not be sufficiently well separated.

Recent developments of instruments for collecting spectrum data of speech have been reported in an earlier publication (1957). One very accurate but rather time-consuming method of deriving amplitude-versus-frequency sections is to make amplitude-versus-time oscilloscopes of the speech wave from the outputs of a number of band-pass filters covering the appropriate part of the frequency scale or from a single band-pass filter that is shifted in mid-frequency after each completed cycle of the entire recording. This was the technique adopted for deriving the sections shown in the *Appendix* and utilized here for comparisons with calculated data. It has been utilized earlier for the analysis of stop sounds (Fant, 1949).

One of the early instrumental methods for spectral analysis of speech was the sweep-frequency analysis, *Suchtonanalyse*, performed with a wave analyzer of continuously variable carrier frequency. This method provides constant bandwidth, but continuously gliding center frequency of the effective filter function of the instrument. The instrumentation utilized by, for instance Sovijärvi (1938a,b) and Barczinski and Thienhaus (1935), required a time of analysis of the order of two minutes. Trained singers were required to sustain a stationary sound of that length. The present techniques of this method (Meyer-Eppler, 1950; Fant, 1957) have been improved so that the sounds to be analyzed need not be sustained more than the theoretical minimum, which is the extent of the frequency scale to be covered divided by the square of the filter bandwidth utilized in the analysis. Sweep frequency analysis covering a frequency range of 8000 c/s thus requires a time of analysis of 2 sec when a 63 c/s bandwidth is utilized, and only 0.5 sec for a 125 c/s filter. It does not pay to reduce the time of analysis below 0.5 sec, since the temporal fine structure from successive voice periods will then be too disturbing. The vowel analysis exemplified in Fig. A.1-2 shows the high quality spectral tracings⁶ obtained by connecting the

⁶ The curves of Fig. A.1-2 were obtained without any high frequency pre-emphasis. It has been found that a simple high-pass RC-network specified by a zero at 200 c/s and a pole at 5000 c/s, both on the negative real axis, provides a generally useful degree of correction for the low frequency emphasis typical of the spectra of most speech sound, thus permitting weaker high frequency formants to be seen better. This correction curve has the desirable property of being simple and symmetric around 1000 c/s and represents a compromise between the demands of a compensation for the long time average speech spectrum and the frequency dependent perception of loudness. The network approximates a 40 phone equal loudness curve for frequencies below 3000 c/s. In our experience it appears to provide better spectral balance than the Sonagraph high frequency pre-emphasis. For routine presentation of spectrum curves it would also be desirable to utilize a frequency scale such as the mel scale (Stevens and Volkman, 1940) so that the visual impression of the

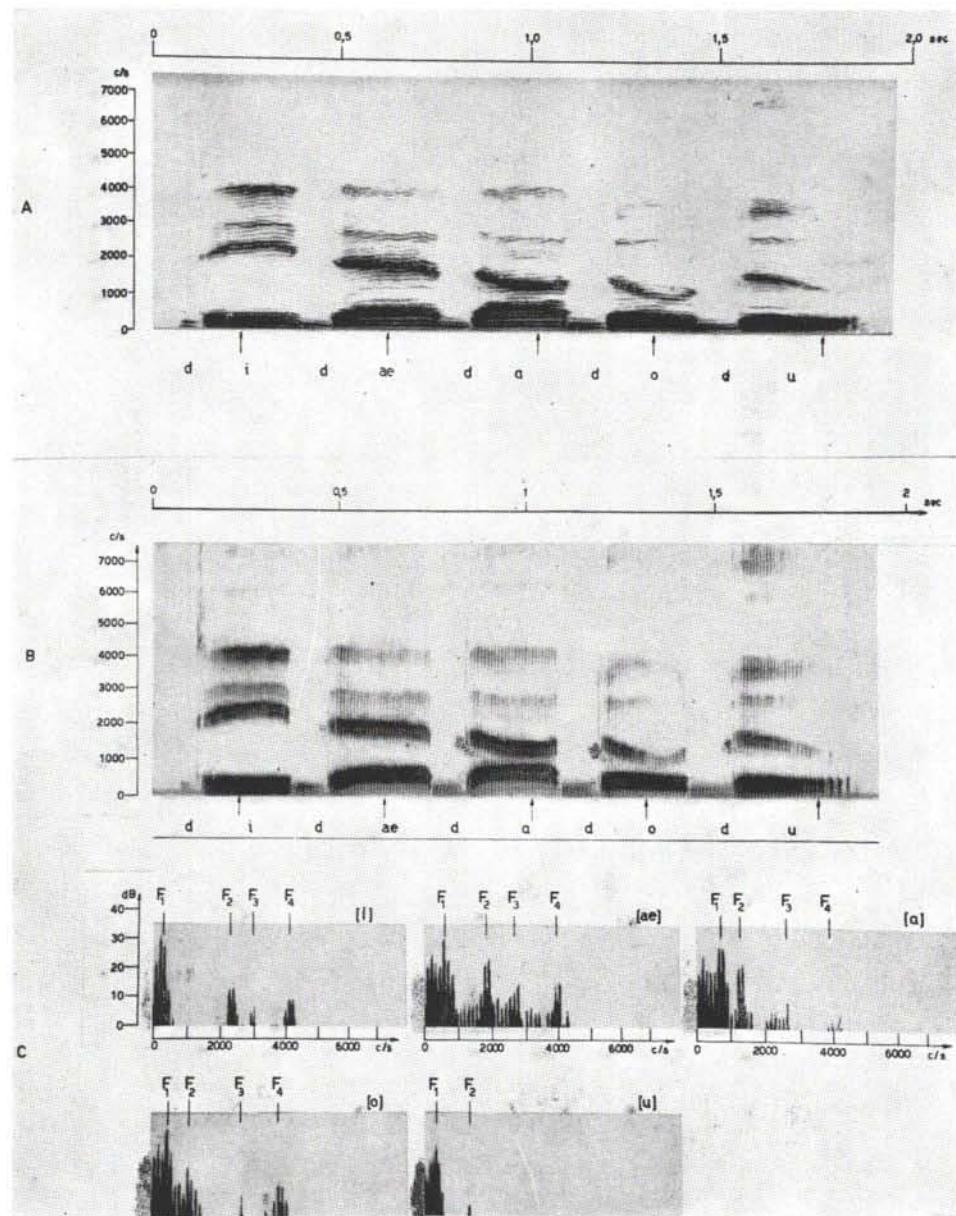


Fig. A.1-1. Spectograms: A. narrow band, B. broad band, and C. sections, obtained with a Sonagraph. Speech material [didædadodu], American subject, H.T. The time locations of the sections are indicated by arrows under the spectrograms.

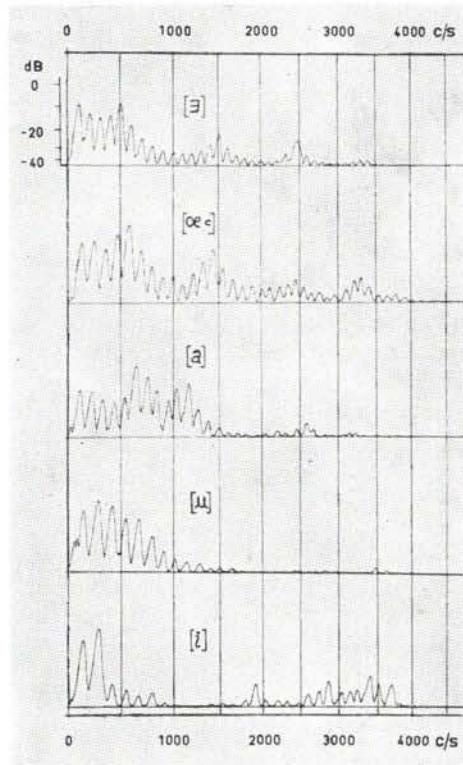


Fig. A.1-2. Sweep frequency analysis of the synthetic neutral vowel and some sustained vowels, Swedish subject, G.F. A 31 c/s bandwidth of the heterodyne analyzer was utilized. Sweep speed 4000 c/s in 3 seconds. Mingographic recording, no high frequency pre-emphasis. The individual harmonics as well as the formant structure are apparent.

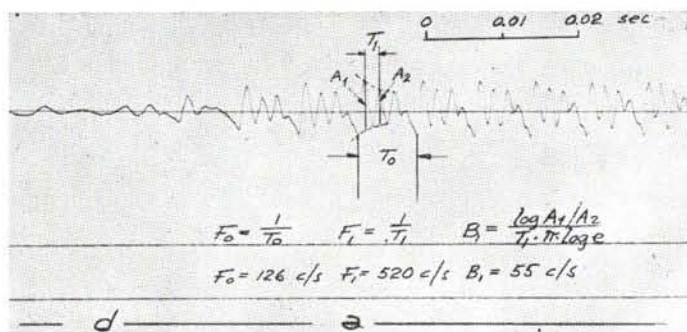


Fig. A.1-3. Waveform analysis of the fundamental pitch F_0 and of the frequency F_1 and bandwidth B_1 of the first formant in the fourth period of the vowel [ə] in [də] Mingograph high-speed oscillogram with half-speed tape-recorder playback.

analyzer output to a direct-writing oscillograph recorder of high frequency response. This method is especially suited for voice quality studies.

From an engineering point of view it seems rational to measure and to define the formant from the time function. Given a high-speed oscillogram of the speech wave, with a prefiltering arranged to isolate the formant of interest, the formant frequency can be measured from the time of a full period of the oscillation. The initial amplitude of the oscillation may then be adopted as an alternative measure of formant intensity. The bandwidth is defined as $B = \sigma/\pi$ where $1/\sigma$ is the time it takes for the oscillation to decay 8.6 dB, i.e., a factor e ; see Fig. A.1-3. This method has the advantage of providing formant frequency data in conformity with the F-pattern definition.

Frequency and intensity measures taken from the time function generally agree quite well with data from spectral sections. Formant bandwidths obtained from frequency domain measurements are constantly greater than the measure obtained from the time domain as described above. The latter are physically more correct. There is always some bandwidth broadening owing to the time-averaging effects in filters or in Fourier series expansions.

The voice fundamental frequency, see Eq. I.1-2, is defined as $1/T_0$ where T_0 is the duration of a fundamental period measured from an oscillogram or from the vertical striations in a broad-band spectrogram. Any observable harmonic of known number can be utilized for direct measurements of fundamental frequency from narrow-band spectrograms or cross-sections. Automatic pitch extracting instruments can also be used, instead of the spectrographic or oscillographic methods. Present devices originate more or less from the Grützmacher and Lottermoser (1937) pitch extractor. It provides a continuous curve of the voice fundamental frequency within voiced sections of the speech wave. Such devices, however, are not always reliable.

The following symbols will be used:

F_n = frequency of formant number n in c/s;

B_n = bandwidth of formant number n in c/s;

L_n = level of formant number n in dB;

F_n = formant number n without specific reference to its dimensions. The first formant is thus denoted by F_1 and its frequency by F_1 and so on;

F_0 = frequency of the voice fundamental in c/s;

F_0 = the voice fundamental without specific reference to its dimensions.

spectral distribution should conform to the auditory importance of various frequency regions. A common linear scale gives too much prominence to higher frequencies and a logarithmic scale will overemphasize the low frequency end. The Koenig (1949) scale is a first order approximation specified as linear below 1000 c/s and logarithmic above 1000 c/s. A better approximation would be to relate the abscissa distance x cm to the frequency f by the formula $x = k \log(1 + f/1000)$, (Fant, 1949). It has the advantage that the interval discontinuity at 1000 c/s is removed. Since most spectrum recording devices are designed for either logarithmic or linear frequency scale and since it is laborious to replot spectral curves, it is generally customary to retain the frequency scale of the original recording.

The natural range of variation of the voice fundamental frequency and of formant frequencies for non-nasal voiced sounds uttered by average male subjects is as follows:

F_0	60-240	c/s
F_1	150-850	c/s
F_2	500-2500	c/s
F_3	1500-3500	c/s
F_4	2500-4500	c/s

Females have on the average one octave higher fundamental pitch but only 17 per cent higher formant frequencies; see Peterson and Barney (1952); Fant (1953b). Children about 10 years of age have still higher formants, on the average 25 per cent higher than adult males, and their fundamental pitch averages 300 c/s. The individual spread is large.

At an average voice level, the level of the first formant measured at a distance of 1 m is 60-65 dB re 0.0002 dynes/cm² (Dunn and White, 1940; Fant, 1949). Higher formants have decreasing levels and they vary according to the particular formant frequencies, as will be discussed later.

The first two formants have bandwidths of the order of 30-100 c/s. Higher formants have increasing bandwidths, B_3 and B_4 , ranging from 40-200 c/s. These data were obtained from oscillographic measurements. They are considerably smaller than the bandwidth data reported by Bogert (1953). The systematic differences are due to the measuring technique as discussed earlier. Recent data provided by House and Stevens (1957) and Stevens (1958) as well as the data of van den Berg (1953) conform closely to the order of magnitudes stated above.

A.13 Spectrographic Illustrations of the Speech Material Utilized for the Control of the Consonant Calculations

The following illustrations labeled A.13-1 to A.13-19 contain speech wave characteristics of monosyllabic test words of the type consonant plus [a] spoken by the subject used for the X-ray photography. The original recording was made in an anechoic chamber and included the syllables

Fig.		Fig.		Fig.	
A. 13-1	[ma]	A. 13-6	[f,a]	A. 13-13	[ča]
A. 13-2	[na]	A. 13-7	[v,a]	A. 13-14	[pa]
A. 13-3	[l,a]	A. 13-8	[sa]	A. 13-15	[ba]
A. 13-4	[r,a]	A. 13-9	[z,a]	A. 13-16	[ta]
A. 13-5	[j,a]	A. 13-10	[x,a]	A. 13-17	[da]
		A. 13-11	[ša]	A. 13-18	[ka]
		A. 13-12	[ža]	A. 13-19	[ga]

spoken in succession and at the rate of approximately two syllables per second.

The analysis comprises spectrograms (*Sonagraph*), intensity curves with various types of pre-filtering, and section samples at places of interest within the consonants

to be studied. The amplitude-versus-frequency sections were compiled from a number of *Mingograph* recordings of the output of a wave analyzer with a bandwidth of 150 c/s and variable center frequency.⁷ These sections are more time-consuming to prepare than those obtainable from the Sonagraph, but they provide a better intensity resolution. Thus, *F*2 and *F*3 of a voiced occlusion may generally be detected. They have been utilized as control material for estimating the practical potentialities of the techniques for predicting speech spectra from X-ray data.

The time location of each section is specified by reference to a specific point on the time scale of the spectrogram. For analysis of unvoiced stops, sections have been taken at three different sampling points or intervals within the burst labeled *I*, *II*, and *III*. These correspond to explosion, frication (fricative interval), and aspiration respectively; see for instance [t,] and [k,]. Generally the first two or the last two of these events are more or less mixed.

The following synchronous time functions were recorded:

- (1) Oscillogram with *B-curve* (see *Section A.11*) suppression of low frequencies. The upper frequency limit is that of the *Mingograph* or approximately 800 c/s;

All intensity-versus-time characteristics (2)-(8) were recorded with an effective integration time of 10 msec, corresponding to the 50 c/s low-pass cutoff frequency of the smoothing filter. Full-wave linear rectification was utilized. The amplitude scale is logarithmic for (2) and (3) and linear for (4), (5), and (6) but is calibrated in dB in all instances. The following pre-filtering was utilized for the separate time functions.

- (2) Sound level meter *A-curve* pre-emphasis. This is essentially a low frequency suppression; see *Section 1.11*;
- (3) High-pass filtering, cutoff frequency 1500 c/s;
- (4) Band-pass filtering, 1400-1800 c/s;
- (5) Band-pass filtering, 2800-3600 c/s;
- (6) High-pass filtering, cutoff frequency 4000 c/s.

Several observations of general interest may be made from these illustrations,⁸ e.g., the difference in temporal intensity envelope comparing stops, affricates, and fricatives. The stops display a relatively short burst interval or at least a short fricative interval within the burst. The duration of the affricate is longer than for the stops but shorter than for continuants. The stop burst has a much shorter onset interval than decay interval. These are of about the same length for the affricate and in the case of continuants in initial position the rising interval occupies a major part of the sound owing to the increasing over-pressure associated with the onset of a breath pulse; compare *Fig. A.2-2*. These onset, duration, and decay characteristics have been summarized quantitatively by Halle (1954).

⁷ Halle and Hughes participated in the initial stage of this work.

⁸ The very weak third formant of the vowel as seen in several of the spectrograms is merely a personal characteristic of the speaker. As shown in *Section 2.42*, there is a theoretical possibility that a critically small degree of nasal coupling can produce this effect, the third formant gaining in intensity at larger or no nasal coupling; compare [la] with [na] and [za], *Fig. A. 13-3*, *A. 13-2*, and *A. 13-9*.

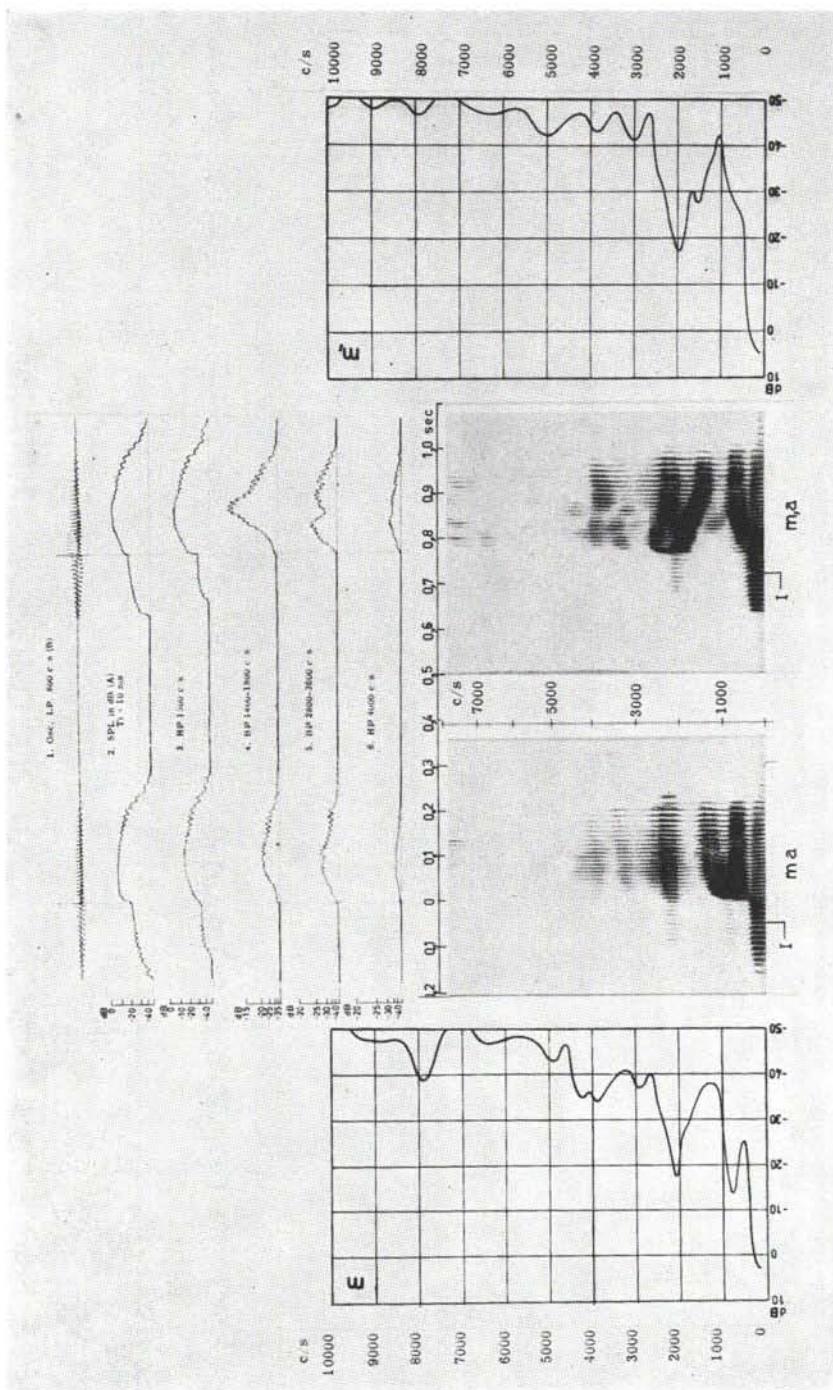


Fig. A.13-1. [m,a], [m,a].

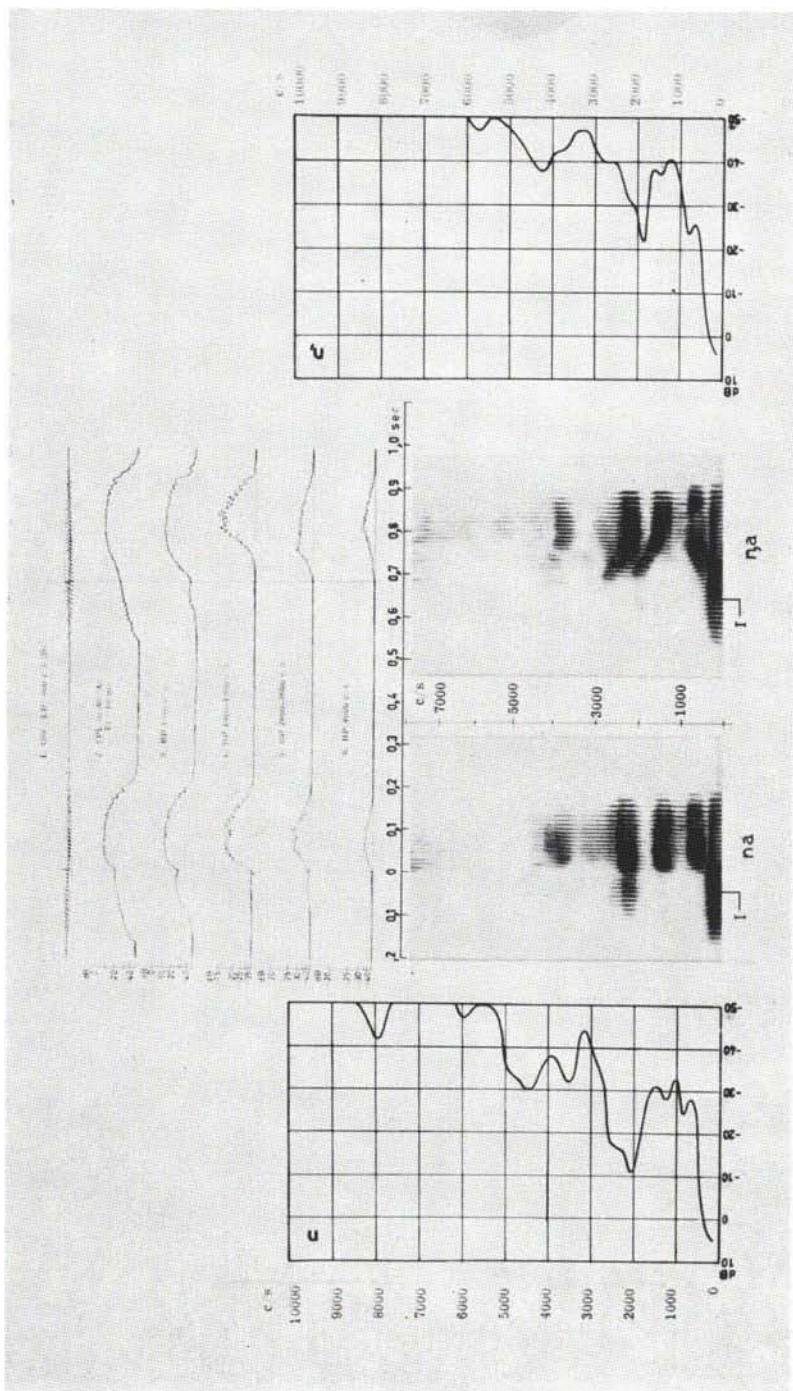


Fig. A.13-2. [ɪ], [ə], [ɑ].

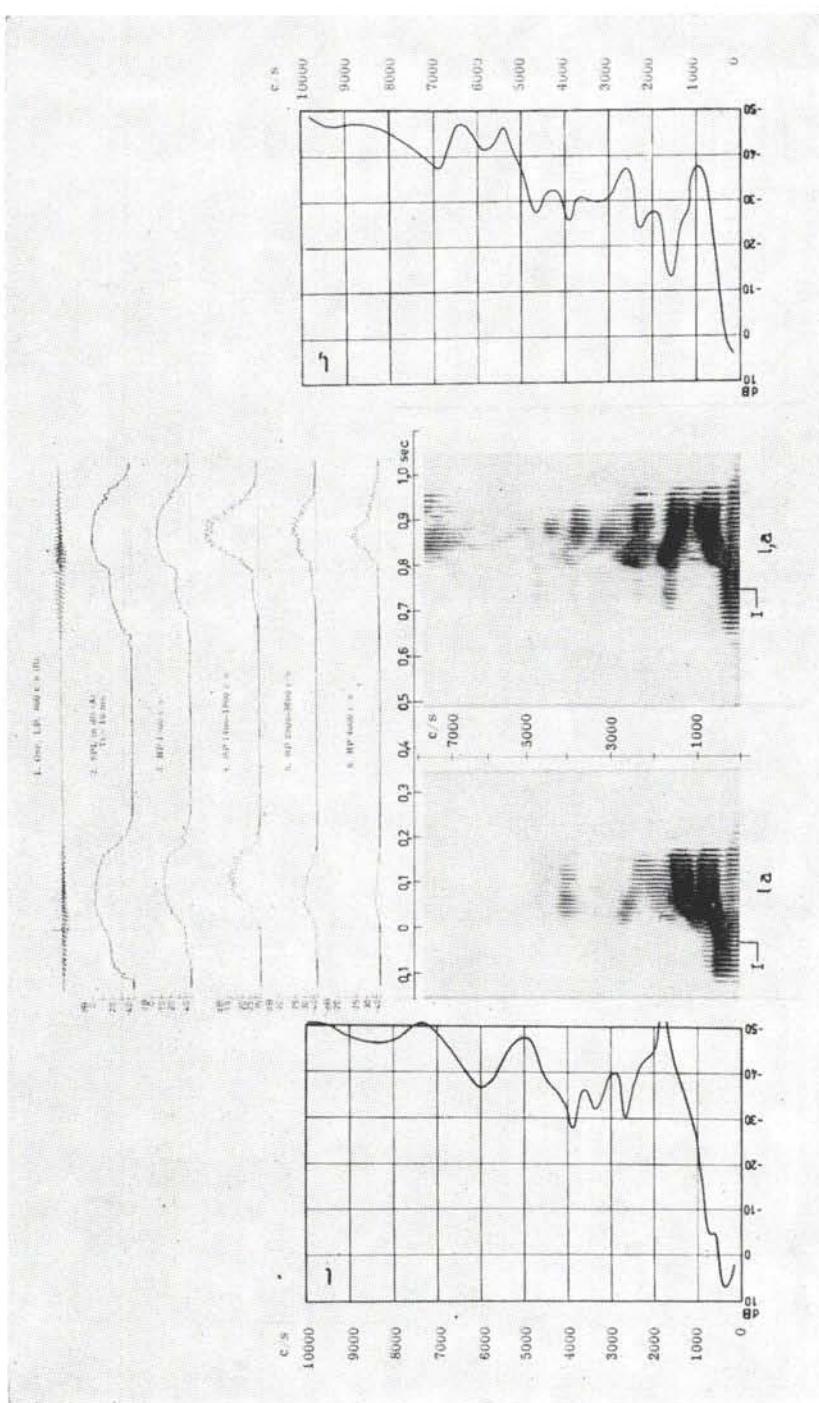


Fig. A.13-3. [ɪ], [ɪ,a].

Appendices

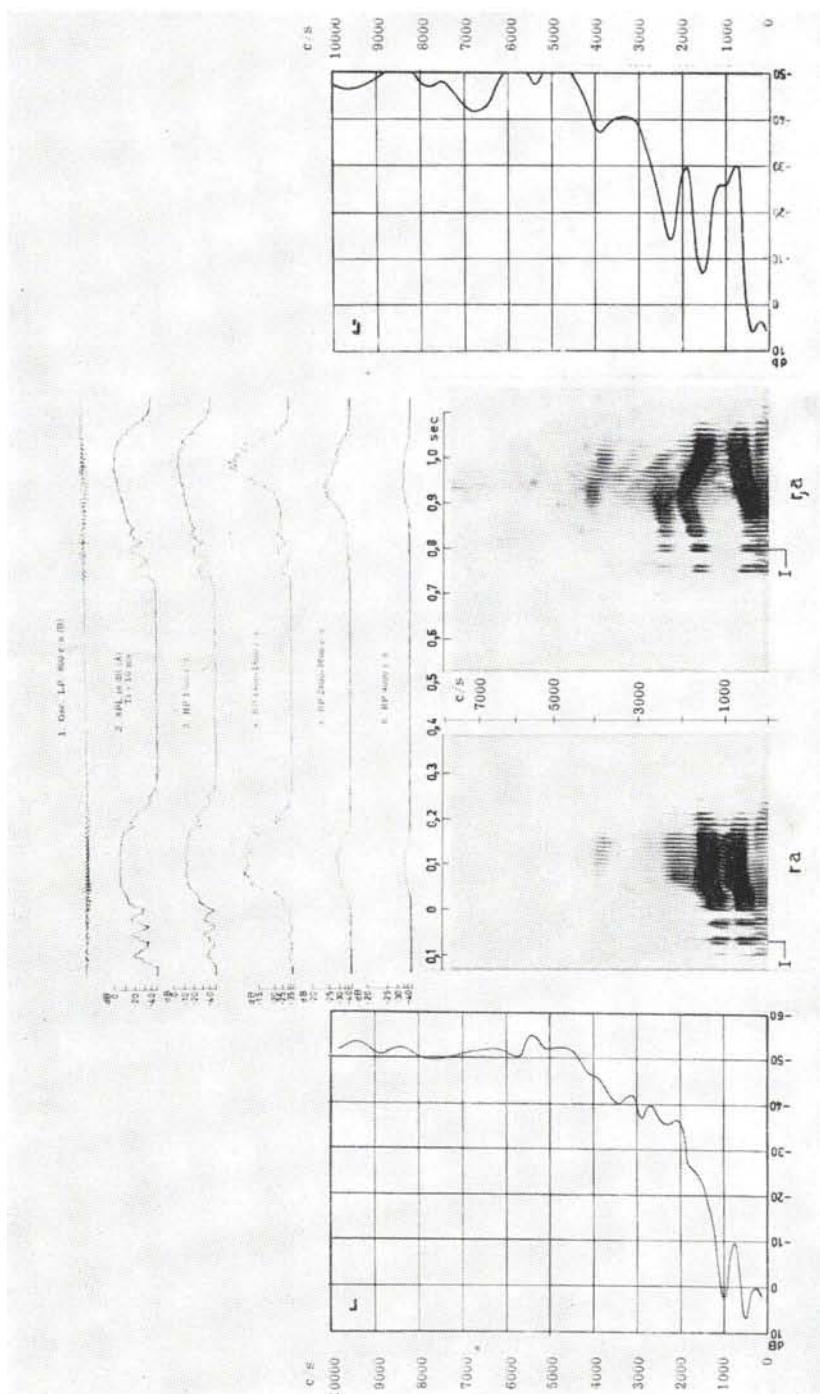


Fig. A.13-4. [ra], [r,a].

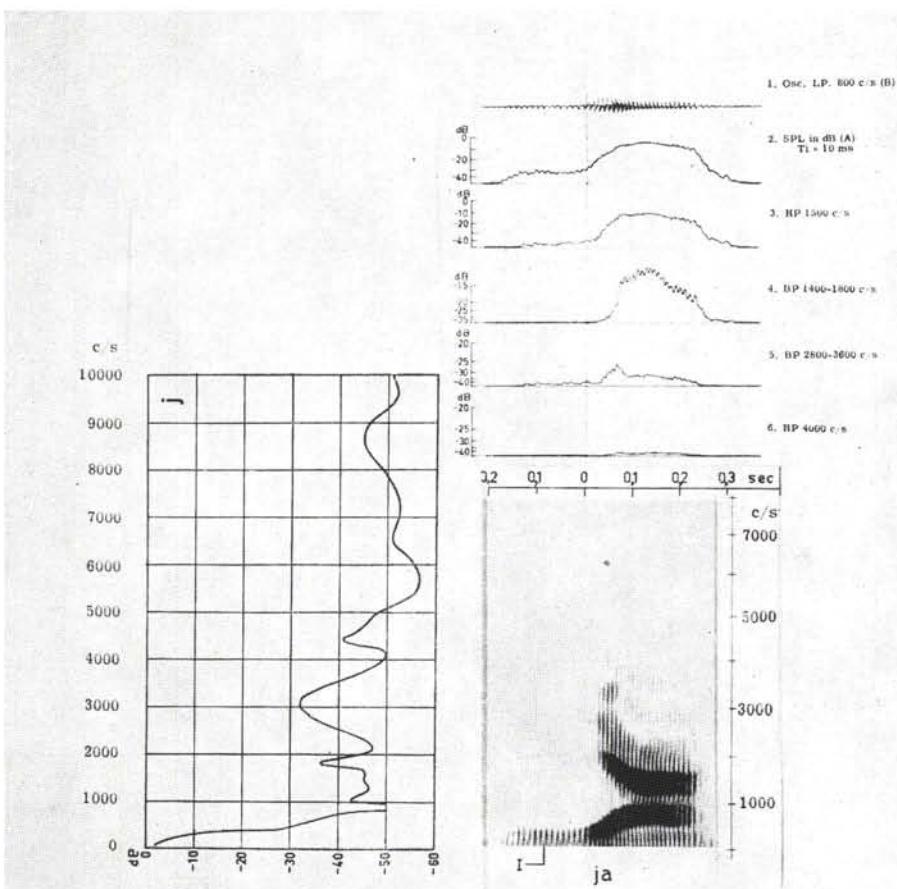


Fig. A.13-5. [ja].

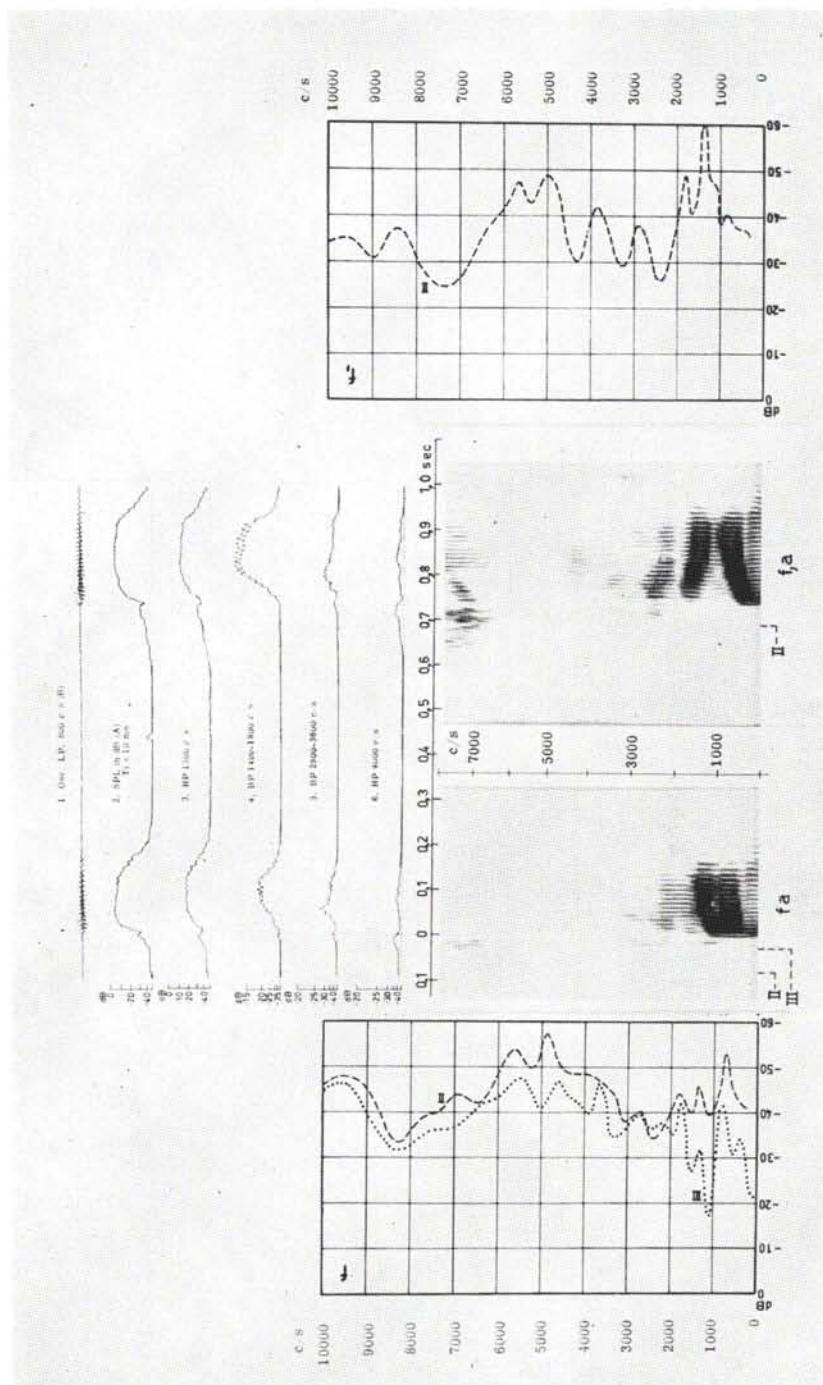


Fig. A.13-6. [f,a], [f,a].

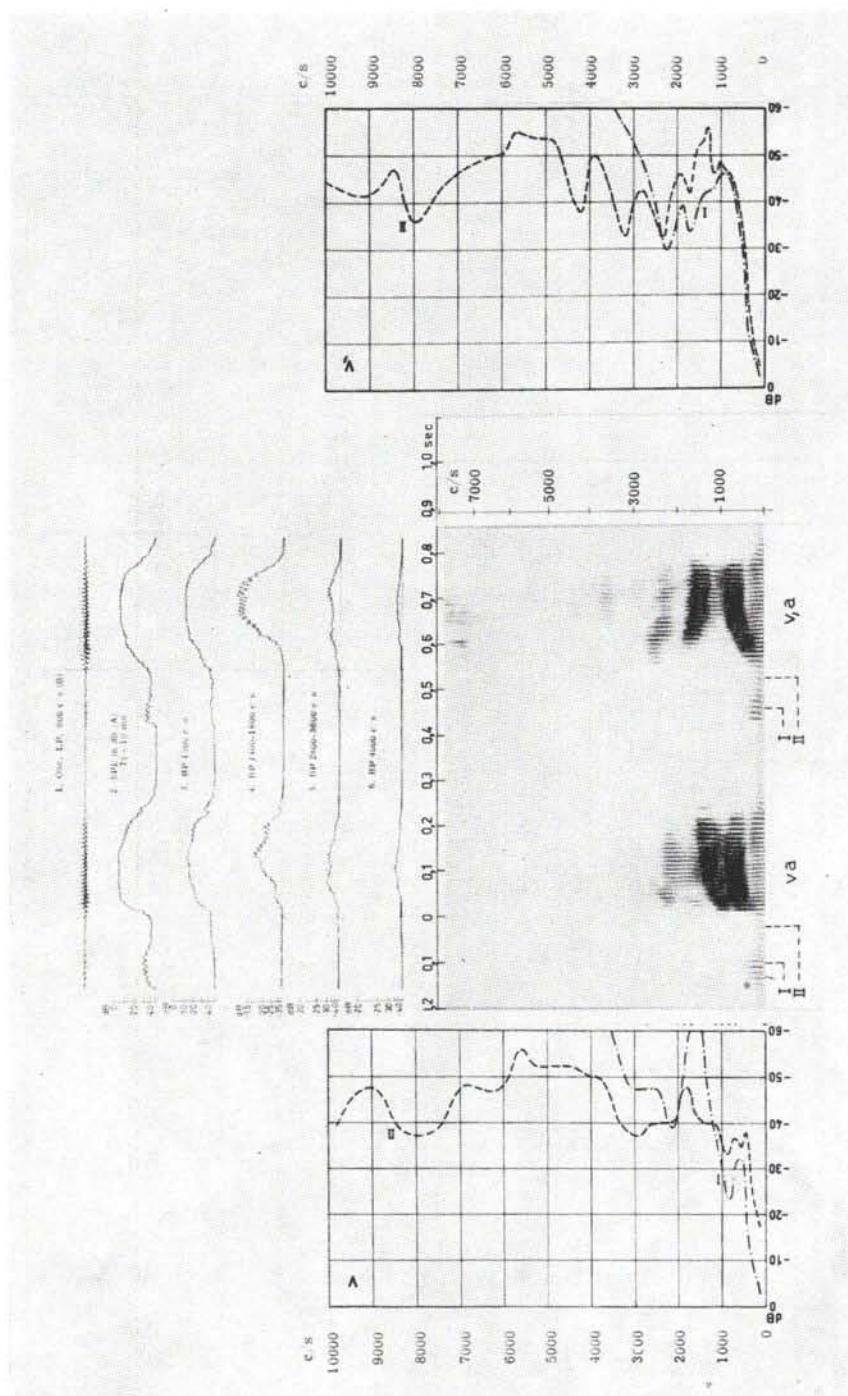


Fig. A.13-7. [v,a].

Appendices

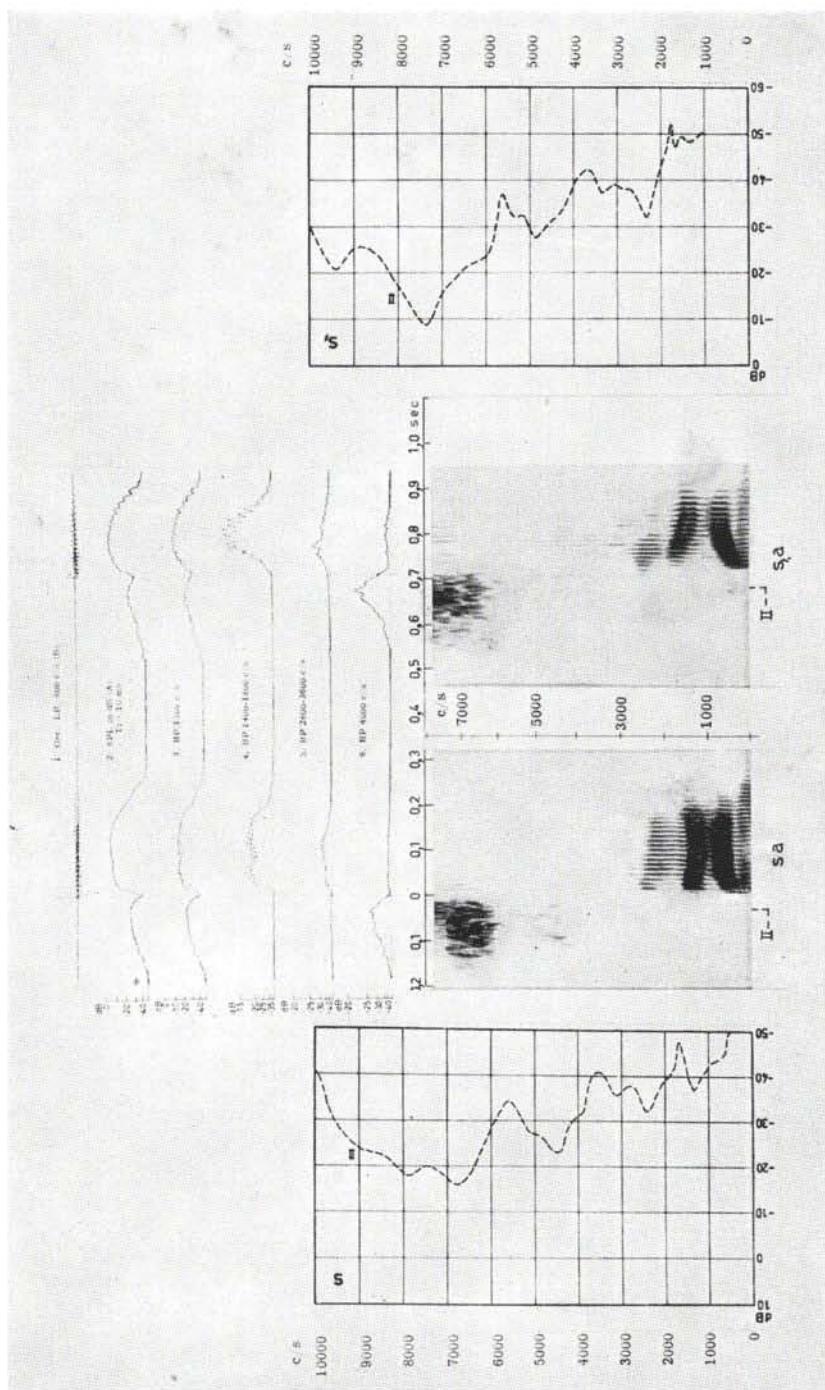


Fig. A.13-8. [sa], [s̪a].

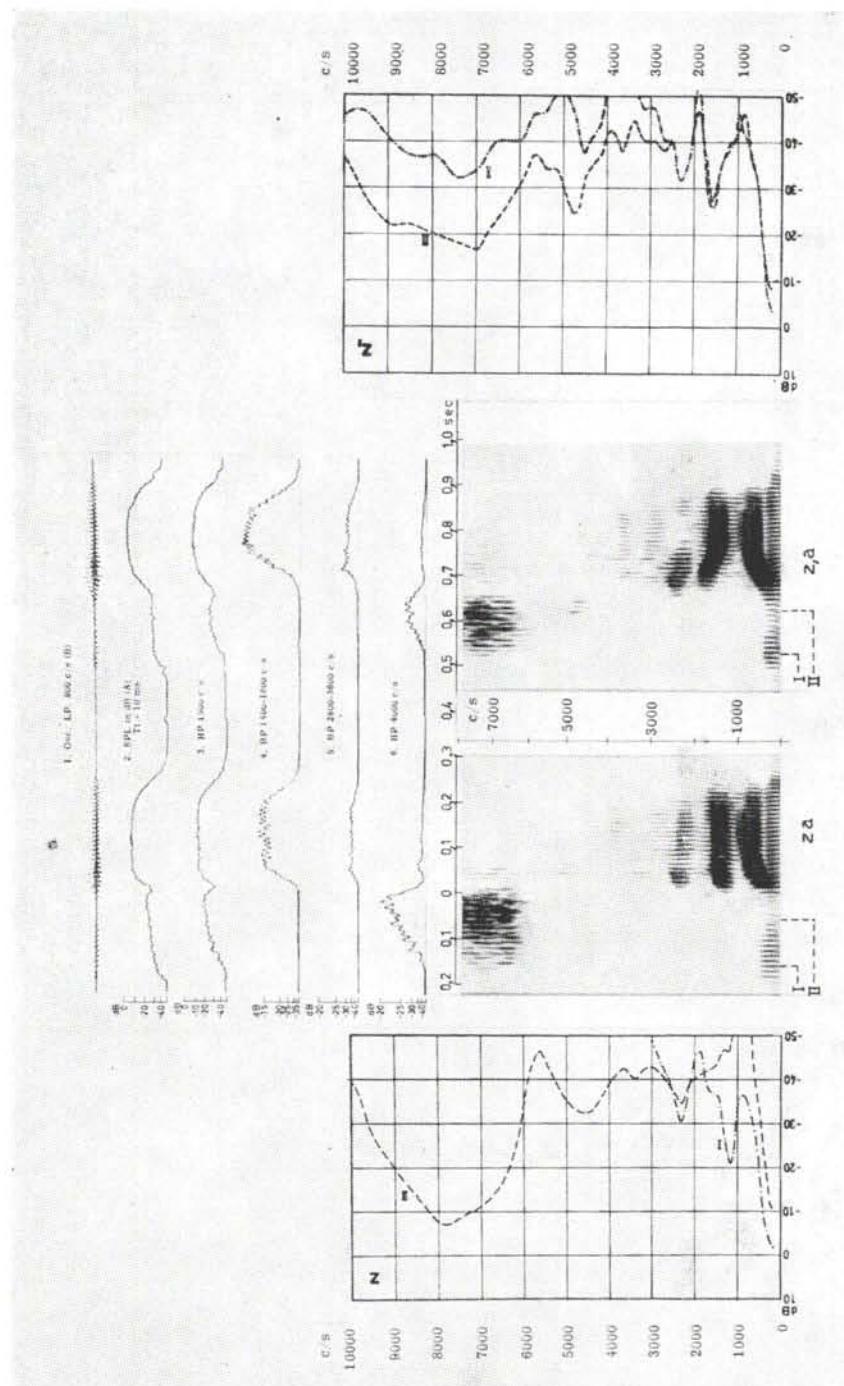


Fig. A.13-9. [za], [za].

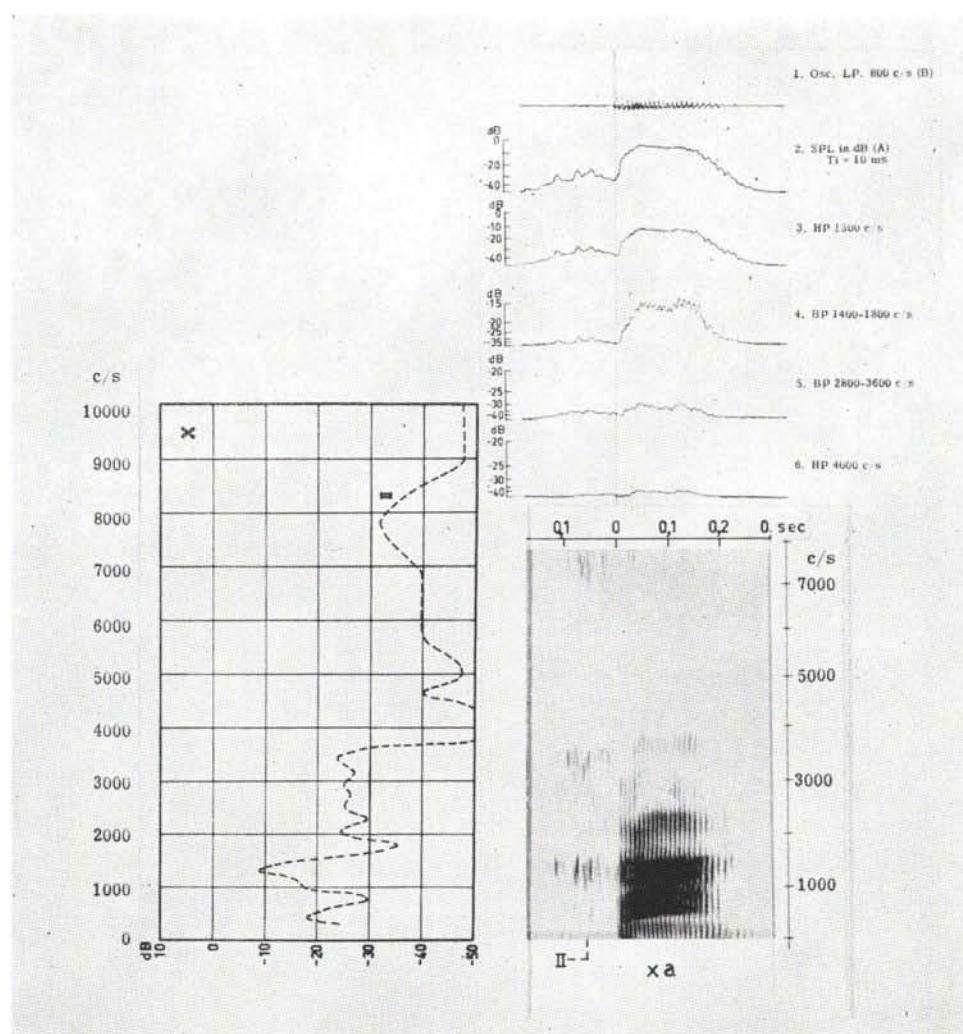


Fig. A.13-10. [xa].

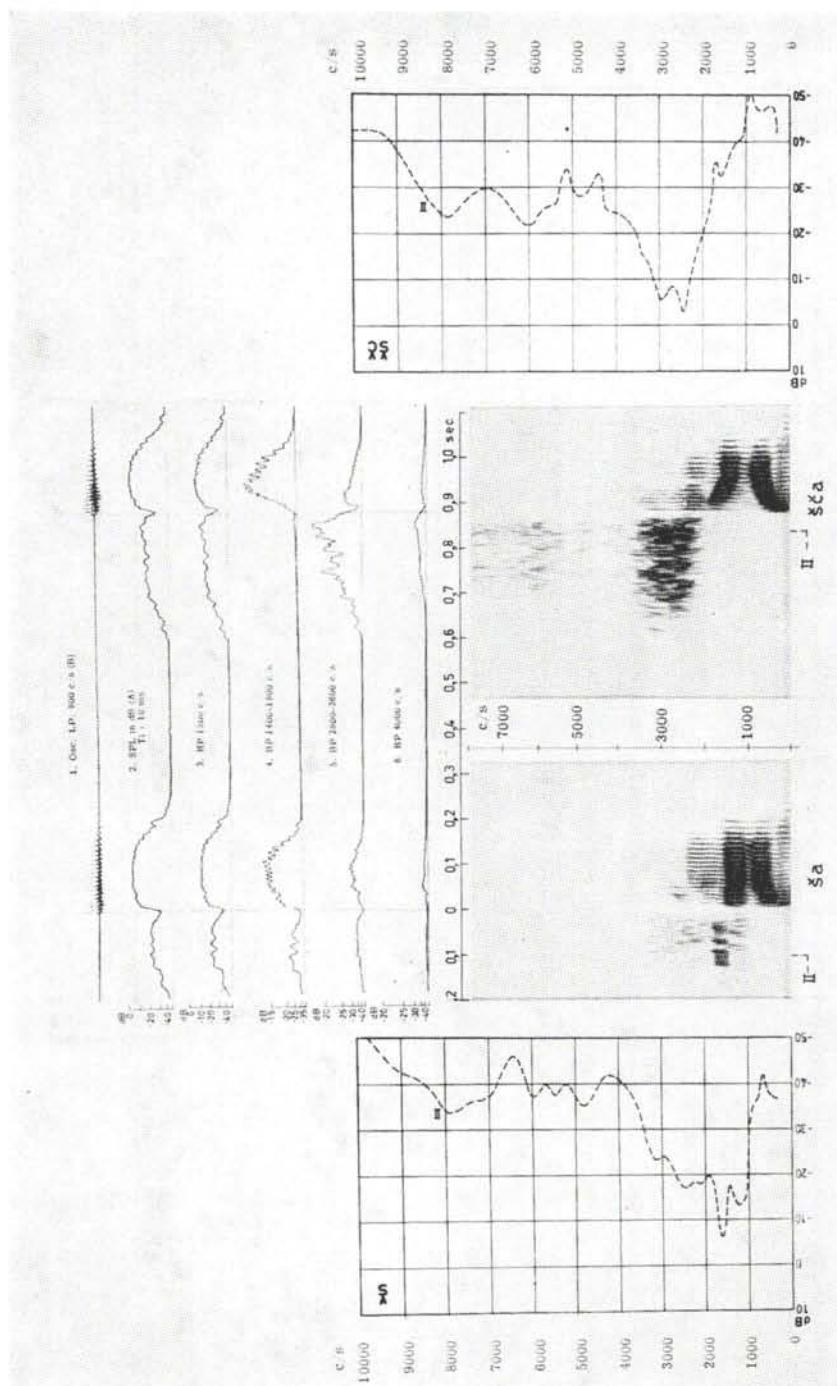


Fig. A.13-11. [šča], [šča].

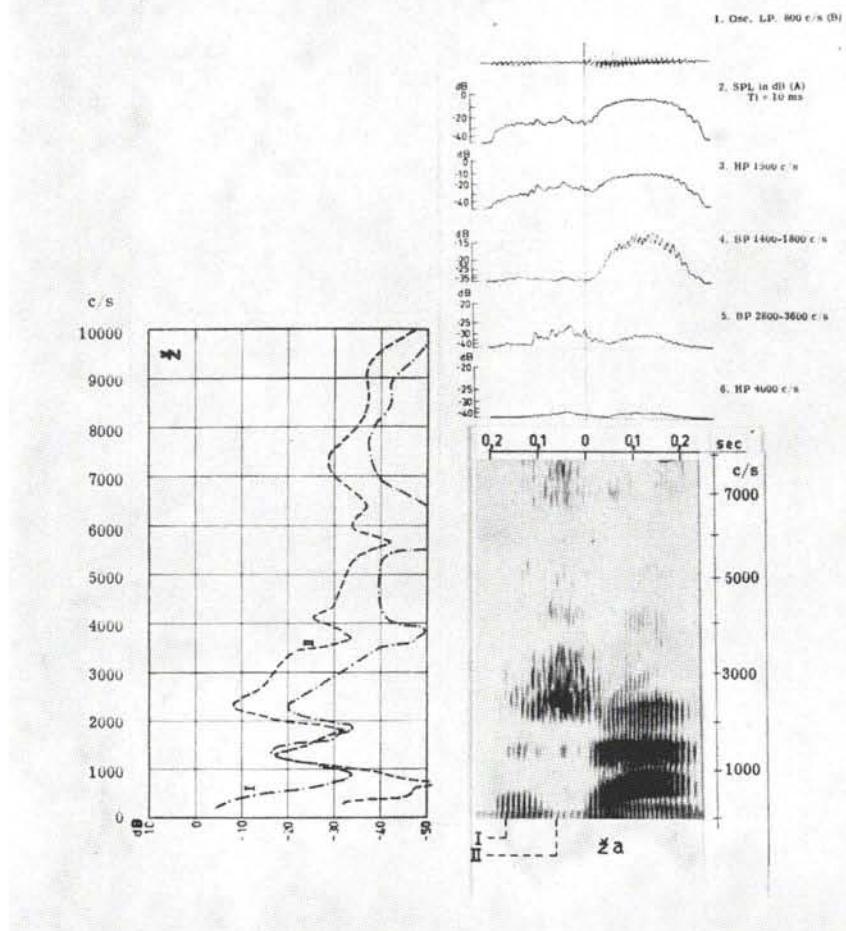


Fig. A.13-12. [ža].

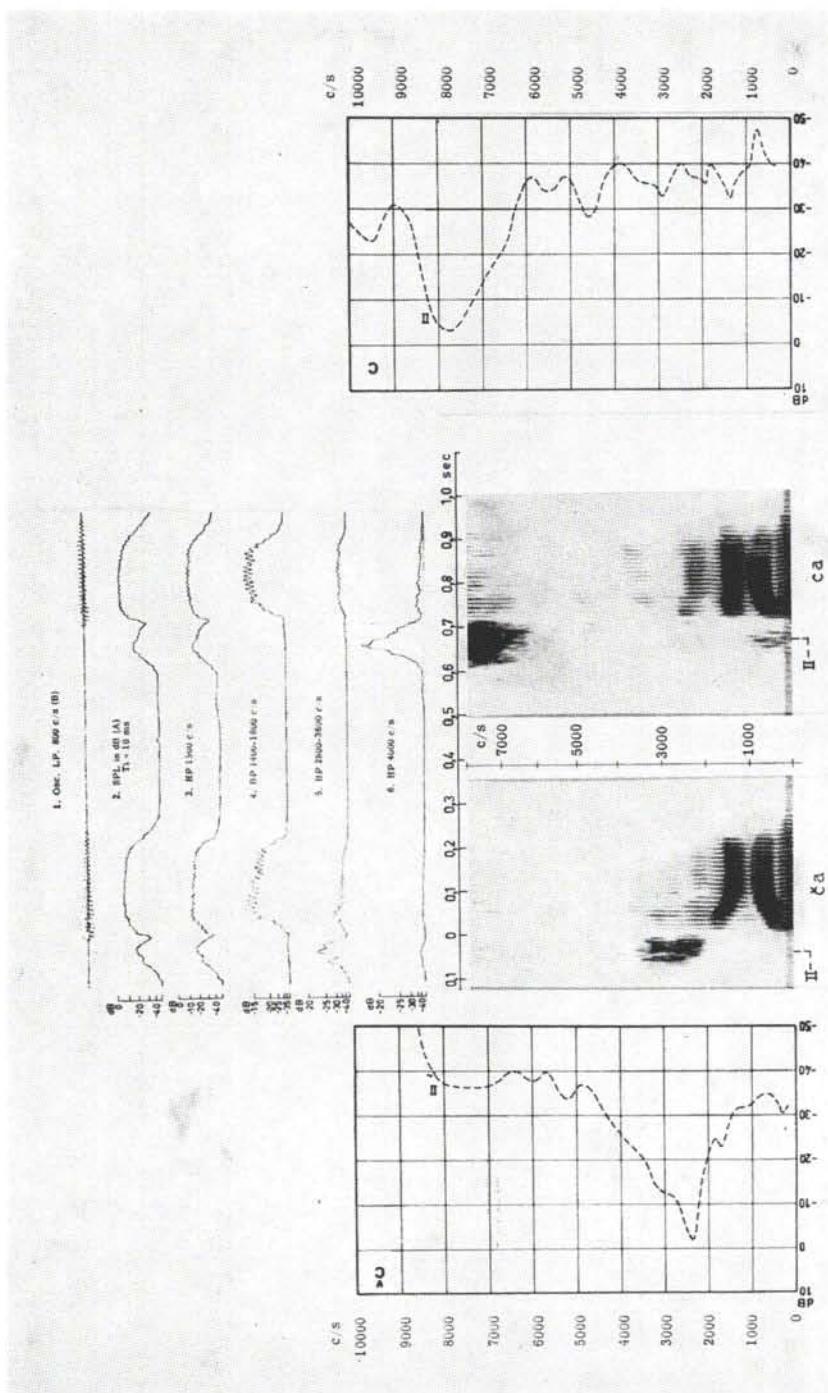


Fig. A.13-13. [ča], [ca].

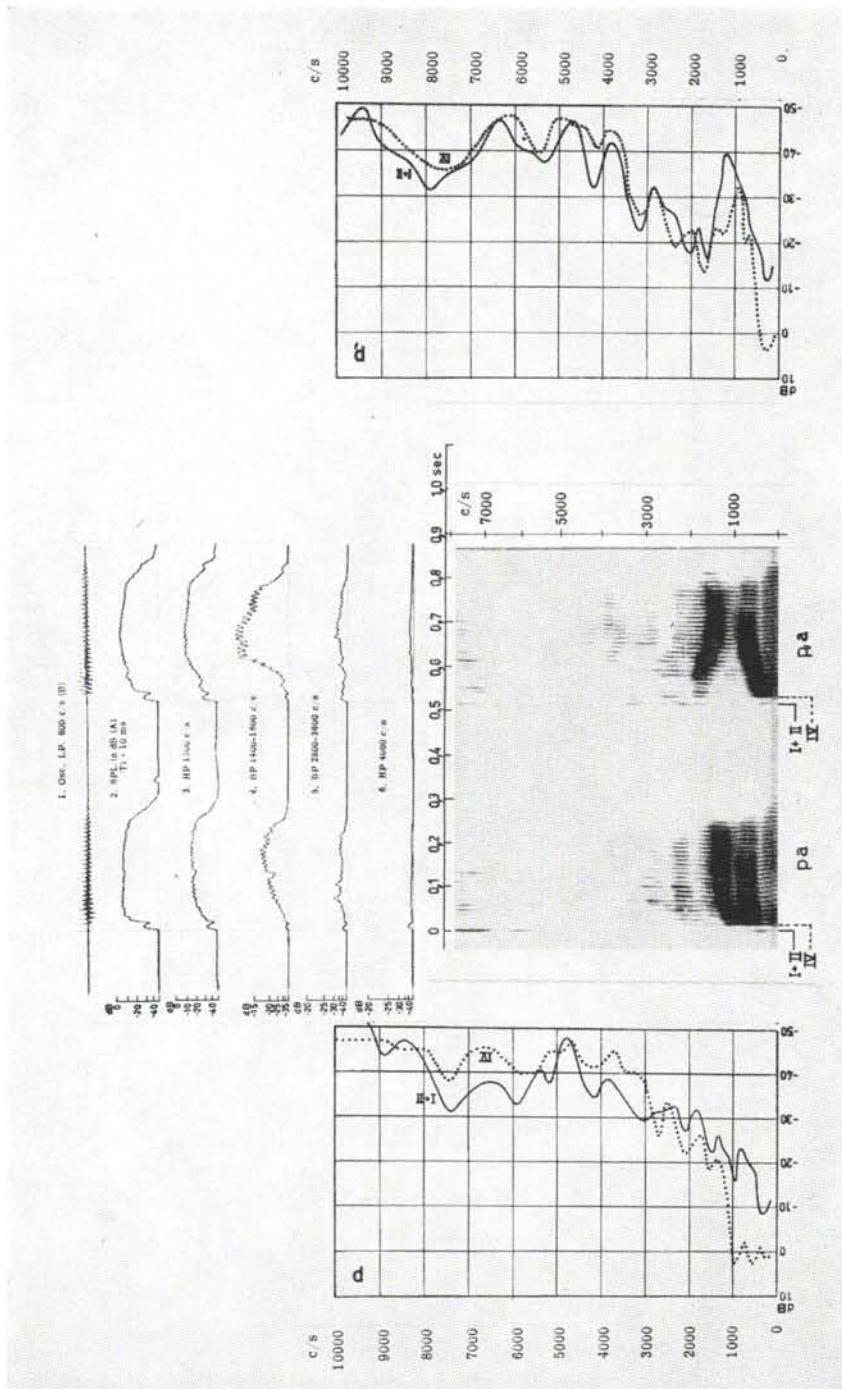


Fig. A.13-14. [pa], [p'a].

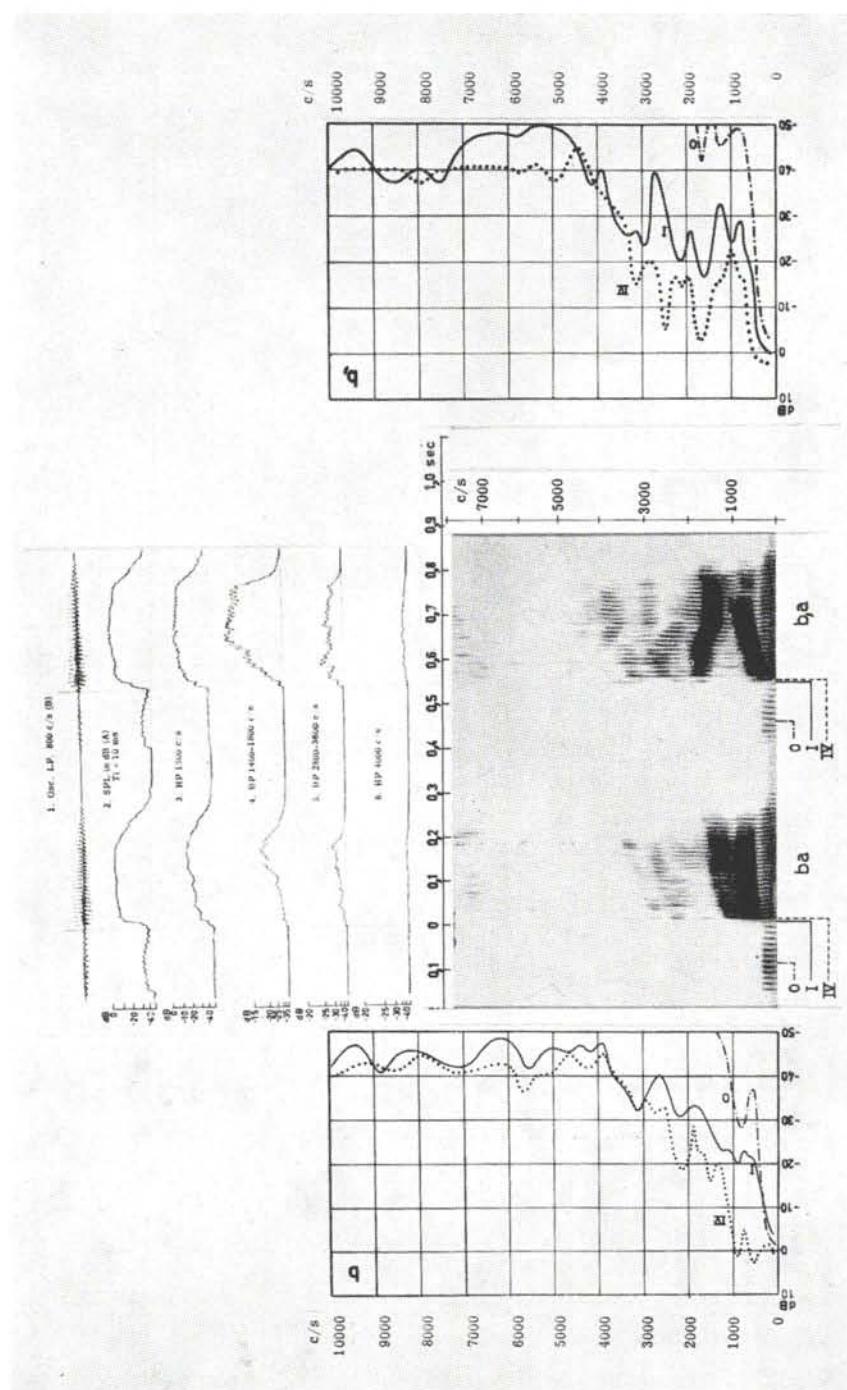


Fig. A.13-15. [ba], [a].

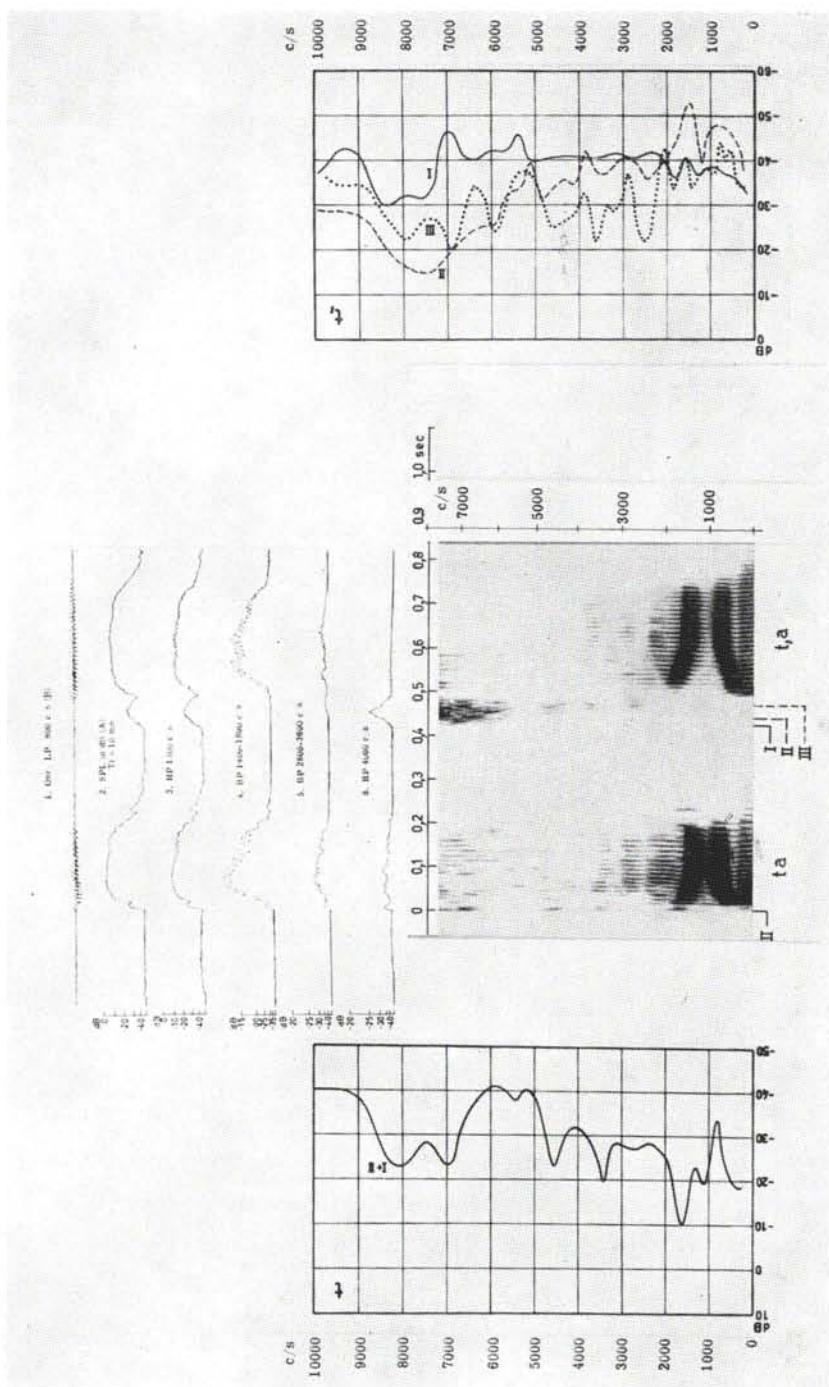


Fig. A.13-16. [ta], [t,a].

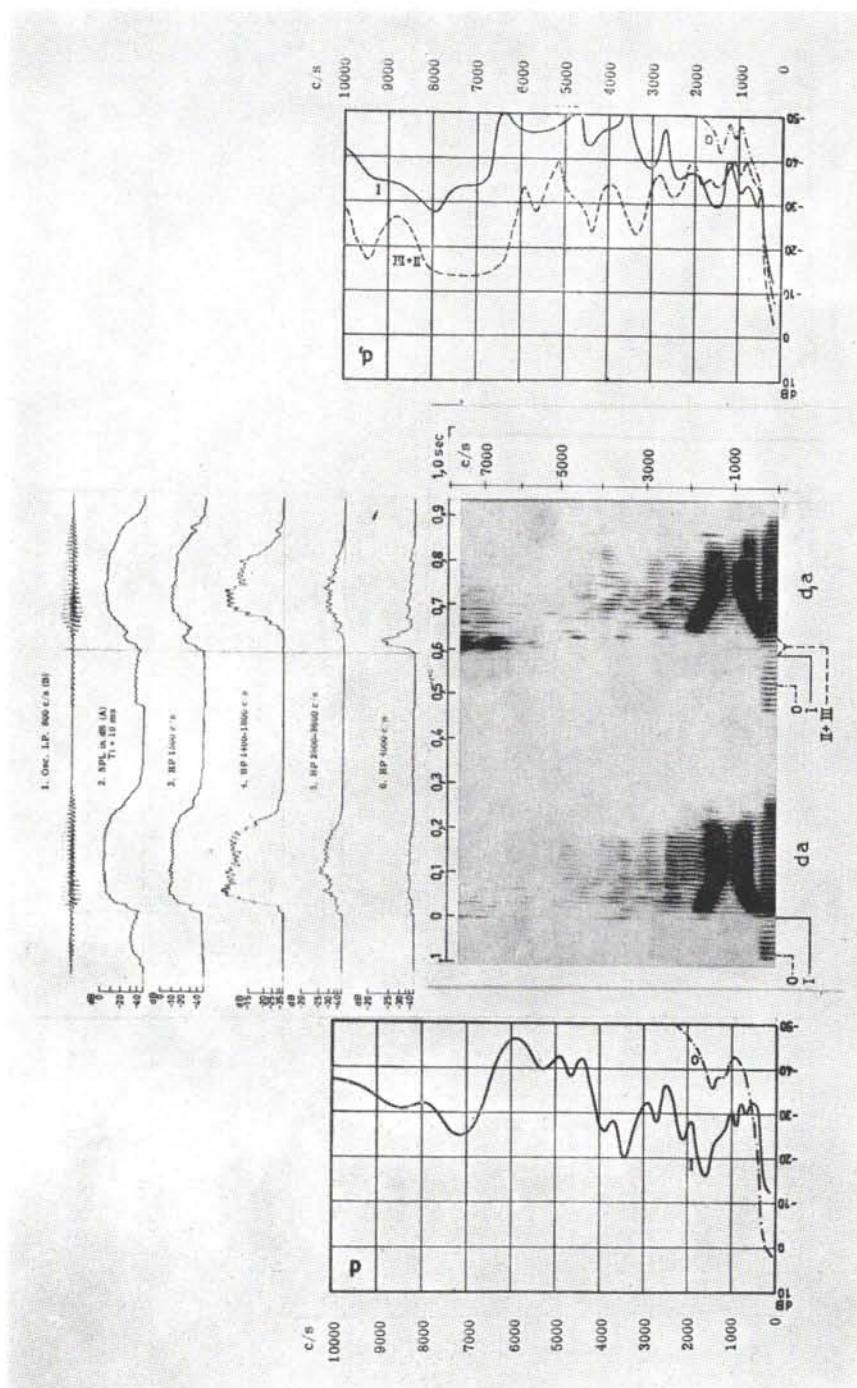


Fig. A.13-17. [da], [d.a].

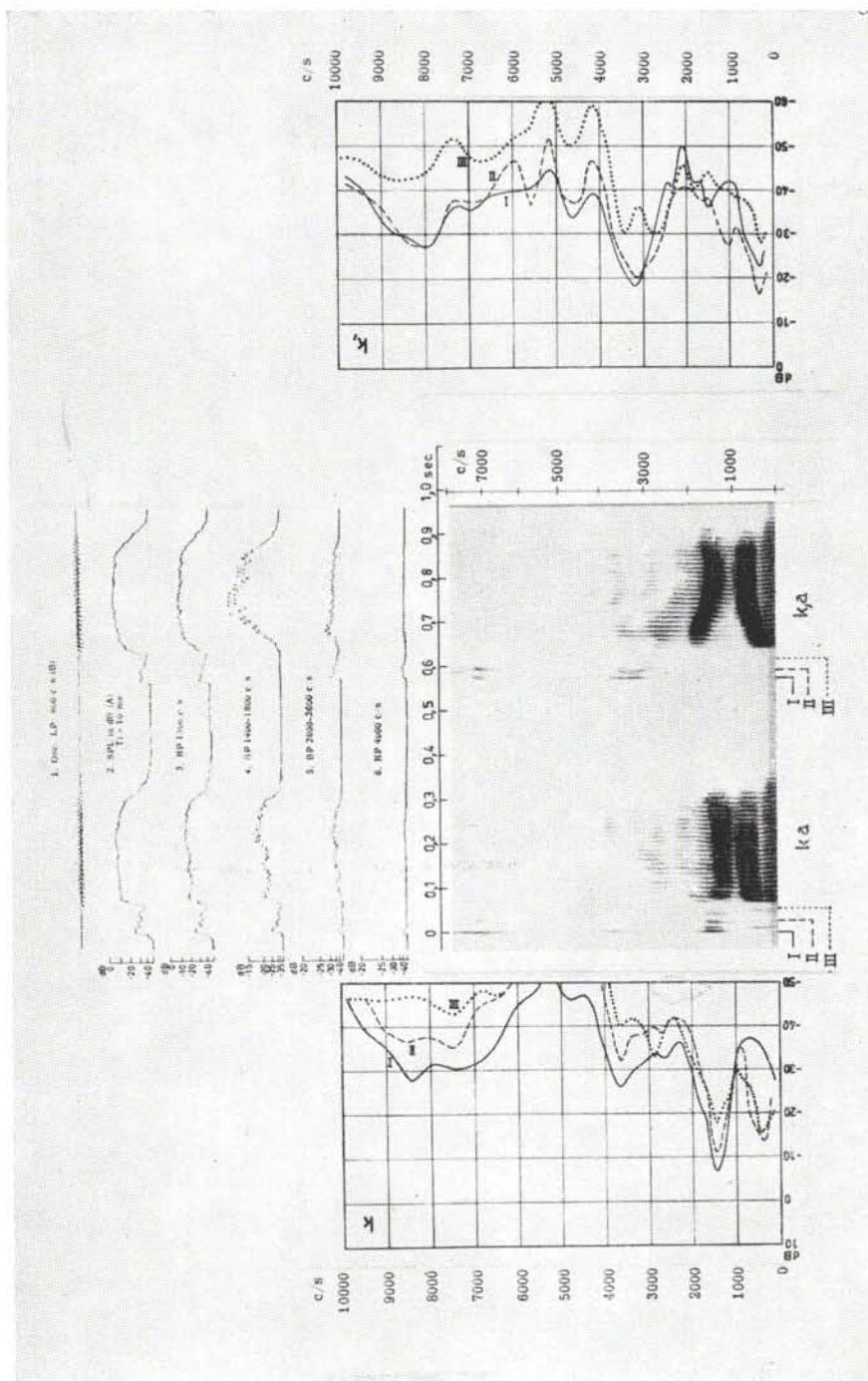


Fig. A.13-18. [ka], [k,a].

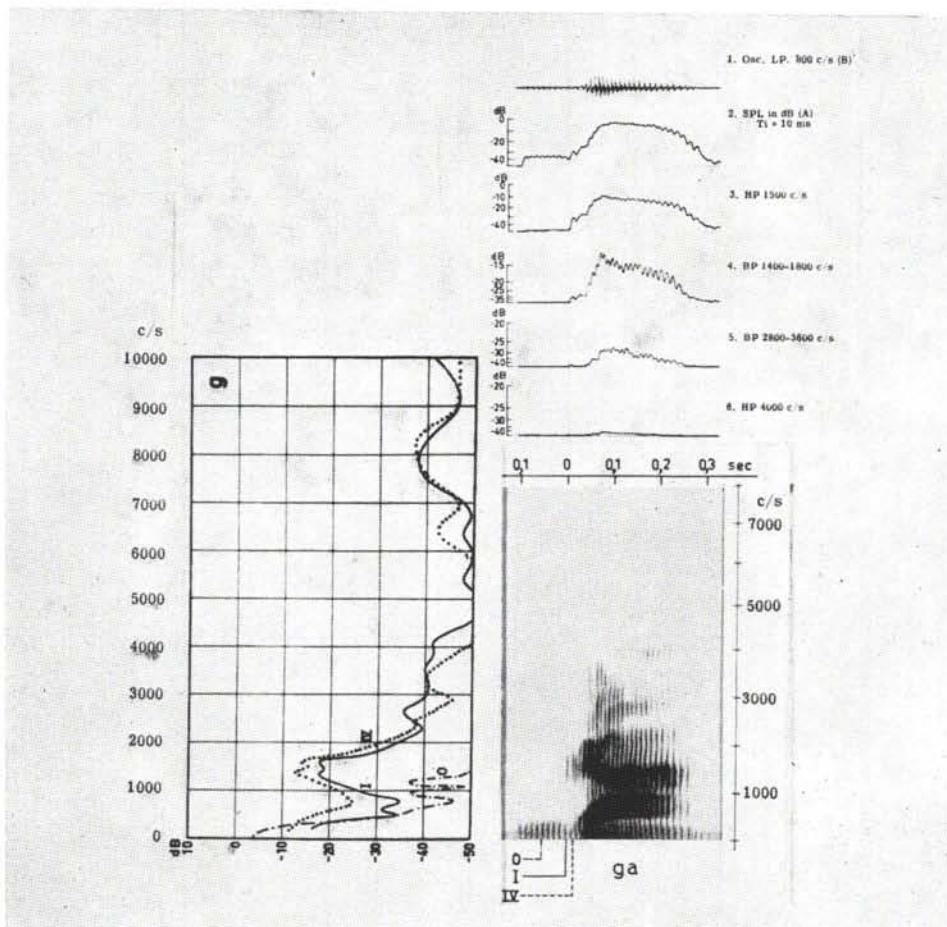


Fig. A.13-19. [ga].

A.2 A STUDY OF SOURCE CHARACTERISTICS

A.21 *The Voice Source*

The primary source of energy for the production of voiced sounds is the contraction of the respiratory muscles resulting in an over-pressure in the lungs and thus in an airflow that is periodically varied in magnitude owing to the opening and closing action of the so-called vocal cords, more recently called the vocal folds, once each fundamental voice period. The acoustic function of these cords or folds should not be regarded in analogy to vibrating membranes. Actually, they cause a modulation of the respiratory air stream, but do not generate sound oscillations of a significant magnitude by a direct conversion of mechanical vibrations to sound.

There are two quite opposite theories concerning the vibrational mechanism. According to Husson (1950), the laryngeal muscles are activated at a rate of the voice fundamental frequency. The vocal cords should accordingly execute forced vibrations in synchrony with the impulse frequency of a centrally determined nerve current. As judged from the model experiments of Smith, who has made successful attempts at imitating the human voice artificially, and from the additional experimental and theoretical evidence put forward by Smith (1954) and by van den Berg (1954, 1954, 1955, 1955a,b, 1957; van den Berg et al., 1957) and from the electromyographic investigations of Faaborg-Andersen (1957), that theory represents a less probable explanation. A simple mechanical explanation can be given on the basis of the alternating force exerted on the vocal folds by the subglottal over-pressure in the closed state and by the negative pressure in the glottis in the open state due to the flow of air. The latter sucking force, the Bernoulli effect, explains why the vocal folds can depart from an initial open state without muscle action. This happens at a soft onset of voice in the chest register, as shown by the Bell Telephone Laboratories *High speed Motion Picture of the Vocal Cords* and discussed by Farnsworth (1940) and by Smith (1954). At a hard onset of the voice the vocal lips are initially pressed against each other and then blown apart. Once in the open state, the vocal lips are brought back to closure by the Bernoulli effect in combination with elastic forces. Earlier versions of the mechanical

theory acknowledged only the latter restoring effect. The sucking effect is, however, alone sufficient to explain the mode of vibration in the chest register.

At a low pitch, the vocal folds have been observed to move from the closed to the open phase by an opening that progresses upward and outward. According to Farnsworth (1940) and Smith (1954), the lower portion is also the first to close. This phase difference of the motion in a vertical direction becomes less pronounced as the pitch is raised, owing to the greater stiffness of the vocal cords and the smaller effective mass-participating in the vibration at the higher pitch. In falsetto register it is mainly the upper edges that participate.

The stiffening of the vocal cords is accomplished by a stretching of their visible part to a greater length, but this is actually followed by a shortening of the maximum length of the glottis opening, i.e., the air passage between the vocal folds (Smith, 1954).

The glottis slit has an effective length of the order of *12 mm* in the chest register and the maximal width is of the order of *2.5 mm* at a moderate voice effort. The depth of the passage in the direction of the air stream that comes into contact during the closed phase is of the order of *2.5 mm*.

Model experiments directed toward the determination of the size of the Bernoulli effect have been made by van den Berg et al. (1957) who have mapped the spatial pressure distribution in the direction of the air stream within fixed larynx models of various glottis cross-sectional areas. For very small values of glottis width, the resistance in the glottis slit is large enough to maintain a positive pressure. For glottis widths above *0.2 mm*, the mean glottis pressure was negative and of the order of one-quarter, and the spatial maximum one-half of the subglottal pressure. From this experiment it can be concluded that the negative pressure of the open glottal phase is directly proportional to the subglottal pressure and of a magnitude large enough to cause an appreciable sucking effect.

From a simplified mechanical analysis disregarding the elasticity of the vocal cords, it follows that the time it takes for the cords to blow apart and come back, is inversely proportional to the square root of the subglottal pressure and proportional to the square root of the vibrating mass and to the small distance the lips have to move away before the mean pressure in the glottis switches to a negative value. An increase of the subglottal pressure will apparently have the same effect on the duration of the voice period and thus on the fundamental frequency as an increase of the elasticity.

In speech the increase of voice intensity from increased subglottal pressure will accordingly be followed by an increase in the voice fundamental frequency if the normal compensation of a decreased tension of the vocal cords is put out of function. As has been demonstrated by van den Berg (1957), the effect may be manifested by striking a person in the stomach while he is singing a steady tone (van den Berg, 1957).

A related but opposite phenomenon is the pitch fall at voiced occlusives in connected speech. If the flow resistance at the upper articulatory constriction consumes a noticeable part of the subglottal pressure, then a reduced Bernoulli effect will result at the glottis. Because of the smaller restoring force opposing the momentum of the

vocal lip excursions it must be concluded that the glottis slit will reach a greater maximum amplitude and that the restoring time will be increased.

An incomplete closure may occur at a soft onset or decay of the voice in consequence of an incomplete inward movement of the vocal folds, or it may be a constant feature in the case of a breathy voice, for instance, in the form of a leakage in the cartilaginous glottis which normally should remain closed even during the open phase. The existence of a phase of complete closure is, of course, not essential for voice production since any periodic disturbance of the airflow constitutes a sound source, the unmodulated part of the exhaled air representing an energy waste.

The spectrum of the voice source can be calculated from data on subglottal pressure during phonation and the time-varying conductivity of the glottis, as determined from the time-varying area dimensions of the passage between the vocal folds. In an electrical analog circuit the pressure in the lungs may thus be represented by a constant *DC* voltage in series with the glottis impedance and the input impedance to the vocal tract seen from the glottis. To a first order of approximation the subglottal pressure is constant which implies that the impedance of the subglottal system may be regarded as small compared with the glottis impedance. The actual flow through the glottis is to some extent also influenced by the vocal tract impedance.

A clearer separation of source and filter characteristics is obtained if the lung pressure and glottis impedance are converted into a constant current source feeding into the input terminals of the vocal tract in a parallel manner with the glottis impedance. This operation is not strictly representative because of the non-linear glottis impedance, but it gives a better approximation than the infinite source impedance treatment since the boundary conditions are retained. The model experiments of van den Berg et al. (1957) show that the glottis flow resistance R_F as a function of glottis area A and particle velocity $v = u/A$, can be decomposed into two terms $R_F = R_L + R_T$, R_L being proportional to A^{-3} and independent of the flow and R_T being proportional to A^{-1} and v . The former is the resistance of a very narrow slit, assuming laminar streaming

$$R_L = \frac{12\mu l}{A^3/b^2} \text{ (dyne sec/cm}^5\text{)}, \quad (\text{A.2-1})$$

where $\mu = 1.84 \cdot 10^{-4}$ is the coefficient of viscosity. The glottis cross-section is assumed to be rectangular and of the width $a = A/b$ across the slit and of the length $b = A/a$ in the direction of the slit. The depth of the slit is l .

When the glottis area has reached about $1/6$ of its maximum value, the second term R_T obtains equal magnitude and dominates at higher area values. This resistance is due to turbulent losses and was found to be $7/8$ of the resistance R_B (see below) associated with the kinetic pressure of the Bernoulli equation

$$p = \rho v^2/2, \quad (\text{A.2-2})$$

where p is the pressure fall at the constriction. The resistance is

$$R_B = p/u = \varrho v/2A = \varrho u/2A^2. \quad (\text{A.2-3})$$

In addition to the resistive elements there is a glottis inductance

$$L_q = \frac{\varrho l}{A} \quad (\text{A.2-4})$$

to be considered.

The source volume velocity $u(t)$ shall, by definition, develop the constant subglottal pressure p_q in the source impedance:

$$R_F(t) \cdot u(t) + L_q(t) \frac{du(t)}{dt} = p_q. \quad (\text{A.2-5})$$

If the resistance term R_L is ignored, $R_F \approx R_B$ and the value from Eq. A.2-3 may be inserted:

$$\frac{\varrho}{2} \frac{u^2(t)}{A^2(t)} + \frac{\varrho l}{A(t)} \cdot \frac{du(t)}{dt} = p_q, \quad (\text{A.2-6})$$

or

$$\frac{\varrho}{2} v^2(t) + \frac{\varrho l}{A(t)} \cdot \frac{d[v(t) \cdot A(t)]}{dt} = p_q. \quad (\text{A.2-7})$$

The solution to this differential equation is not simple unless either the resistance or the inductance term dominates. In the former case the particle velocity $v(t)$ is a constant, $\sqrt{2p_q/\varrho}$. Thus

$$u(t) = A(t)\sqrt{2p_q/\varrho}, \quad (\text{A.2-8})$$

and the source flow becomes proportional to the glottis area. If the resistive term is negligible,

$$u(t) = \frac{p_q}{\varrho l} \int A(t) dt. \quad (\text{A.2-9})$$

The integral implies that the spectrum of the source flow slopes 6 dB/octave faster than under purely resistive conditions. Under normal voice conditions, assuming both a resistive and an inductance impedance term, the crossover from the resistive to the inductive conditions is estimated to occur at an approximate frequency of 2000 c/s.¹ This additional low-pass filtering effect is in part counteracted by the sharpening of the onset and the decay of a glottal pulse owing to the transition from laminar to turbulent streaming. However, the area dependency of the flow resistance in the laminar region is probably not so negative as the A^{-3} proposed by Van den Berg et al. (1957), since there are glottis width variations in the chest register in addition to the length variations.

¹ A more detailed analysis of this problem, following a similar approach, has been undertaken by J. L. Flanagan (1958). His data support the findings above.

The assumption of a resistive term of turbulent origin governing the major part of the airflow checks very well with measurements. According to Chiba and Kajiyama (1941), there is a normal air consumption of $140 \text{ cm}^3/\text{sec}$ at a subglottal pressure of $p_g = 16 \text{ cm H}_2\text{O}$, i.e., $16 \cdot 980 = 15700 \text{ dyne/cm}^2$ during phonation at medium intensity and $F_0 = 144 \text{ c/s}$. From Eq. A.2-2 the corresponding particle velocity is 5200 cm/sec and the mean glottis opening 0.027 cm^2 . On the basis of a peak factor of 3.75, given by these authors, relating peak to average flow during a period, the maximum glottis area reaches 0.10 cm^2 . Assuming elliptical shape and 12 mm width the maximum breadth would be 1.1 mm , which is of the right order of magnitude but rather small. Van den Berg derives, by a similar calculation, a maximum glottis breadth of 1.3 mm , assuming $p_g = 10 \text{ cm H}_2\text{O}$ subglottal pressure and $150 \text{ cm}^3/\text{sec}$ flow. However, these theoretical values are smaller by a factor of 2 than what can be directly measured from the BTL film. The neglected glottis inductance may, in part, account for this discrepancy.

During the open phase the glottis resistance consumes a part of the oscillatory power corresponding to a formant. The mean amplitude of this damped oscillation leaking through the glottis is small compared with the mean amplitude of the glottal airflow and decreases with an increase of formant frequency essentially in conformity with the voice source spectrum envelope. It is thus necessary to adopt the differential resistance for estimation of the formant damping. From a differentiation of A.2-2 it can be seen that the differential resistance $R_D = \Delta p / \Delta v A$ is twice the flow resistance $R_B = p/vA$ (Westervelt, 1950):

$$R_D = \frac{\Delta p}{\Delta v A} = \rho u / A^2. \quad (\text{A.2-10})$$

This relation apparently holds whenever the flow resistance is proportional to the volume velocity. On the basis of the above-mentioned data, the average differential resistance is thus $2 \cdot 15700 / 140 = 225$ acoustical ohms or $5 \rho c$.

A substantial reaction from a supraglottal constriction on the airflow presumes a narrowing to a greater extent than is generally found in vowels. The major part of the flow resistance in voiced sounds is thus confined to the glottis. This holds also true for most voiced consonants. The maximum reaction of the supraglottal system on the vocal cords occurs at the voiced interval of complete closure preceding a voiced stop, during which air is forced into the oral cavities causing the walls to expand. Besides the pressure loss at the glottis associated with a supraglottal source resistance the effect of the articulatory narrowing will be to tune F_1 to a lower frequency. If F_0 coincides with F_1 there results an increased resistive load on the vocal cords from the vocal tract input impedance which is resistive and high at the frequency of a formant. This reaction has been studied in detail by van den Berg (1954). However, the associated changes of the voice source spectrum are still to be studied. Low frequency emphasis is to be expected.

The flow pulsations at a narrow supraglottal constriction are essentially in phase

with the glottal pulse train but are somewhat delayed and smoothed out, primarily owing to the low-pass filter effect of the vocal tract fundamental resonance, i.e., F_1 . The instantaneous value of the flow resistance at the articulatory constriction is at a maximum during the peak of the flow and can be estimated from *Eq. A.2-3* with the understanding that the particular shape and surface conditions also enter as determinants, though to an unknown degree, as further discussed in *Section A.22*. The peak of an air pulse from the glottis is associated with a minimum resistance at the glottis due to a maximum opening, and with a maximum resistance at the constriction due to a maximum velocity, at both places causing a maximum of formant damping within the voice period; see *Section A.36-A.3*.

At the articulatory constriction, the oscillating flow of the frequency of the first formant will be smaller than, or comparable in size with, the non-oscillatory voice pulsations, the difference in level decreasing with decreasing F_1 . Since the transfer from volume velocity through the lips to the pressure in the soundfield is proportional to the frequency it is possible to estimate the constriction flow spectrum at and below the frequency of the first formant by an integration, i.e., a -6 dB/octave correction applied to the particular sound pressure spectrum of the acoustic field in front of the speaker. Only if the oscillating flow of formant frequency is small relative to the mean value of the airflow will the differential resistance be as high as twice the flow resistance.

Available data on glottis area and airflow variations within a fundamental period have been utilized for the calculation of corresponding source flow spectra.

The results are summarized in *Fig. A.2-1*, where the first two curves pertain to data obtained from the Bell Telephone Laboratories high-speed motion film. Each complete voice period of the flow curve has been constructed from thirty successive picture frames, and it was assumed that the flow was proportional to the glottis area. The spectrum envelopes were calculated with the aid of a mechanical Fourier analyzer.

It can be seen that the spectrum envelopes fall at an approximate rate of 12 dB/octave and that the rate of fall above 500 c/s is greater for the lower voice effort, as could be expected, because of the less abrupt transitions from the closed to the open phase and vice versa. The relative opening time defined by the ratio of the duration of the open glottis state to the period duration affects primarily the lower part of the spectrum, including the voice fundamental and the direct current component, i.e., the amount of air exhaled during a period.

From earlier measurements of the spectra of voiced sounds at varying intensity it is known that there are appreciable variations of the average spectrum slope. As a rule of thumb, a change of voice level at constant pitch which causes the level of the first formant to increase 10 dB will cause an increase of 4 dB in the level of the voice fundamental. Similar relations hold for the shift from a medium to a low voice effort.

These findings conform with those of *Fig. A.2-1* except for the change of the calculated voice source spectrum slope below 1000 c/s which is opposite to that above 1000 c/s . The expected change in glottis impedance from essentially resistive, during

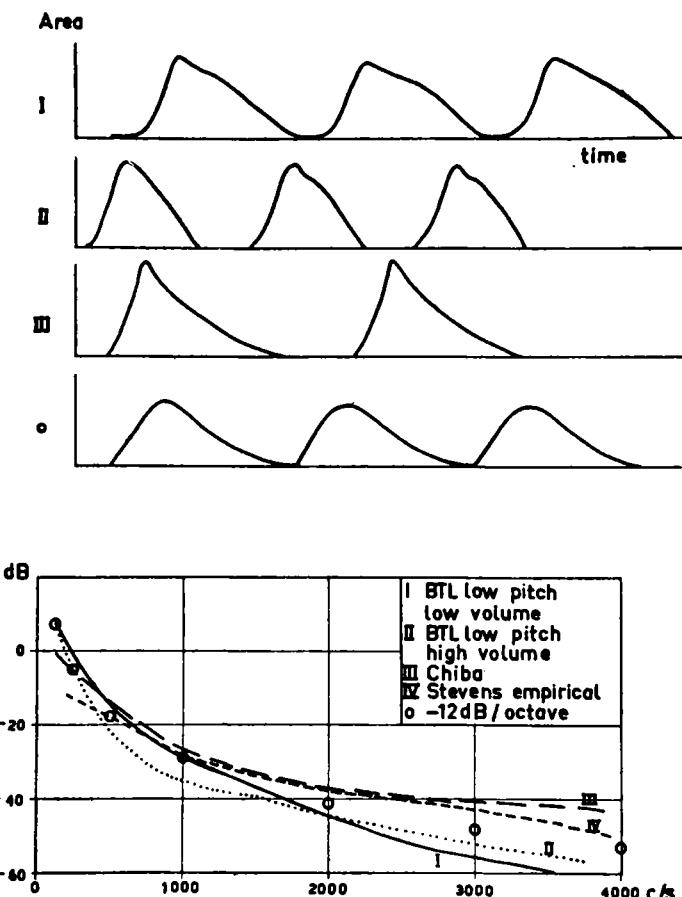


Fig. A.2-1. Calculated waveforms and spectrum envelopes of the voice source. Curves I and II have been derived from area measures taken from the Bell Telephone Laboratories film of the vocal cords. Curve III has been derived from the glottis area versus time pictures given by Chiba and Kajiyama (1941) and the curve marked o is a critically damped exponential wave characterized by the -12 dB/octave spectrum envelope slope adopted as a standard for calculations in this work. The spectrum envelope of the voice source utilized by Stevens et al. (1953) in early experiments on an electrical vocal tract is also shown for comparison.

high voice levels and turbulent air-streaming, to essentially inductive at a very low voice level and laminar streaming, as discussed above, might in part account for the discrepancy. This impedance change alone causes a -6 dB/octave drop of the spectrum. This is merely a hypothetical suggestion, and further experiments and calculations are needed as a basis for a more rigid theory.

The voice source spectrum utilized by Stevens et al. (1953) in the early work on the M.I.T. line analog is included for comparison in *Fig. A.2-1*. It differs from the other

data by a less rapidly sloping spectrum envelope in the frequency range 100-1000 c/s. It gives a sharper voice quality than the more low frequency emphasized source spectra. Fig. A.2-1 also contains the standard source of -12 dB/octave constant slope adopted as a standard for the calculations in this work. There are indications that this source provides fairly natural speech except that the levels of the third and higher formants may become somewhat too high.

Experiments have been reported by R. L. Miller (1956) and by Chang (1956) on the derivation of the waveform of the glottal airflow by a technique of inverse filtering. The formants may accordingly be removed by anti-resonances.² Our own experiments along these lines have provided preliminary results which indicate that the glottal flow may be studied in appreciable detail in spite of the indirect method of investigation. The waveform variations observed confirm the general statements made above. The relative role played by the opening and closing transients and the flow peak as determinants of the source spectrum still remains to be ascertained for different voice categories.

The efficiency of voice generation is not very great. According to van den Berg (1956), the overall efficiency, defined as the ratio of radiated acoustic power to the product of the subglottal over-pressure and volume velocity of the air exhaled during phonation, rises with voice effort from $0.45 \cdot 10^{-5}$ at a sound pressure level of 55 dB at 25 cm speaking distance to the value $45 \cdot 10^{-5}$ at 40 dB higher sound pressure level. He also found that the volume velocity varied in proportion to the subglottal pressure and that, at constant voice fundamental frequency, the sound pressure of the radiated sound increased with the square of the volume velocity of the airflow.

From the studies quoted earlier relating to spectrum changes following an increase of voice effort of constant F_0 (Fant, 1949), it is apparent that the amplitude of the voice fundamental varies approximately with the square root of the overall sound pressure amplitude. Since the mean flow, i.e., the harmonic number zero of the glottis source spectrum, should vary in a fashion similar to the amplitude of the voice fundamental, there is thus reason to conclude that the increased efficiency at higher voice efforts is largely a matter of a shortening of the air pulses within a fundamental period of fixed duration and a sharpening of the onset, the peak, or the closure.³

A.22 Turbulent and Transient Sources

Our present knowledge of the physical characteristics of noise sources is less extensively developed than that of the voice source. Sustainable sounds of noise character are produced from a turbulent source located at or near a constricted passage within the

² Experiments performed along these lines by Lawrence of SRDE, England, with his "speech microscope" indicate that the damping constant (bandwidth) of the first formant varies within a voice fundamental period. (Personal communications.)

³ Van den Berg is in favor of a filter-function explanation rather than the source function origin suggested here. He emphasizes the importance of a natural correlation between voice effort and mouth-opening. However, such articulatory changes would have to be rather great and would cause a substantial increase of F_1 and thus of phonetic quality.

vocal tract. The contraction of the flow causes the air particles to accelerate, forming a jet of air shot at high speed through the passage. The jet is associated with circulation effects and eddies, partially of a random nature. The place within the constriction where they are created is dependent on the flow and the geometry including surface conditions. An obstacle hit by the jet of air will give rise to a turbulent source that can be of greater intensity than the noise produced in the passage. This would be the role, for example, of the upper incisors in the production of dentals.

The production of eddies around a free obstacle at low flow rates can give rise to a whistling noise of a quasi-periodic nature referred to as *Aeolian tones* or *Schneidetöne* (Meyer-Eppler, 1953). The tendency towards periodicity is increased if the source is enclosed by a cavity resonator. At low flow rates, the source will be influenced by one of the resonance frequencies of the cavity that determines the periodicity of the eddies. Seen from the source, the impedance of the cavity system is low and resistive at a resonance frequency, and the flow has the effect of a negative resistance causing the system to oscillate. This is the case with either ordinary lip-whistling or whistling between the teeth. However, at an increased rate of flow the periodicity can no longer be maintained, and the whistling changes into a random noise source. The source now rules the cavities.

Of basic importance for the discussion of turbulent streaming and noise is the Reynold's number

$$Re = \frac{vh}{\nu}, \quad (\text{A.2-11})$$

which is a dimensionless parameter proportional to the particle velocity v cm/sec and to the effective width h cm of the passage. The constant $\nu = 0.15$ cm²/sec is the kinematic coefficient of viscosity defined as the ratio of the viscosity coefficient to the density of the gas. Depending on the particular geometry and surface properties of the passage, there is a critical Reynold's number above which turbulence sets in.

In a relatively short constriction the turbulent flow resistance is much more influenced by the minimum area than by the length. According to measurements performed by Heinz (1956), the major part of the pressure drop along a narrow tube of length 5 cm and diameter 0.25 cm inserted as a constriction in a mechanical vocal tract model, will occur at the posterior end hit by the air stream. Inside the narrow tube the pressure falls off at a slower rate.

These data seem to agree within 10 per cent with those predictable from the theory on pressure drop within pipes of a length large enough to insure homogeneous flow. According to Schlichting (1951); Eck (1944), the pressure drop per unit length under these ideal conditions is

$$\frac{\Delta p}{\Delta l} = \frac{\lambda}{d} \cdot \frac{\rho}{2} \cdot v^2, \quad (\text{A.2-12})$$

where $\lambda = 0.3164 Re^{-1}$, $Re = \frac{vd}{\nu}$, and d is the diameter. The pressure drop is thus proportional to $v^{7/4}$. Under conditions of laminar streaming,

$$\frac{\Delta p}{\Delta l} = \frac{64}{Re \cdot d} \cdot \frac{\rho v^2}{2} = 32 \frac{\mu v}{d^2}, \quad (\text{A.2-13})$$

This formula could have been derived from *Eq. 1.1-8*.

The velocity-squared dependent pressure drop at the inlet can be written

$$p = \frac{\rho}{2} \left(\frac{u}{AK} \right)^2, \quad (\text{A.2-14})$$

where K is a constant, A the area, and u the volume velocity. The average particle velocity is smaller than in the center of the air stream because of the contraction of the flow. The constant K represents, according to Westervelt (1950), in part this effect and in part frictional losses. It was found to be of the order of 0.7-0.9 in his experiments on thin orifices, and the same order of magnitude was found by Heinz (1956) for the flow proportional term of the total resistance of his constriction simulating tubes. In speech production, however, an articulatory constriction is seldom abrupt, and it is then more probable that the constant K is closer to or smaller than 1, as was the case for van den Berg's et al. (1957) glottis model.

Meyer-Eppler (1953) has shown, by means of experiments with constricted plastic tubes and verified by experiments on the human production of the three fricative consonants [s], [ʃ], and [f], that there appears to be a critical Reynold's number Re_c for the onset of the source. The sound pressure P_l of the noise measured at some distance l from the speaker or the model could be put in the form

$$P_l = \alpha (Re - Re_c)^2, \quad (\text{A.2-15})$$

where α is a constant and Re_c was found to be of the order of 1800 for the plastic tube models, but smaller for the real speech sounds.

Under turbulent conditions the pressure drop p_d in the constriction, approximately given by *Eq. A.2-12*, is proportional to the square of the particle velocity v . For studies of speech production it is generally more convenient to measure the over-pressure behind the constriction than the particle velocity. *Eq. A.2-15* may accordingly be reformulated

$$P_l = k_1 h^2 p_d - k_2, \quad (\text{A.2-16})$$

where k_1 and k_2 are constants. The sound pressure of the generated noise is apparently proportional to the over-pressure or to the particle velocity squared, in excess of the particular threshold value, and further, to the effective width h . For sections of elliptical shape the constant h is approximately equal to $4A/S$, where S is the circumference and A the area. For the special case of circular sections, h equals the diameter and for an ellipse of larger eccentricity, h approaches a value $\pi/2$ greater than the smaller diameter. It is thus apparent, that for a given over-pressure and effective width, the air consumption is much smaller for a circular than for an elliptically shaped constriction with large eccentricity. Conversely, the Reynold's number is larger and the noise generation more efficient for a circular than for an elliptical

section shape, provided the power of expiration, defined by the product of over-pressure and volume velocity, remains constant. The large width to height ratio of the slit could accordingly be one of the factors responsible for the weak intensity of the interdental fricative [θ].

A few figures may be quoted here as an indication of the order of magnitude of the variables involved. According to Meyer-Eppler (1953) the minimum over-pressure⁴ needed for producing the fricatives [s], [ʃ], and [f] was $p_d = 1, 1.5$, and $6 \text{ cm } H_2O$ respectively, corresponding to the particle velocities $1300, 1600$, and 3100 cm/sec , and the volume velocities $130, 250$, and $500 \text{ cm}^3/\text{sec}$ respectively. At strong force of expiration the over-pressure was of the order of $20 \text{ cm } H_2O$ and the particle velocities of the order of 5000 cm/sec .

If a closed passage in the vocal tract is gradually opened, it is possible to maintain a constant over-pressure behind the constriction only at the expense of an airflow increase in proportion to the minimum area of the constriction. This requires an increasing power consumption from the expiratory muscles which is limited and thus determines the maximum rate of flow at a particular over-pressure. Conversely, the maximum over-pressure that can be maintained will decrease with increasing constriction area. The maximum value occurs at complete obstruction and is of the order of $130 \text{ cm } H_2O$.

These limitations on the variables of *Eq. A.2-16* imply that there exists an optimum effective width above which a further increase of h is followed by a decrease of the product $h^2 p_d$ and thus of the noise level.

An apparent implication of the flow dependency of noise sources is that voiced fricatives will be intensity modulated at the rate of the voice fundamental frequency, which is superimposed on the direct current component of the flow. More specifically, the noise pressure fluctuations will be dependent on the square of the flow variations above a threshold value set by the critical particle velocity for the generation of turbulent noise. Spectra of voiced fricatives may be studied in the *Appendix*. The vertical striations of the high frequency formants in a broad-band spectrogram of, for instance [z], are periodic and synchronous with the vertical pulses constituting the fine structure of the voiced formant bars $F1, F2$, and $F3$.

The fricative element of a stop sound is quite similar to a short fricative continuant of the same articulation, but the later part of the noise interval preceding the onset of voicing in aspirated stops is to be considered as a short [h]-sound, as previously mentioned. In general, the onset and decay characteristics of a turbulent sound and its duration are to a considerable extent influenced by the speed of movement of the articulators and the time they remain in a position appropriate for an effective noise generation. Another determinant of the temporal intensity characteristics is the position of the sound within a breathpulse from the expiratory organs.

⁴ The existence of a threshold of this order of magnitude as well as the representability of *Eq. A.2-12* has been verified recently by our own experiments. However, in a logarithmic plot of pressure versus flow the threshold effect is not very apparent.

The temporal characteristics of the flow following the release of the air accumulated during the build-up interval of a stop sound are also dependent on the volume of the particular air container and the initial over-pressure. The conditions for the flow are apparently analogous to the discharge of a condenser through a non-linear time-variable resistance. The problem has been treated by K. N. Stevens (1956) on the basis of Boyle's law and the law of conservation of mass combined into the equation

$$\frac{d}{dt} V \left(1 + \frac{P(t)}{P_0} \right) = -u(t), \quad (\text{A.2-17})$$

where P_0 is the atmospheric pressure and V the volume behind the constriction. It is assumed that the expiratory muscles merely balance the over-pressure and do not deliver any additional energy for driving out the air. The amount of air exhaled during the release will then be small, and it is possible to make use of the approximation of constant volume V . At sufficiently high flow ratings, the pressure drop $p(t)$ and the particle velocity $u(t)$ are to a first approximation related through Eq. A.2-2, and the differential equation is simply

$$dv = -\frac{P_0}{\rho V} A(t) dt. \quad (\text{A.2-18})$$

It is assumed that the minimum constriction cross-sectional area $A(t)$ varies from zero to the final value A_0 by means of the exponential function $A(t) = A_0(1 - e^{-t/t_i})$. The solution to Eq. A.2-18 may be put in the form

$$v_0 - v(t) = \frac{P_0 A_0 t}{\rho V} \left[1 - \frac{t_i}{t} (1 - e^{-t/t_i}) \right], \quad (\text{A.2-19})$$

where $v_0 = \sqrt{2p_0/\rho}$ is the initial particle velocity at $t = 0$ and p_0 the initial over-pressure. At the time $t = t_i$, the area $A(t)$ has reached a value of $1/e$ or 37 per cent below the final value A_0 , and the pressure drop at $t = t_i$ is only 37 per cent of that which would have occurred in the simplified case of an abrupt change of area from 0 to A_0 at $t = 0$. If the opening time t_i is negligibly short, the particle velocity $v(t)$ decays linearly with time and the over-pressure $p(t)$ and thus the noise pressure amplitude may be expected to decay according to a parabolic function. In the case of a finite t_i and small A_0 the rise time for the onset of the noise will be of the same order of magnitude as t_i since the increase of the effective width h of the constriction will be the determining factor for noise intensity variations, Eq. A.2-16, as long as h increases faster than the decay of v . The duration of the onset time is found to be greater for dentals than for other stops.

In the last part of the pressure equalization period the streaming is laminar, and the constant resistance will determine an exponential decay of the flow. An approximate measure of the duration of the turbulent interval may be defined from the time it takes for the particle velocity, and thus the over-pressure, to decay to zero if the final interval of laminar flow is ignored. According to Eq. A.2-19 this time will be

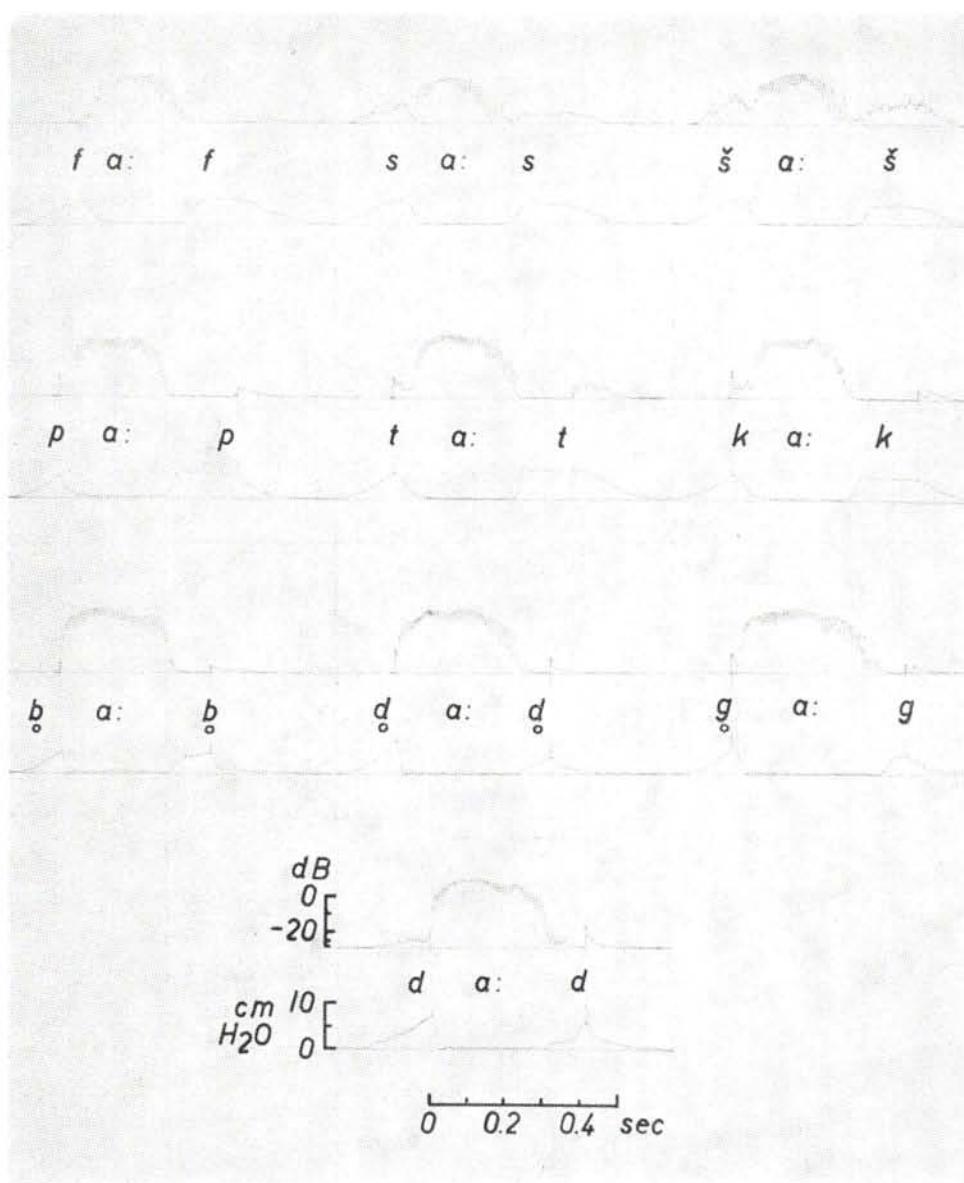


Fig. A.2.2. Simultaneous oscillographic display (*Mingograph* recorder) of (1) upper curve: sound intensity (*A*-curve weighting and compressed amplitude scale) and; (2) lower curve: pharyngeal air pressure (via a nose catheter) within mono-syllabic test words of the type Ca:C where the consonant C is [f] [s] [š] [p] [t] [k] [b] [d] [g]. These were spoken by a Swedish subject. The variants [b̥] [d̥] [g̥] are devoiced, i.e., no vocal cord vibrations occur during the pressure build-up period (occlusion).

$$t_T = \frac{\rho V}{P_0 A_0} \sqrt{2p_0/\rho}. \quad (\text{A.2-20})$$

When the glottis is open, the volume V representing the lungs is of the order of 4000 cm^3 . Assuming an over-pressure of $6 \text{ cm H}_2\text{O}$ appropriate for the production of an unvoiced strong stop and a constriction area of $A_0 = 0.1 \text{ cm}^2$ applicable to a dental, the decay time t_T will be 130 msec , assuming that the constriction area increases abruptly. According to Eq. A.2-2, the ideal linear decay starts from a velocity of 3200 cm/sec . Calculations show that the high constriction resistance causes a much higher damping than that for critical damping of $F1$. The constant R/L of the constriction constitutes a rise time of the order of 1 msec , which is much less than the time it takes for the tongue to move back from the closure under actual speaking conditions. After 50 msec the particle velocity v has reached the value 1300 cm/s which, according to Meyer-Eppler (1953), is a probable lower limit for dental noise generation. At this instant the amount of air exhaled will only be 11 cm^3 . These time measures are of the correct order of magnitude as shown by spectrograms; see also Fig. A.2-2.

The status of the glottis is of fundamental importance for the duration of the discharge interval as mentioned by Stevens (1956). Assuming a closed glottis $V = 70 \text{ cm}^3$ and $p_t = 3 \text{ cm H}_2\text{O}$ which might be valid for a weak voiced stop, Eq. A.2-19 is no longer applicable since the initial velocity, approximately 800 cm/sec will be barely sufficient to create turbulence. The vocal tract will be less than critically damped because of the small volume, and the decay thus oscillatory and largely determined by the first formant and of the order of 5 msec .

It is of some interest to note that the opposition of tense/lax for stops, sometimes referred to as fortis/lenis and defined by the noise duration, can, according to Eq. A.2-20, be maintained by either an open/closed glottis, which is the most effective means, or a slower/faster rate of area increase at the articulatory constriction and by a greater/smaller over-pressure behind the constriction. Any of these factors can cause a prolongation of the decay time. A superimposed expiratory breath-pulse will also cause a prolongation by the maintenance or at least support of the over-pressure. In case the constriction opening is kept narrow, there results an affrication, and if the constriction rapidly opens past the critical width, the breath-pulse will result in a very marked aspirative sound interval.

The turbulent noise is a secondary effect of the airflow. The impact of the air itself at the release constitutes a source which in electrical engineering terminology is called a transient source. This is represented in the electrical analog circuit by the closing of a switch at the constriction. The source and the back cavity system are thus transformed into a two-terminal branch containing the switch in series with the source impedance and the back cavity impedance and a voltage corresponding to the initial over-pressure. This is evidently a step function in case of an abrupt area shift, characterized by a spectrum envelope of -6 dB/octave . Owing to the finite opening time the fall will be greater at higher frequencies.

In acoustic terminology the initial over-pressure may be referred to as a uniformly distributed density source within the back cavities (Ingård, 1956). The electrical analog circuit source defined with the aid of Thevinin's theorem gives a conceptually clearer starting point for the calculation of the frequency characteristics.

The validity of the assumption of a step function source has been tested earlier in calculation and has been further verified in this work. The spectra of turbulent sound sources are not so well known but the results of the calculations in this work support the conclusions from model experiments performed by the M.I.T. group, see e.g., Heinz (1957a), that the essential part of the noise source spectrum is flat and that there is a fall at low and high frequencies.

The source impedance will influence the spectrum of the produced sound, especially at low frequencies. The source impedance of stop sounds and continuant fricative sounds is composed of the constriction inductance plus a resistive term which in first approximation, *Eq. A.2-10*, is twice the flow resistance assuming turbulent conditions; see also *Eq. A.2-12, 13, 14*.

Synchronous recordings of sound intensity and over-pressure during the production of stops and fricatives⁶ are exemplified in *Fig. A.2-2*. The burst interval is found to be shorter for initially positioned [g] than for [k] and is almost absent in initial [b] and [d], while apparent for [p] and [t]. At the instant of explosion the over-pressure is larger for the devoiced [d] than for [t] in this example, but the discharge time for [d] is of the same order of magnitude or smaller than the 15 msec integration time of the air pressure probe. The discharge time for initial [p], [t], and [k] is of the order of 70 msec and this is also the duration of the burst interval from explosion to the onset of vocal cord vibrations. The pressure decay of terminal [p], [t], and [k] starts from a constant value of 5-7 cm H₂O and has a duration of the order of 100-200 msec.

The duration of the occlusion interval of the terminal stop is twice as long for [p], [t], [k], as for [b], [d], [g], but the duration of the interval from onset of the vowel to the explosion of the terminal stop is very nearly the same.

The sound intensity curves of the fricatives reflect the shape of the corresponding pressure curves and have a threshold corresponding to a pressure of the order of 1 cm H₂O. The intensity of the [f] is typically low as compared with that of [s] or [š].

⁶ Synchronous oscillographic recordings of sound and over-pressure have earlier been performed by Fischer-Jørgensen (personal communication). A nose catheter was used by Malécot (1955) for kymographic recordings of air pressures. A mouth probe was used by Stetson (1951).

A.3 ANALYTICAL STUDY OF SIMPLE RESONATOR MODELS WITH APPLICATIONS TO SPEECH PRODUCTION

A.31 *The Single Helmholtz Resonator*

A. ONE OPENING

The most commonly used theoretical model for acoustic interpretation of resonance phenomena in speech production is the Helmholtz resonator. There exists extensive literature on the physical details of this class of resonators. The damping of simple resonators and their end corrections have been treated in detail by Nielsen (1949) and by Ingård (1953), the latter with special emphasis on non-linear effects. Single and compound resonators of dimensions large enough for wave propagation to be taken into account, have not, however, been studied to the same extent. The following presentation is intended to provide a more general inventory of simple resonators both of the Helmholtz type (lumped impedance element systems) and the transmission line type (distributed parameter systems).

The mathematical treatment will be based on the equivalent network concept. One of the main objectives of the analysis is to determine the resonance and anti-resonance frequencies and the associated damping constants constituting the poles and the zeros of the system. The first step is to determine a closed-form expression of the transfer function relating output current through the branch containing the radiation resistance to the current or voltage of the active source. The frequency variable $j\omega$ is substituted for the complex frequency variable $s = \sigma + j\omega$. The real and imaginary parts of the characteristic equation are then separately equated to zero.

The current transfer function of the single resonator, *Fig. A.3-1*, has the form

$$H(s) = \frac{U_0(s)}{U_1(s)} = \frac{1}{1 + RG + s(LG + RC) + s^2LC} = \frac{\omega_{01}^2}{(s - \hat{s}_1)(s - \hat{s}_1^*)}. \quad (\text{A.31-1})$$

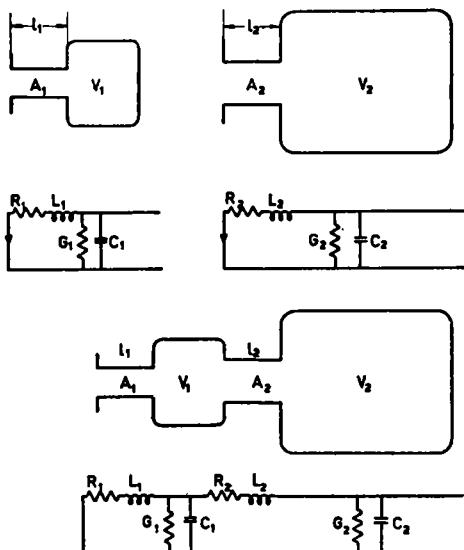


Fig. A.3-1. The classical double Helmholtz resonator and its equivalent electrical network.

1. Less than critical damping:

$$\frac{1+RG}{LC} > \frac{1}{4} \left(\frac{R}{L} + \frac{G}{C} \right)^2,$$

$$\hat{s}_1, \hat{s}_1^* = -\frac{1}{2} \left(\frac{R}{L} + \frac{G}{C} \right) \pm j \sqrt{\frac{1+RG}{LC} - \frac{1}{4} \left(\frac{R}{L} + \frac{G}{C} \right)^2}, \quad (\text{A.31-2})$$

$$\hat{s}_1 = \sigma_1 + j\omega_1,$$

$$\hat{s}_1^* = \sigma_1 - j\omega_1.$$

Vocal resonances generally have high $Q = \omega_0/2\sigma$ which implies $RG \ll 1$ and $\omega_0^2 = 1/LC \gg \frac{1}{4}(R/L+G/C)^2$. Under these conditions the frequency and bandwidth of the resonance are

$$F_1 = \frac{1}{2\pi} \omega_1 = \frac{1}{2\pi\sqrt{LC}}, \text{ and} \quad (\text{A.31-3})$$

$$B_1 = \frac{-1}{\pi} \sigma_1 = \frac{R}{2\pi L} + \frac{G}{2\pi C}.$$

In terms of the acoustic quantities $L = \rho l_e/A$ and $C = V/\rho c^2$,

$$F_1 = \frac{c}{2\pi} \sqrt{\frac{A}{l_e V}}, \quad (\text{A.31-4})$$

as stated before. *The resonance frequency is inversely proportional to the square root of the volume V , the effective length l_e of the neck, and directly proportional to the square root of the neck area A .*

The dependency of resonance bandwidth B on the circuit constants according to Eq. A.31-3 will next be exemplified by an investigation of the differential effect of a narrowing of the resonator neck.

If the resistance element R is proportional to $f^{\frac{1}{2}} A^{-\frac{1}{2}}$ as for classical viscosity losses, Eq. 1.21-13, it is apparent that the bandwidth of the resonance increases in proportion to $f^{-\frac{1}{2}}$ or as $A^{-\frac{1}{2}}$, i.e., it is inversely proportional to the square root of the constriction radius at constant resonator volume and neck length. If the resistance R is considered to be independent of frequency, the bandwidth increases as $1/f$, i.e., the Q , decreases as $1/f^2$. The damping effect of radiation is of importance at frequencies above 1000 c/s only, contributing in proportion to A and to f^2 at constant volume. There exists apparently a maximum Q at an intermediate frequency of the simple resonator, as calculated by Ingård (1953). Similarly there is a constriction radius for minimum bandwidth of the ideal resonator, as can be calculated from the data above. A more detailed discussion of the damping problem is given in Section A.36-A.

Formants of very low frequency, such as F_1 of voiced consonants, could theoretically be more than critically damped. Critical damping implies $\hat{s}_1 = \hat{s}_1^*$ and $\omega_1 = 0$. Another technical reference of some interest is a matched prototype low-pass half-section filter in which case $\sigma_1 = \omega_1$.

2. More than critical damping:

$$\frac{1+RG}{LC} < \frac{1}{4} \left(\frac{R}{L} + \frac{G}{C} \right)^2,$$

$$\hat{s}_1, \hat{s}_1^* = -\frac{1}{2} \left(\frac{R}{L} + \frac{G}{C} \right) \pm \sqrt{\frac{1}{4} \left(\frac{R}{L} + \frac{G}{C} \right)^2 - \frac{1+RG}{LC}}. \quad (\text{A.31-5})$$

The roots are real and different, one increasing and the other decreasing as the losses increase. If the main cause of the damping is the element R , the two poles will occur at $-R/L$ and $-1/RC$ respectively, provided the damping is large.

This overcritical damping probably applies to the first formant of a voiced constrictive. It is obviously difficult to measure the pole frequencies under these conditions.

B. TWO OPENINGS

The addition of a branch R_2+sL_2 paralleling the capacitance causes an increase in resonance frequency and introduces a pole a_1 and a zero a_2 on the negative real axis. Assuming small losses:

$$H(s) = \frac{L_2(s-a_2)}{(L_1+L_2)(s-a_1)} \cdot \frac{(\omega_1^2+\sigma_1^2)}{(s-\hat{s}_1)(s-\hat{s}_1^*)}, \quad (\text{A.31-6})$$

where

$$\begin{aligned} a_2 &\simeq -R_2/L_2, \\ a_1 &\simeq -(R_1+R_2)/(L_1+L_2), \\ \omega_1 &\simeq \frac{L_1+L_2}{L_1 L_2} \cdot \frac{1}{C}, \\ \sigma_1 &= \frac{1}{2} \left(\frac{G}{C} + \frac{R_2}{L_2} \cdot \frac{L_1}{L_1+L_2} + \frac{R_1}{L_1} \cdot \frac{L_2}{L_1+L_2} \right), \text{ and} \\ \hat{s}_1, \hat{s}_1^* &= \sigma_1 \pm j\omega_1. \end{aligned}$$

In terms of acoustic quantities the resonance frequency is apparently

$$f_1 = \frac{c}{2\pi} \sqrt{\frac{1}{V} \left(\frac{A_1}{l_{1e}} + \frac{A_2}{l_{2e}} \right)}, \quad (\text{A.31-7})$$

where l_{2e} is the length and A_2 the area of the second opening. The resonance frequency increases with an increase in the conductivity index A/l of either opening. If losses are uniformly distributed, i.e., $R_1/L_1 = R_2/L_2$, the transform is the same as for a simple resonator. Compare the comments on nasal coupling in the previous chapter.

If the source is a constant voltage inserted in series with $R_2 L_2$, the only change in the system function is that the zero factor $s-a_2$ is removed and replaced by the constant $1/L_2$:

$$\begin{aligned} H(s) &= \frac{1}{R_1+R_2} \cdot \frac{a_1}{s-a_1} \cdot \frac{(\omega_1^2+\sigma_1^2)}{(s-\hat{s}_1)(s-\hat{s}_1^*)} = \\ &= \frac{1}{L_1+L_2} \cdot \frac{1}{s-a_1} \cdot \frac{(\omega_1^2+\sigma_1^2)}{(s+\sigma_1)^2+\omega_1^2}. \end{aligned} \quad (\text{A.31-8})$$

A.32 The Double Helmholtz Resonator

The loss-free double resonator was first treated by Raleigh (1877). It has played a very important role in the phonetic literature, e.g., for the interpretation of the cavity-formant relations of the first two formants of vowels, e.g., Paget (1930); Crandall (1927); Chiba and Kajiyama (1941); Essner (1947); and Joos (1948).

The obvious limitation of double resonator theory to the study of speech production is the fact that the dimensions of the vocal tract—when compared with wavelengths $\lambda = c/f$ of interest, are generally not small. The relation between frequency and permissible cavity length is $f < c/8l$. The applicability cannot be judged from a low frequency criterion alone. There should also be a distinct resonator neck separating the two main cavities and one at the front end. The first two formants of back vowels and the first formant of not-too-open front vowels can approximately be calculated by double or single Helmholtz resonator models. On the other hand, there may be a complete failure when attempting to apply the Helmholtz resonator theory to F_2 of front vowels as shown in more detail in *Section 2.33*.

The bottom circuit of *Fig. A.3-1* with the addition of a branch R_3L_3 , paralleling C_2 , is applicable to the complete double resonator with openings in both cavities. The current transfer ratio is

$$H_s = \frac{L_3}{(L_1+L_2+L_3)} \cdot \frac{(s-a_2)}{(s-a_1)} \cdot \frac{(\omega_1^2+\sigma_1^2)(\omega_2^2+\sigma_2^2)}{[(s+\sigma_1)^2+\omega_1^2][(s+\sigma_2)^2+\omega_2^2]}, \quad (\text{A.32-1})$$

where

$$a_2 \simeq -R_3/L_3,$$

$$a_1 \simeq -(R_1+R_2+R_3)/(L_1+L_2+L_3),$$

$$\omega_{1,2} = \left[\frac{\omega_a^4}{2} \mp \left(\frac{\omega_a^4}{4} - \omega_b^4 \right)^{\frac{1}{2}} \right]^{\frac{1}{2}},$$

$$\omega_a^2 = \omega_1^2 + \omega_2^2 = \frac{1}{L_1C_1} + \frac{1}{L_2C_2} + \frac{1}{L_3C_2} + \frac{1}{L_2C_1},$$

$$\omega_b^4 = \frac{L_1+L_2+L_3}{L_1L_2L_3C_1C_2} = \omega_1^2\omega_2^2, \text{ and}$$

$$\begin{aligned} \sigma_1 \simeq & \frac{1}{2(\omega_2^2-\omega_1^2)L_1C_1} \left\{ R_1 \left(\frac{L_1}{L_2} + \frac{L_1C_1}{L_2C_2} + \frac{L_1C_1}{L_3C_2} - \omega_1^2 L_1 C_1 - \frac{L_1/L_3}{\omega_1^2 L_2 C_2} \right) + \right. \\ & + \frac{R_2}{L_2} \left(1 + \frac{L_1C_1}{L_3C_2} - \omega_1^2 L_1 C_1 - \frac{L_2/L_3}{\omega_1^2 L_2 C_2} \right) + \frac{R_3}{L_3} \left(1 + \frac{L_1}{L_2} + \frac{L_1C_1}{L_2C_2} - \right. \end{aligned}$$

$$\begin{aligned} & - \omega_1^2 L_1 C_1 - \frac{1}{\omega_1^2 L_2 C_2} \Big) + \frac{G_1}{C_1} \left(\frac{L_1C_1}{L_3C_2} + \frac{L_1C_1}{L_2C_2} - \omega_1^2 L_1 C_1 \right) + \\ & \left. + \frac{G_2}{C_2} \left(1 + \frac{L_1}{L_2} - \omega_1^2 L_1 C_1 \right) \right\}. \quad (\text{A.32-2}) \end{aligned}$$

σ_2 is found by exchanging ω_1 in the expression above for ω_2 . The effect of a_2 and a_1 is restricted to very low frequencies as was the case in example *B*.

The relations between the impedance elements of the equivalent circuit and the resonances of the double resonator will be discussed next.

The resonance frequency of the front resonator, regarded as a separate unit, will be denoted by

$$F_{01} = \frac{1}{2\pi\sqrt{L_1C_1}}. \quad (\text{A.32-3})$$

Similarly, the *uncoupled* resonance of the back cavity will be denoted by

$$F_{02} = \frac{1}{2\pi\sqrt{L_2C_2}}. \quad (\text{A.32-4})$$

The glottis branch L_3R_3 represents a high impedance termination of the vocal tract. A finite value of the glottis inductance L_3 causes an increase in F_{02} :

$$F_{02} = \frac{1}{2\pi\sqrt{L_2C_2}} \sqrt{\frac{L_2+L_3}{L_3}} \quad (\text{A.32-5})$$

In the following analysis the branch L_3R_3 will be neglected or will be taken into account as an increase in the conductivity of the branch L_2R_2 as a high impedance parallel element. Providing $L_3+L_2 \gg L_1$, the analysis can then proceed as if the branch L_3R_3 did not exist. The resonance frequencies of the compound system after interconnection are denoted by F_1 and F_2 . By definition $F_2 > F_1$ but F_{01} is quite independent of F_{02} . From Eq. A.32-2.

$$\begin{aligned} F_1^2 + F_2^2 &= \frac{1}{4\pi^2} \left(\frac{1}{L_1C_1} + \frac{1}{L_3C_2} + \frac{1}{L_2C_1} \right), \text{ and} \\ F_1^2 F_2^2 &= \left(\frac{1}{4\pi^2} \right)^2 \frac{1}{L_1C_1 L_3 C_2} = F_{01}^2 F_{02}^2. \end{aligned} \quad (\text{A.32-6})$$

Introduce the inductance ratio

$$k_L = L_1/L_2 = l_{1e}A_2/l_{2e}A_1, \quad (\text{A.32-7})$$

and the capacitance ratio

$$k_C = C_1/C_2 = V_1/V_2, \quad (\text{A.32-8})$$

which also is a volume ratio.

The resonance frequencies are then obtained from

$$F_{1,2}^2 = \frac{1}{4\pi^2 L_1 C_1} \left[\frac{1}{2} (1+k_L+k_L k_C) \mp \sqrt{\frac{1}{4} (1+k_L+k_L k_C)^2 - k_L k_C} \right]. \quad (\text{A.32-9})$$

After interconnection of the two resonators, the higher resonance frequency increases and the lower decreases. Their product is, however, the same. Fig. A.3-2 shows the frequency shift factor $F_{01}/F_1 = F_2/F_{02}$ corresponding to the conditions $F_{02} > F_{01}$ (solid line) and $F_{02}/F_1 = F_2/F_{01}$ corresponding to the condition $F_{02} < F_{01}$ (broken line). Each of the uncoupled frequencies F_{01} and F_{02} has to be determined numerically. The resonance frequencies can then be determined from the frequency shift factor. The lower diagram of Fig. A.3-2 shows the ratio F_2/F_1 . Under the conditions that either F_2 and F_1 or the volumes differ appreciably, the frequency shift will be small.

The articulatory conditions for small F_2/F_1 , as for the back vowels [o] [ɔ] [ɑ], are apparent from these data. If F_2/F_1 is to be reasonably small, both cavities must have about the same uncoupled resonance frequencies, i.e., ideally $F_{01} = F_{02}$, which also implies a fairly equal dependency of F_1 and F_2 on the front and the back resonators. If F_2/F_1 reaches a value close to 1, it is also necessary that the volume ratio V_1/V_2 shall be as large as possible, i.e., the back of the tongue must approach the pharynx

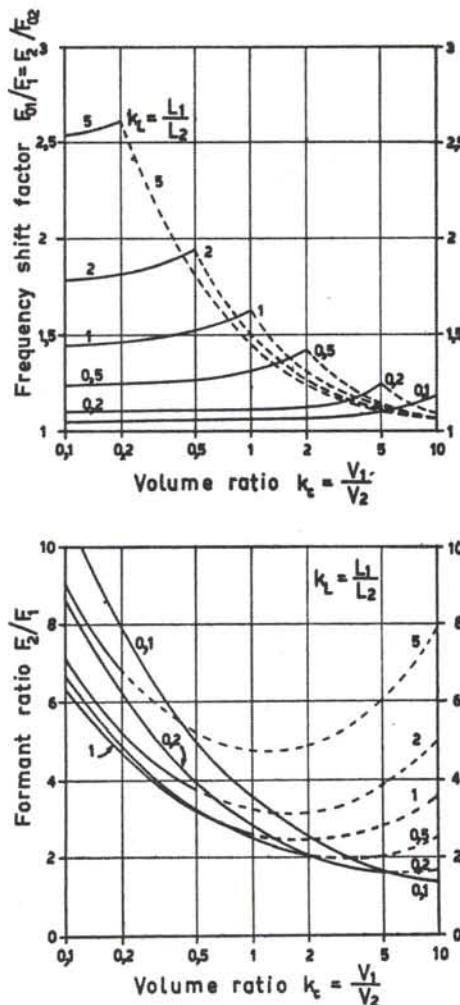


Fig. A.3-2. Nomograms for calculating the frequencies of the first formant F_1 and the second formant F_2 of a double Helmholtz resonator and their ratio. F_{01} indicates the lower and F_{02} the higher of the uncoupled resonance frequencies of the two resonators analyzed separately. Observe that either F_{01} or F_{02} can be the uncoupled front cavity resonance frequency. V_1 and V_2 are the volumes of the front and back cavities respectively, and $L_1 = \varrho l_1/A_1$ and $L_2 = \varrho l_2/A_2$ are the inductances of the front and back resonator necks respectively. Thus, $k_L = L_1/L_2 = l_1 A_2/l_2 A_1$. The frequencies F_{01} and F_{02} have to be determined numerically. F_1 and F_2 are then obtained from the upper graph which shows how much the resonances of the two cavities separate owing to the coupling.

wall to reduce the back cavity volume. A ratio of $V_1/V_2 = 5$ is needed if a formant ratio $F_2/F_1 = 1.6$ is required as for instance in a vowel [a] of $F_1 = 630$ c/s and $F_2 = 1000$ c/s. It is not possible to compensate by making the front cavity small

and the back cavity large. Theoretically there remain two possible means of compensation. One is from a finite glottis inductance, L_3 , which can play some role providing the open interval within a fundamental period is relatively long and the airflow slow enough to minimize the non-linear increase of the glottis resistance. An increasing glottis conductivity will raise F_1 more than F_2 . The other is that a minor coupling to the nasal cavities can, according to our calculations, have the same effect.

It is of some general interest to perform an analytical investigation of how, and to what extent, each of the two resonances of a double resonator is associated with each of the two resonator units. The relations are apparently not as simple as the F_1 -back-cavity, F_2 -front-cavity affiliations generally proposed in phonetic theory. The following mathematical treatment will be limited to the Helmholtz resonator model. First the terminology should be made clear. The word "cavity" above is generally used in the sense of resonator, i.e., cavity plus associated neck. The back cavity enters in the formula via its volume V_2 or by its electrical equivalent, the capacitance C_2 . The main constriction between the two resonators is the neck of the back resonator, entering via the inductance L_2 of the analog electrical circuit. Each of F_1 and F_2 can be related to any of the four electrical circuit elements $L_1L_2C_1C_2$ or to the corresponding acoustical parameters I_1/A_1 I_2/A_2 V_1 V_2 . From a differentiation of Eq. A.32-6 it is possible to compute the increase in each of F_1 and F_2 caused by a decrease in one of the four circuit elements of the double resonator. These calculations can be summarized as follows:

Incremental variation in		Relative frequency shift	
Acoustical system	Electrical system	F_1	F_2
V_1	C_1	$\frac{1}{2}A$	$\frac{1}{2}B$
V_2	C_2	$\frac{1}{2}B$	$\frac{1}{2}D$
I_1/A_1	L_1	$\frac{1}{2}C$	$\frac{1}{2}D$
I_2/A_2	L_2	$\frac{1}{2}D$	$\frac{1}{2}C$

where

$$A = -2 \frac{\delta F_1}{F_1} \frac{C_1}{\delta C_1} = -2 \frac{\delta F_2}{F_2} \frac{C_2}{\delta C_2} = \frac{F_{02}^2 - F_1^2}{F_2^2 - F_1^2},$$

$$B = -2 \frac{\delta F_2}{F_2} \frac{C_1}{\delta C_1} = -2 \frac{\delta F_1}{F_1} \frac{C_2}{\delta C_2} = \frac{F_2^2 - F_{02}^2}{F_2^2 - F_1^2},$$
(A.32-10)

$$C = -2 \frac{\delta F_1}{F_1} \frac{L_1}{\delta L_2} = -2 \frac{\delta F_2}{F_2} \frac{L_2}{\delta L_2} = \frac{F_2^2 - F_{01}^2}{F_2^2 - F_1^2}, \text{ and}$$

$$D = -2 \frac{\delta F_2}{F_2} \frac{L_1}{\delta L_1} = -2 \frac{\delta F_1}{F_1} \frac{L_2}{\delta L_2} = \frac{F_{01}^2 - F_1^2}{F_2^2 - F_1^2}.$$

The constants A , B , C , and D vary between 0 and 1. These limits correspond to zero and maximal formant-resonator dependency respectively. Thus the condition $A = C = 0$, $B = D = 1$ applies to a single resonator. A one per cent variation of its inductance or capacitance causes a one-half per cent change of resonance frequency.

In general

$$A+B = C+D = 1, \quad (\text{A.32-11})$$

which means that the sum of the percentage increase in the two resonance frequencies is always one-half of the percentage decrease in the resonator element that was varied to cause this detuning. The resonator-resonance dependency is symmetrical. The effect, on a percentage basis, of a variation of C_1 on F_1 , is the same as the effect on F_2 of a variation of C_2 . Also, the effect of C_1 on F_2 is the same as of C_2 on F_1 . Analogous relations hold for the effects of a variation in L_1 and L_2 .

For the special case of the front cavity being tuned to the same uncoupled resonance frequency as the back cavity, i.e., $F_{01} = F_{02}$ or in other words $1/L_1C_1 = 1/L_2C_2$, the constants are

$$\begin{aligned} A &= D = 1/(1+F_2/F_1), \\ B &= C = 1/(1+F_1/F_2), \end{aligned} \quad (\text{A.32-12})$$

thus

$$A+B = B+D = 1. \quad (\text{A.32-13})$$

Under the special conditions of $F_{01} = F_{02}$ it can be seen that the relative frequency shifts in F_1 and F_2 due to a one per cent decrease in both cavity volume and the degree of neck constriction l/A of one of the resonators, equals +1 per cent in both F_1 and F_2 . This rule can also be interpreted as a statement of equal dependency of F_1 and F_2 on the back and front resonators when these have the same uncoupled resonance frequencies. In this case the first resonance is, however, somewhat more dependent on the back cavity volume and the front orifice, and the second resonance is to the same degree more dependent on the front cavity volume and the neck of the back resonator. The dependency of both F_1 and F_2 on the front and back cavity volumes apparently becomes more equally divided when F_1 approaches F_2 as a result of a change in the resonator dimensions. A comparison between resonator-formant dependency factors measured with a transmission line analog to the vocal tract, and those calculated from Eq. A.32-10, is performed in Section 2.33.

The cavity-formant dependency can also be described by the amount of damping contributed by the resistive elements of any particular part of the resonator system to one of the resonances. The effect of damping is quantitatively related to the bandwidths of the resonances. The contribution from each resistive element within the double resonator to the bandwidths of each of the two resonances, as stated by the expression A.32-2, can be rewritten in the following form, disregarding the separate effect of the branch R_3L_3 , and adopting the notations introduced in the expression A.32-10.

$$\begin{aligned} B_1 &= \frac{G_1}{2\pi C_1} A + \frac{G_2}{2\pi C_2} B + \frac{R_1}{2\pi L_1} C + \frac{R_2}{2\pi L_2} D, \\ B_2 &= \frac{G_1}{2\pi C_1} B + \frac{G_2}{2\pi C_2} A + \frac{R_1}{2\pi L_1} D + \frac{R_2}{2\pi L_2} C. \end{aligned} \quad (\text{A.32-14})$$

The resonances are damped in proportion to the R/L - or G/C -values of a branch and in proportion to the factor A , B , C , or D denoting its relative importance for the tuning of the frequency of the resonance. The damping criteria for resonance-cavity relations are thus exactly the same as the frequency criteria that could be expected from energy relations.

If the two resonances are far apart, e.g., $F_2 \gg F_1$ and the ratio $F_2/F_{01} = F_{02}/F_1$ is close to unity it can be seen that the factors A and C become very small and the factors B and D close to unity, denoting the close dependency of F_1 on the back resonator and of F_2 on the smaller front resonator.

From the equivalent circuit of the double Helmholtz resonator it is also possible to calculate the ratio of the sound pressure in the front cavity to the pressure in the back cavity. The result is

$$\frac{P_1}{P_2} = 1 - F_1^2/F_{02}^2 \quad (\text{A.32-15})$$

at the frequency F_1 of the first resonance and

$$\frac{P_1}{P_2} = 1 - F_2^2/F_{01}^2 \quad (\text{A.32-16})$$

at the frequency of the higher resonance, F_2 .

Since F_2 is always greater than F_{02} and F_1 is always smaller than F_{02} , it can be concluded that the pressures in the two cavities are of the same sign at the frequency F_1 and of opposite sign at the frequency F_2 . If the back cavity uncoupled resonance F_{02} is close to F_1 it may be seen that at the frequency F_1 the back cavity has the largest pressure and that the front cavity pressure P_1 dominates at F_2 . Results from detailed calculations on actual vocal tract models on the basis of a distributed parameter analog will appear in *Section 2.34*.

A.33 The Single Tube as an Acoustic Resonator

Single or compound Helmholtz resonators have a restricted applicability, not only because of the relatively large dimensions involved but also because the true configuration of the vocal tract often departs from the requirements of the theoretical model. Thus the area of the lip-opening may be the same or larger than the front part of the mouth. The lacking front orifice is, however, not the most serious objection. The main difficulty is that the neck of the back resonator may either fail altogether or be so closely associated with the front cavity that it will be difficult or meaningless to separate the two, as is the case in many front vowels. There may also be more

than one internal constriction of importance as in vowels and consonants produced with combined retroflex and back tongue articulation.

An approximation of the vocal tract configuration by means of a large number of cylindrical sections is needed if a maximum amount of articulatory information is to be preserved. The matrix calculation technique described in the previous chapter must then be used.

A representation of the area function by a few cylindrical sections is useful for explaining the essentials of speech production, since formulas in closed form can be derived. This approach is physically better founded than the double resonator theory since it allows for more than two formants. When discussing the cavity dependence of a particular formant, either approach may be preferable. The single Helmholtz resonator can be thought of as a special case of a system of two cylindrical tube sections. The general small loss solution to this twin-tube resonator will be derived, since it is an important model for the study of formant damping. The single tube with arbitrary impedance termination will first be treated.

A single tube circularly or arbitrarily shaped, but constant cross-sectional area, behaves like an electrical transmission line up to frequencies of $f = 20000/d$, where d is the largest cross-dimension in cm. Above this frequency, transverse vibrations can occur with an effect similar to that of a shunting cavity system. They can generally be disregarded in speech analysis.

The current transfer ratio relating the current through the short-circuited end of the line, i.e., the open end of the tube, to the constant current, supplied at the opposite open-circuited end, i.e., the approximately closed end of the tube, is

$$U_0/U_q = H(s) = \frac{1}{\cosh \Gamma(s)} = \frac{1}{\cosh [al + sl/c]} = \frac{K}{\prod_{n=1}^{\infty} (1 - s/s_n)(1 - s/s_n^*)}, \quad (\text{A.33-1})$$

where, as before, l is the length of the line, c is the velocity of sound, and α is the attenuation constant

$$\alpha = \alpha_R + \alpha_G = \frac{1}{2c} \left(\frac{R}{L} + \frac{G}{C} \right); \quad (\text{A.33-2})$$

and R , L , G , and C are the *resistance, inductance, conductance, and capacitance per unit length of the line*.

The poles are found at the complex frequencies $s = \sigma + j\omega$ where $H(s)$ goes to infinity. This implies

$$\cosh \Gamma(s) = 0;$$

or

$$\Gamma(s) = \pm j \frac{\pi}{2} (2n-1); \quad (\text{A.33-3})$$

$$\omega_n = \pm \frac{\pi}{2} (2n-1)c/l;$$

$$\sigma_n = -(\alpha_R + \alpha_G)c. \quad (\text{A.33-4})$$

In other words, the resonance frequencies and bandwidth are

$$F_n = (2n-1)c/4l; \quad (A.33-5)$$

$$B_n = c(a_R + a_G)/\pi.$$

Assuming small losses,

$$K = 1/\cosh al \approx 1. \quad (A.33-6)$$

The ideal neutral vowel [ə] has been defined as the sound produced from an idealized vocal tract of constant cross-sectional area and an effective length of approximately 17.6 cm or more precisely the length $c/2000$ cm, where c is the velocity of sound. The formant frequencies occur at

$$F_n = (2n-1) 500 \text{ c/s}, \quad (A.33-7)$$

i.e.,

$$F_1 = 500, F_2 = 1500, F_3 = 2500 \text{ c/s, etc.}$$

The infinite factorized pole expansion, Eq. A.33-1, can be approximated by the poles corresponding to the $g = 2, 3$, or 4 lower formants plus a rest factor $k_{rg}(s)$, as described in Section 1.23, Eq. 1.23-1.

$$U_0/U_g = \frac{k_{rg}(s)}{\prod_{n=1}^g (1-s/\hat{s}_n) (1-s/\hat{s}_n^*)} \quad (A.33-8)$$

Since the most important property of the higher poles with regard to their low frequency residues is their average density along the $j\omega$ -axis, it is apparent, that the k_{rg} factor of the single tube can be used irrespective of the particular resonator configuration and that the total length is the most important parameter. This approximation is practically valid up to the frequency of the pole $g-1$.

The resonance conditions for a single tube, closed or approximately closed at both ends are

$$Z \coth \Gamma(s) = \infty. \quad (A.33-9)$$

Similarly, the resonance conditions for a tube open at both ends or terminated into sections of a relatively large cross-sectional area are

$$Z \tgh \Gamma(s) = 0, \quad (A.33-10)$$

which is identical with Eq. A.33-9. The solution for the loss free case is

$$\omega_n = (n-1)\pi c/l_e, \quad (A.33-11)$$

i.e.,

$$F_n = (n-1)c/2l_e, \quad (A.33-12)$$

where

$$l_e = l + l_t.$$

The half-wavelength resonances in these tubes are dependent on the effective length l_e only, i.e., physical length l plus end corrections l_t . The same was true of the quarter-

wavelength resonator, i.e., the tube open at one end and approximately closed at the other end, see *Eq. A.33-9*. In general, there are two end corrections to be determined, one for each end of the tube. In the case of a spherical baffle of radius 9 cm terminating the open end, the end correction length is $l_t = 0.8(A/\pi)^{1/2}$, according to *Eq. 1.21-17*. If there is no baffle $l_t = 0.6(A/\pi)^{1/2}$, according to Morse (1948). If the open end is abruptly terminated into a larger cavity, *Eq. 1.21-18* should be used.

A partial closure at the end of a tube constitutes a high impedance termination, paralleling the equivalent transmission line. This termination may be expressed in terms of a short open tube of length l_g and cross-sectional area $A_g \ll A$. The end correction length is negative:

$$l_t = -\frac{-c^2 A_g}{\omega^2 A l_g}. \quad (\text{A.33-13})$$

This end correction vanishes at high frequencies, and its use is limited to frequencies where $\omega L_g \gg Z$. If the terminating tube contains a resistance R_g in series with L_g such that either $\omega L_g > Z$ or $R_g > Z$,

$$l_t = -\frac{c Z L_g}{R_g^2 + \omega^2 L_g^2}, \quad (\text{A.33-14})$$

where the inductance $L_g = \rho l_g / A_g$. The requirement that l_g be short can be formulated as $\omega l_g / c < 1$ and small enough that $\tanh \omega l_g / c \approx \omega l_g / c$.

If $\omega L_g \ll Z$, the end correction is positive, as in the case of the radiation reactance:

$$l_t = l_g A / A_g. \quad (\text{A.33-15})$$

When the same terminating tube is closed at the end opposite to the main tube it will behave as a high impedance capacitance termination

$$C_g = l_g A_g / \rho c^2, \quad (\text{A.33-16})$$

providing $\omega l_g / c < 1$. The end correction is positive and equal to

$$l_t = l_g A_g / A. \quad (\text{A.33-17})$$

Finally if the terminating cavity acts as a low impedance capacitance termination, e.g., a closed cavity of length l_g and area $A_g \gg A$.

$$l_t = \frac{-c}{Z \omega^2 C_g} = \frac{-c^2 A}{\omega^2 A_n l}, \quad (\text{A.33-18})$$

which is a negative end correction.

Any end correction is associated with resistive elements, e.g., a radiation resistance or a viscous resistance, *Eq. 1.21-14*. These can also be transformed into the main tube as an increase in its attenuation constant by the amount

$$\alpha_t = \frac{R_T}{Z l_t}, \quad (\text{A.33-19})$$

providing R_T is a small series resistance and

$$a_t = \frac{ZG_T}{l_e}, \quad (\text{A.33-20})$$

provided $1/G_T$ is a large resistance paralleling the line. Thus in case of a high impedance termination the resistance must first be expressed as a parallel element.

The transformations above do not change the phase angle of the characteristic impedance, i.e., a_t is divided into equal halves on the shunt losses a_G and the series losses a_R . The transformation of the terminating reactive load into the transfer constant of the tube is, however, associated with a small redistribution of losses, which will next be evaluated. The transmission properties of the tube will be specified by the following low-loss expressions; see Eq. 1.21-3.

$$\begin{cases} Z = Z_0[1+j\frac{c}{\omega}(a_G-a_R)]; \\ \Gamma = l[a_R+a_G+j\omega/c]; \\ Z_0 = \rho c/A. \end{cases} \quad (\text{A.33-21})$$

The input impedance Z_t to the tube at the end opposite to the acoustic termination Z_T is put in the well-known form

$$\begin{cases} Z_t = Z \operatorname{tgh}(\Gamma + \Gamma_t) = Z \operatorname{tgh} \Gamma_t; \\ \Gamma_t = \operatorname{artgh} Z_T/Z. \end{cases} \quad (\text{A.33-22})$$

When $Z_T/Z > 1$,

the alternative expression

$$\begin{cases} Z_t = Z \coth(\Gamma + \Gamma_t); \\ \Gamma_t = \operatorname{artgh} Z/Z_T; \end{cases} \quad (\text{A.33-23})$$

is preferable.

The general expression for the additional transfer constant is

$$\Gamma_t = l_t (\mp a_R \pm a_G + j\omega/c) + (l+l_t)a_t. \quad (\text{A.33-24})$$

Termination	Impedance specification of termination	End correction length l_t	Change in initial attenuation constants	Additional attenuation constant a_t
A Small inductance	$Z_T = R_T + j\omega\varrho l_g/A_g$	$l_g A_g / A$	$l_t (-a_R + a_G)$	$R_T/Z(l + l_t)$
B Large inductance	$\frac{1}{Z_T} = G_T + \frac{A_g}{j\omega\varrho l_g}$	$-\frac{c^2 A_g}{\omega^2 A l_g}$	$l_t (a_R - a_G)$	$G_T Z(l + l_t)$
C Small capacitance	$\frac{1}{Z_T} = G_T + j\omega l_g A_g / \varrho c^2$	$l_g A_g / A$	$l_t (-a_R + a_G)$	$G_T Z(l + l_t)$
D Large capacitance	$Z_T = R_T + \varrho c^2 / j\omega l_g A_g$	$-\frac{c^2 A}{\omega^2 A_g l_g}$	$l_t (a_R - a_G)$	$R_T/Z(l + l_t)$

According as the termination at the end of the tube opposite to Z_T is open-circuited or short-circuited, the conditions for resonance will be one of the two following alternatives:

$$\Gamma_e(s) = j\frac{\pi}{2}(2n-1), \quad (\text{A.33-25})$$

or

$$\Gamma_e(s) = j\pi(n-1). \quad (\text{A.33-26})$$

The solution for the damping constants is the same for the quarter-wavelength resonance, Eq. A.33-25, and the half-wavelength resonance conditions, Eq. A.33-26, described above.

Terminating condition	Damping constant σ
A	$-c[a_R(1-2l_t/l_e) + a_G + a_t]$
B	$-c[a_R(1-2l_t/l_e) + a_G + a_t(1-2l_t/l_e)]$
C	$-c[a_R + a_G(1-2l_t/l_e) + a_t]$
D	$-c[a_R + a_G(1-2l_t/l_e) + a_t(1-2l_t/l_e)]$

These expressions can also be derived as approximations of the general twin-tube formula, which will be given in Eq. A.35-3. It can be seen that if l_t is small, the damping constants are approximately

$$\sigma = -c(a_R + a_G + a_t). \quad (\text{A.33-28})$$

The negative end correction l_t for a tube terminating into a large volume cavity implies that the resonance frequencies are lower if l_t is disregarded. However, a terminating cavity behaves only approximately as a capacitance. There is a series inductance determined by one-third of the mass of air in the terminating cavity to be taken into account plus an additional series inductance due to the radiation effects, Eq. 1.23-18. This means that the end correction l_t of a termination type D should more properly be expressed as follows:

$$l_t = l_g \frac{A}{A_g} \left[\frac{1}{3} - \frac{c^2}{\omega^2 l_g^2} + \frac{0.48 A_g}{l_g A^{\frac{1}{2}}} \left(1 - 1.25 \left(\frac{A_g}{A} \right)^{\frac{1}{2}} \right) \right]. \quad (\text{A.33-29})$$

The contributions to the damping constant σ of very small or very large impedance elements, reactive or resistive, terminating either end of the tube are approximately independent of each other and may thus be added.

It is obvious that the transformations of the terminating impedances into the transfer constant of the main tube do not permit an unrestricted use of the equivalent network. The poles come out correctly, but the zeros may be incorrect, e.g., in the impedance at the end where Z_T is attached. For calculations of vocal tract transmission, the transformations are allowable as far as the current transfer ratios and forward transfer conductances are concerned, i.e., when the ratio or the output volume velocity to the glottis volume velocity source or to any pressure source is to be calculated.

The possibility of including the whole radiation impedance and, in case of large lip-openings, also the front orifice impedance in the tube simulating the mouth cavity is very useful when the vocal tract shall be represented by a simple compound-tube system. Similarly, the glottis resistance may be transformed into the back tube, and the same operation may also be performed on the glottis inductance at higher frequencies. In the frequency range below the third formant, the larynx tube may be transformed into the pharynx as an increase in length of $l_g A_g / A$, where A is the pharynx area and $l_g A_g$ the larynx volume.

The current transfer ratio relating the current U_T through a low impedance termination Z_T of a tube to the input current U delivered at the opposite end is

$$\frac{U_T}{U} = \frac{1}{\cosh \Gamma + \frac{Z_T}{Z} \sinh \Gamma}, \quad (\text{A.33-30})$$

or

$$\frac{U_T}{U} = \frac{\cosh \Gamma_t}{\cosh(\Gamma + \Gamma_t)}, \quad (\text{A.33-31})$$

where

$$\Gamma_t = \operatorname{artgh} Z_T/Z.$$

The factor $\cosh \Gamma_t \approx 1$, provided $Z_t \ll Z$. Analogous conditions exist for calculating the voltage at a high impedance termination as a response to a source at the opposite end.

The above transformations are not limited to extreme values of the loading impedance Z_T . One practical application is the inclusion of the radiation impedance $R_0 + j\omega L_0$ into the front tube at higher frequencies, i.e., above $f = 3000 (10/A_0)^{1/2}$ in which case R_0 is no longer small compared with Z_0 . The initial losses in the front tube may then be neglected. Under these conditions, and also if $a_g = a_R$, the exact expression for the increase in transfer constant owing to the radiation impedance is

$$\Gamma_t = \frac{j}{2} \operatorname{artg} \frac{2\omega L_0/Z}{1 + \left(\frac{R_0}{Z}\right)^2 + \left(\frac{\omega L_0}{Z}\right)^2} + \frac{1}{2} \operatorname{artgh} \frac{2R_0/Z}{1 + \left(\frac{R_0}{Z}\right)^2 + \left(\frac{\omega L_0}{Z}\right)^2} = j\omega l_t/c + ia_t. \quad (\text{A.33-32})$$

The ratio R_0/Z may be both smaller and greater than unity. The damping constants of the poles are approximately

$$\sigma_n = -ca_t. \quad (\text{A.33-33})$$

The simultaneous growth of ωL_0 and $R_0(\omega)$ prevents exact impedance matching, and the damping constants will remain small compared with the frequencies of the poles up to rather high frequencies, provided the length of the tube is fairly large. This can be seen in Fig. 1.2-3 where the higher resonances are not damped out completely.

Under the conditions $L_0 = 0$, $R_0 = Z$ the line is perfectly matched, $a_t = \infty$, and the damping constants $\sigma_n = -\infty$. The pole frequencies ω_n are not influenced by a purely resistive load.

A.34 Four-Tube Systems. Transform Equations for Arbitrary Source Locations

Compound resonator systems composed of up to four tubes of different cross-sectional areas can be handled by transcendental formulas. The transfer function relating output current at the lips to the input volume velocity current at the glottis end is

$$U_0/U_q = H_p(s) = \frac{1}{\cosh \Gamma_1 \cosh \Gamma_2 \cosh \Gamma_3 \cosh \Gamma_4 (AB+CD)}, \quad (\text{A.34-1})$$

where

$$A = \left(1 + \frac{A_2}{A_1} \tgh \Gamma_2 \tgh \Gamma_1 \right),$$

$$B = \left(1 + \frac{A_4}{A_3} \tgh \Gamma_4 \tgh \Gamma_3 \right),$$

$$C = \frac{A_3}{A_2} \left(\tgh \Gamma_2 + \frac{A_2}{A_1} \tgh \Gamma_1 \right), \text{ and}$$

$$D = \frac{A_4}{A_3} \left(\tgh \Gamma_4 + \tgh \Gamma_3 \right).$$

It has been assumed that $Z_m/Z_{m-1} = A_{m-1}/A_m$, i.e., that all the characteristic impedances have the same phase angle. Assuming small losses

$$H_p(s) = \frac{1}{\prod_{n=1}^{\infty} (1-s/\hat{s}_n)(1-s/\hat{s}_n^*)}, \quad (\text{A.34-2})$$

which means that the function is composed of conjugate poles and approaches the unity value at zero frequency. The transfer function relating output current to the voltage of a series source $E_s(s)$ of arbitrary position within the system, will be expressed by

$$\frac{U_0(s)}{E_s(s)} = H(s) = H_p(s)H_z(s), \quad (\text{A.34-3})$$

where $H_p(s)$ is the glottis transfer function according to Eq. A.34-2 and $H_z(s)$ is the zero function, including constant factors. If the source is placed in series with the radiation impedance,

$$H_z(s) = \frac{\sinh \Gamma_1 \sinh \Gamma_2 \sinh \Gamma_3 \sinh \Gamma_4 (ef+gh)}{Z_4}, \quad (\text{A.34-4})$$

where

$$e = \frac{A_1}{A_3} \left(1 + \frac{A_3}{A_4} \coth \Gamma_4 \coth \Gamma_3 \right),$$

$$f = \coth \Gamma_2 + \frac{A_2}{A_1} \coth \Gamma_1,$$

$$g = \frac{A_1}{A_2} \left(1 + \frac{A_2}{A_1} \coth \Gamma_1 \coth \Gamma_2 \right), \text{ and}$$

$$h = \frac{A_3}{A_4} \coth \Gamma_4 + \coth \Gamma_3.$$

At very low frequencies, disregarding losses:

$$\lim_{\omega \rightarrow 0} H_z(\omega) = j\omega \sum_{v=1}^4 l_v A_v / \rho c^2, \quad (\text{A.34-5})$$

as required by Eq. 1.23-7.

In general

$$H_z(s) = sC_b \prod_{n=1}^{\infty} (1 - s/\bar{s}_n) (1 - s/\bar{s}_n^*), \quad (\text{A.34-6})$$

where C_b is the capacitance of the total cavity volume behind the source.

If the source is placed at a distance of l_s cm anterior to the front end of section 2, the only change made in the zero function, Eq. A.34-4, is that Γ_1 is substituted for $\frac{l_s}{l_1} \Gamma_1$. Similar substitutions may be undertaken for any arbitrary source location.

Impedance elements in front of the source do not enter the zero function. If the source is placed in the middle of section 2 for instance, the procedure is to set $\Gamma_1 = 0$ and to substitute Γ_2 for $\Gamma_2/2$. The following expression is valid for a source at the boundary between sections 1 and 2

$$H_z(s) = \frac{\sinh \Gamma_2 \sinh \Gamma_3 \sinh \Gamma_4 \left(e \frac{A_2}{A_1} + h \cdot \coth \Gamma_2 \right)}{Z_4}. \quad (\text{A.34-7})$$

The twin-tube resonator is a useful vocal tract model. Its pole function is obtained by setting $\Gamma_3 = \Gamma_4 = 0$:

$$H_p(s) = \frac{1}{\cosh \Gamma_1 \cosh \Gamma_2 \left[1 + \frac{A_2}{A_1} \operatorname{tgh} \Gamma_1 \operatorname{tgh} \Gamma_2 \right]}. \quad (\text{A.34-8})$$

For consonant studies the source can be placed anywhere within the compound model. Assuming a source at the radiating end

$$H_z(s) = \frac{\sinh \Gamma_1 \sinh \Gamma_2 \left(\coth \Gamma_1 + \frac{A_1}{A_2} \coth \Gamma_2 \right)}{Z_2}, \quad (\text{A.34-9})$$

and if the source is moved to the boundary between sections 1 and 2,

$$H_z(s) = \frac{\sinh \Gamma_2}{Z_2}. \quad (\text{A.34-10})$$

The simple case of $A_2 = A_1$ will next be investigated more closely.¹ Without regard to radiation end correction,

¹ The single tube model of stop sound production has been treated by Ingård (1956). The equivalent network method of analysis followed here leads to the same results in a more compact form. It has independently been used by Heinz (1957b) for a twin-tube model of fricatives.

$$H_p(s) = \frac{1}{\cosh(\Gamma_1 + \Gamma_2)}; \quad (\text{A.34-11})$$

$$H(s) = H_p(s)H_z(s) = \frac{\sinh \Gamma_2}{Z_2 \cosh(\Gamma_1 + \Gamma_2)}. \quad (\text{A.34-12})$$

With radiation load included

$$H_p(s) = \frac{1}{\cosh(\Gamma_1 + \Gamma_2) + \left(\frac{R_0}{Z_1} + s \frac{\Gamma_0}{Z_1}\right) \sinh(\Gamma_1 + \Gamma_2)}, \quad (\text{A.34-13})$$

or

$$H_p(s) = \frac{\cosh \Gamma_0}{\cosh(\Gamma_1 + \Gamma_2 + \Gamma_0)}, \quad (\text{A.34-14})$$

where

$$\Gamma_0 = \operatorname{artgh} \left(\frac{R_0(s)}{Z_1} + s \frac{L_0}{Z_1} \right),$$

as previously described in the single tube analysis. In most applications of interest the low frequency approximation of Γ_0 being small is valid, i.e., $\cosh \Gamma_0 \simeq 1$.

It is assumed that at $t = 0$ a switch is closed connecting the parts of the tube in front of and behind the source. The initial over-pressure in the back tube is Δp . The source transform is thus a step function

$$E(s) = \frac{\Delta p}{s}, \quad (\text{A.34-15})$$

sloping -6 dB/octave . The energy spectrum is thus

$$W(\omega) = R_0(\omega)U_0^2(\omega) = E^2(\omega)R_0(\omega)H^2(\omega) = \left(\frac{\Delta p}{\omega}\right)^2 \left(\frac{A_2}{\rho c}\right)^2 R_0(\omega) \frac{\sinh^2 \Gamma_2 \cosh^2 \Gamma_0}{\cosh^2(\Gamma_1 + \Gamma_2 + \Gamma_0)}. \quad (\text{A.34-16})$$

The absolute value is

$$|W(\omega)| = \left(\frac{\Delta p}{\omega}\right)^2 \left(\frac{A_2}{\rho c}\right)^2 R_0(\omega) \frac{\sin^2 \varphi_2}{\left[\left(1 - \frac{\omega L_0}{Z_1} \operatorname{tg} \varphi\right)^2 + \frac{R_0^2(\omega)}{Z_1^2} \operatorname{tg}^2 \varphi\right] \cos^2 \varphi}, \quad (\text{A.34-17})$$

where $\varphi_1 = \omega l_1/c$, $\varphi_2 = \omega l_2/c$, $\varphi = \omega(l_1 + l_2)/c$.

The simplicity of Eq. A.34-12 is retained in its low frequency approximation:

$$|W(\omega)| = \left(\frac{\Delta p}{\omega}\right)^2 \left(\frac{A_2}{\rho c}\right)^2 R_0(\omega) \frac{\sin^2 \varphi_2}{\cos^2 \varphi_e + R_0^2(\omega)/Z_1^2}, \quad (\text{A.34-18})$$

where $l_e = (l_1 + l_2 + l_t)$ is the effective length of the tube, and $\varphi_e = \omega l_e/c$.

The bandwidths of the resonances are

$$B_n = -\sigma_n/\pi = \frac{cR_0(\omega)}{\pi l_e Z_1}, \quad (\text{A.34-19})$$

and their frequencies are

$$F_n = \omega_n/2\pi = (2n-1)c/4l_e. \quad (\text{A.34-20})$$

The zeros, i.e., the anti-resonances, occur at

$$F_m = \omega_m/2\pi = nc/2l_2. \quad (\text{A.34-21})$$

For practical applications, it is more useful to specify the pressure response at a certain point in front of the speaker than the total radiated energy. It follows from *Eq. 1.23-13, A.34-3, and A.34-15* that the pressure versus frequency response of an arbitrary compound resonator system to a step function pressure source of arbitrary location is

$$P(\omega, l_d) = \frac{\Delta p}{4\pi l_d} K_T(\omega) H_z(\omega) H_p(\omega), \quad (\text{A.34-22})$$

where $H_z(s)$ and $H_p(s)$ are the zero and pole functions, in the case of four-tube resonators of the form given by *Eq. A.34-1* and *A.34-4*. $K_T(\omega)$ combines the frequency correction increase of $R_0(\omega)$ in excess of ω^2 and the directivity factor.

Results from calculations of stop sound spectra, earlier performed by Fant (1950b), on the basis of a true twin-tube configuration, *Eq. A.34-8, 9, 22*, assuming small internal losses, have been included in *Chapter 2.6*; see *Fig. 2.6-17*. Damping constants were calculated from the empirical formula

$$\alpha = 0.007(\pi/A)^{1/2} \text{ neper/cm}, \quad (\text{A.34-23})$$

assuming $\alpha_G = \alpha_R = \alpha/2$. This ideal distribution of shunt and series losses is rather hypothetical. For an ideal hard-walled tube $\alpha_G \approx 0.45 \alpha_R$, as shown in *Section 1.21*. The attenuation constants of the vocal cavities are of the order of 2-8 times as large.

A.35 The Damping Effect of Series and Shunt Losses Within Twin-Tube Resonators

The following derivation of the relations between the twin-tube constants and the damping constants and pole frequencies of the resonator response is based on the small loss assumption but is general enough to provide the necessary mathematical basis for future investigations of the series and shunt losses within the vocal tract.

The pole frequencies are found from the approximation

$$\frac{A_2}{A_1} \operatorname{tg} \frac{\omega l_1}{c} = \cot \frac{\omega l_2}{c}. \quad (\text{A.35-1})$$

The damping constants are calculated by the usual technique of introducing the complex frequency $s = j\omega + \sigma$, which divides the complete characteristic equation

$$\frac{Z_1(j\omega, \sigma)}{Z_2(j\omega, \sigma)} \operatorname{tgh} \Gamma_1(j\omega, \sigma) + \coth \Gamma_2(j\omega, \sigma) = 0, \quad (\text{A.35-2})$$

into a real and an imaginary part that are separately equated to zero.

The solution can be put in the form

$$\sigma = -c[k_{R_1}a_{R_1} + k_{G_1}a_{G_1} + k_{R_2}a_{R_2} + k_{G_2}a_{G_2}], \quad (\text{A.35-3})$$

in which the sum of the damping coefficients for the series losses as well as the sum of the damping coefficients for the parallel losses equals unity, that is

$$k_{R_1} + k_{R_2} = k_{G_1} + k_{G_2} = 1. \quad (\text{A.35-4})$$

For convenience set $\omega l_1/c = \varphi_1$ and $\omega l_2/c = \varphi_2$

$$\left. \begin{array}{l} k_{R_1} \\ k_{G_1} \end{array} \right\} = \frac{1 + \operatorname{tg}^2 \varphi_1 \pm \operatorname{tg} \varphi_1 / \varphi_1}{1 + \frac{A_1 l_2}{A_2 l_1} + \left(1 + \frac{l_2 A_2}{l_1 A_1}\right) \operatorname{tg}^2 \varphi_1}, \quad (\text{A.35-5a})$$

$$\left. \begin{array}{l} k_{R_2} \\ k_{G_2} \end{array} \right\} = \frac{1 + \operatorname{cot}^2 \varphi_2 \pm \operatorname{cot} \varphi_2 / \varphi_2}{1 + \frac{A_2 l_1}{A_1 l_2} + \left(1 + \frac{l_1 A_1}{l_2 A_2}\right) \operatorname{cot}^2 \varphi_2}, \quad (\text{A.35-5b})$$

where the minus sign applies to k_{G_1} and to k_{R_2} . The relation A.35-4 implies complimentary damping effects of the losses within the two tubes. The damping coefficients are also indicative of the resonance-tube dependency in complete analogy to the reasoning in connection with Eq. A.32-14. For the ideal Helmholtz resonator $k_{R_1} = k_{G_2} = 1$ and $k_{R_2} = k_{G_1} = 0$. In the case of a standing wave resonance that is essentially confined to tube 2, the coefficients will approach the values $k_{R_1} = k_{G_1} = 0$ and $k_{R_2} = k_{G_2} = 1$.

The single tube formulas for a large or a small inductance termination, Eq. A.33-24, A and B, can be derived from Eq. A.35-5 above by setting $l_1 \ll l_2$ and $\operatorname{tg} \varphi_1 = \varphi_1$.

If the resonator neck is very narrow, $A_2/A_1 \gg 1$ and, provided $l_2 A_1/l_1 A_2 < 1$ as in a simple resonator with small but finite wave propagation in the cavity, the fundamental resonance is in first approximation

$$\omega \simeq c(A_1/A_2 l_1 l_2)^{\frac{1}{2}}, \quad (\text{A.35-6})$$

as required. The damping constant is

$$\sigma \simeq -c \left[\frac{2a_{R_1} + a_{R_2} l_2 A_1 / l_1 A_2}{2 + l_2 A_1 / l_1 A_2} + a_{G_2} \right]. \quad (\text{A.35-7})$$

The series losses of the main tube, entering through a_{R_2} are reduced to a small value. When $l_2 A_1 / l_1 A_2 \ll 1$, the tube behaves like a capacitance and

$$\sigma = -c(a_{R_1} + a_{G_2}). \quad (\text{A.35-8})$$

The second and higher resonances are of standing wave character. Because of the high neck impedance, Eq. A.35-1 reduces to

$$\frac{A_2}{A_1} \omega l_1/c = \frac{1}{\omega l_2/c - (n-1)\pi}, \text{ and} \quad (\text{A.35-9})$$

$$\omega_n = \pi(n-1)c/l_2e, \quad (\text{A.35-10})$$

where

$$l_{2e} = l_2 \left(1 - \frac{1}{\frac{A_2}{A_1} \omega^2 l_1 l_2 / c^2} \right) \quad (\text{A.35-11})$$

in agreement with the previously derived end correction length for a large inductance termination, *Eq. A.33-24*. For accurate determinations of the damping constant *Eq. A.35-5* has to be used. If $l_{2e} = l_2$ with sufficiently good approximation the following simplified formula may be used:

$$\sigma_n = -c \left[a_{R1} \frac{2 + \left(\frac{l_1}{l_2} \right)^2 (n-1)^2 \pi^2}{(n-1)^2 \pi^2 l_1 A_2 / l_2 A_1} + a_{R2} + a_{G2} \right]. \quad (\text{A.35-12})$$

Provided $(n-1)\pi l_1/l_2 < 1$ the contribution from a_{R1} to σ_n varies inversely to frequency squared. The damping effect of a series resistance in the neck is thus much smaller than if the same resistance alone or in series with a small inductance terminates Section No. 2, as in the open tube case. Applied to speech production, this means that the damping effect of the radiation resistance on the second and higher formants is much reduced by lip-rounding. The ratio of the damping constants for the small and large inductance termination at constant a_{R1} may be derived from either *Eq. A.35-5a* or *A.33-24*:

$$\sigma_{ns}/\sigma_{nl} = (n-1)^2 \pi^2 \left(\frac{l_1}{l_2} \right)^2 \left(\frac{A_2}{A_1} \right)^2. \quad (\text{A.35-13})$$

The length l_1 includes the end correction of the resonator neck. Provided the effective length l_1 is not very small compared with l_2 , the factor $(n-1)\pi l_1/l_2$ can become larger than 1. Under these conditions the damping contribution from a_{R1} is at a minimum value:

$$\sigma_n = -c \left[a_{R1} \frac{l_1 A_1}{l_2 A_2} + a_{R2} + a_{G2} \right]. \quad (\text{A.35-14})$$

If the physical length of l_1 is relatively large, the minimal loading conditions above will occur when the phase shift $\omega l_1/c = (2n-1)\pi/2$. The maximal loading effect occurs when $\omega l_1/c$ is close to $n\pi$, i.e., when the loading tube has a resonance frequency close to that of the main tube. The transformer effect is then at a maximum. The input impedance of a loss-less line terminated by a small resistance R at the far end is

$$Z_t = Z \operatorname{tgh} \left(j\omega l/c + \operatorname{artgh} \frac{R}{Z} \right), \quad (\text{A.35-15})$$

which is

$$Z_t = R \quad (\text{A.35-16})$$

under the conditions of $\omega l/c = m\pi$, and

$$Z_t = Z^2/R \quad (\text{A.35-17})$$

when

$$\omega l/c = (2n-1)\pi/2.$$

In general, the damping effect of one tube on the next of a compound system can be evaluated by determining the impedance of the loading tube seen from the main cavity. This impedance is expressed as a reactance in parallel with a resistance if the load is of high impedance, i.e., when the main cavity is partially closed by the loading tube. These elements are then included as end corrections, which vary relatively little in the frequency range of the resonance peak. Similarly, a loading tube of larger cross-sectional area than the main tube is treated as a low impedance termination of a resistance in series with a reactance. The reactance provides the length end correction and the resistance contributes to the attenuation constant, as described earlier.

A.36 Summary of Twin-Tube Formulas for the Study of Resonator Damping. Applications to Vocal Tract Models

The following section is devoted to an evaluation and summary of formulas for calculating the bandwidths of the resonances of one- or two-tube ideal resonators given their dimensions and the data on the resistive elements causing dissipation. The radiation resistance, *Eq. 1.21-16*, contains the factor $K_s(f)$, *Fig. 1.2-3*, which is the increase of $R_o(f)$ in excess of f^2 , assuming a spherical baffle. The classical losses within a hard-walled tube have been decomposed into α_R -series losses due to viscosity and α_G -parallel losses due to heat conduction, see *Eq. 1.21-13*.

The bandwidth B_n of a resonance number n is defined from the real part σ_n of the corresponding pole in the vocal tract transfer function, as specified in *Section A.12* and by *Eq. 1.23-19*

$$B_n = -\sigma_n/\pi. \quad (\text{A.36-1})$$

When several dissipative elements contribute to the damping of a resonance, their effects may be linearly superimposed, provided two resonances do not come too close and no particular resonance bandwidth is too large. These small-loss conditions are generally fulfilled for the first three formants of human speech.

The theory of hard-walled systems can be tested by measurements on mechanical models, as reported in *Section A.36 C*. The predictability is fairly good, but the results are only in part applicable to speaking conditions since there are presumably other dissipative sources in the vocal tract to be taken into account and unspecified surface and cavity wall constants to be inserted in the formulas. The theoretical basis for calculating the effects of the glottis termination, of cavity wall vibrations, and of turbulent losses due to a superimposed DC air stream, is provided here, but a systematic experimental check of the appropriate vocal tract constants remains to be undertaken.

The relative roles of various vocal tract dissipative elements as estimated theoretically are discussed in connection with the calculations on the vocal tract models of six Russian vowels, *Section 2.34*.

A. THE HELMHOLTZ RESONATOR

According to the lumped element approximation, *Eq. A.31-3*

$$B = \frac{R}{2\pi L_e} + \frac{G}{2\pi C} = B_R + B_G, \quad (\text{A.36-2})$$

where $L_e = \rho l_e/A$ is the effective inductance of the resonator neck and $C = V/\rho c^3$ is the capacitance of the resonator volume V . The resonance frequency is

$$F = \frac{1}{2\pi} \left(\frac{1}{L_e C} \right)^{\frac{1}{2}} = \frac{c}{2\pi} \left(\frac{A}{l_e V} \right)^{\frac{1}{2}}. \quad (\text{A.36-3})$$

1. Radiation Resistance

The bandwidth contribution $R_0/2\pi L_e$ from the radiation resistance can be expressed in the following alternative forms:

$$B_{R_0} = \frac{c A^2 K_s(f)}{8^2 \pi^2 V l_e^2} = 45(A/l_e)^2(10/V)K_s(f), \quad (\text{A.36-4})$$

or

$$B_{R_0} = \frac{f^2 A K_s(f)}{2 c l_e} = 14(f/1000)^2(A/l_e)K_s(f), \quad (\text{A.36-5})$$

when resonator volume V is eliminated and

$$B_{R_0} = \frac{2\pi f^4 V K_s(f)}{c^3} = 4.5(f/1000)^4(V/10)K_s(f) \quad (\text{A.36-6})$$

after elimination of the conductivity index A/l_e of the neck, where l_e is the effective length and A the cross-sectional area of the neck.

The damping effect of radiation on simple resonators is negligible at low frequencies and decreases with the fourth power of frequency when the opening of a constant volume resonator is constricted, as seen by *Eq. A.36-6*. According to *Eq. A.36-4*, the damping increases with the square of the A/l_e ratio and in inverse proportion to the volume V when each of these parameters is varied separately. At high values of the neck area A , the resonator no longer behaves like a lumped element system.

2. Frictional Losses in the Resonator Neck Under Ideal Conditions

The contributions from frictional losses to the bandwidth of the Helmholtz resonator is, according to *Eq. A.36-2* and *1.21-13*

$$B_{Rf} = S_A(\mu f/\rho A)^{\frac{1}{2}} = 13(f/1000A)^{\frac{1}{2}} S_A. \quad (\text{A.36-7})$$

After elimination of f ,

$$B_{Rf} = (\mu c/2\pi\rho)^{\frac{1}{2}}(1/l_e A V)^{\frac{1}{2}} S_A = 9.7(100/A l_e V)^{\frac{1}{2}} S_A. \quad (\text{A.36-8})$$

When A is eliminated,

$$B_{Rf} = \frac{c}{2\pi} (\mu/\varrho l_e V f)^{\frac{1}{4}} = 7.3(1000/f)^{\frac{1}{4}} (100/Vl_e)^{\frac{1}{4}} S_A. \quad (\text{A.36-9})$$

It has been assumed that the neck length l is so large that the difference in the resistance and inductance end corrections can be neglected. The shape factor

$$S_A = S/(4A)^{\frac{1}{4}} \quad (\text{A.36-10})$$

is unity for circular cross-sections and of the order of 2 for narrow constrictions within the vocal tract (width to height ratio 9). An additional surface factor is to be expected. There is also an amplitude dependency of the damping of low frequency resonances extending down to relatively moderate sound levels, see further Eq. 1.21-15.

From Eq. A.36-7 it can be seen that the damping is inversely proportional to the fourth root of the neck area, neck length, and volume.² A decrease of neck area A by a factor of 4 lowers the pitch of the resonator one octave. This is followed by a 40 per cent increase in resonance bandwidth, i.e., an $f^{-\frac{1}{4}}$ dependency of frequency. This may in part explain the inverse bandwidth-versus-frequency relation found for the first formant at frequencies below 300 c/s. Cavity wall vibration and non-linear flow-dependent constriction-damping can also cause similar effects.

3. The Effect of a Superimposed Turbulent Air Stream. The Glottal Shunt

Investigations on the flow resistance of more or less sharp-edged orifices and tubes, see Section A.22, have shown that the main part of the flow resistance under turbulent conditions is confined to the entrance of the constricted passage where the streamlines converge and is thus independent of the length l of relatively short tubes ($l < 5 \text{ cm}$). According to Eq. 1.21-15 the differential resistance in the presence of the DC flow is of the order of $R_D = \varrho v/A$, where v is the particle velocity of the DC flow and A the constriction area. Since the inductance is $L = \varrho l/A$ the contribution to the bandwidth of the fundamental resonance should simply be

$$B_{D_1} = \frac{R_D}{2\pi L} = v/2\pi l. \quad (\text{A.36-11})$$

The corresponding time constant $1/\pi B_{D_1}$ is apparently of an order of magnitude equal to the time it takes a particle in the superimposed DC flow to be propagated down the tube.

A narrowing of the constriction area A at constant volume velocity u will cause an increase of B_{D_1} in proportion to A^{-1} of $f^{-\frac{1}{2}}$, compared with the $f^{-\frac{1}{4}}$ factor for the linear viscous losses, Eq. A.36-9. The damping associated with the turbulent DC bias appears to affect substantially the bandwidth of the first formant of voiced continuant sounds articulated with a constriction area A of the order of 0.2 cm^2 , or smaller, and

² This damping problem, applied to vocal transmission, has also been treated in part by van den Berg (1953). His statement that the bandwidth increases in inverse proportion to the third power of the constriction radius implies constant frequency conditions which can be realized only if the cavity volume is simultaneously decreased in order to compensate for the increase of constriction inductance following the narrowing of the resonator neck.

for DC volume velocities above $100 \text{ cm}^3/\text{sec}$. This damping is periodically time variable with the voice flow pulsations. Since the latter are not much larger than the F1-oscillation, it can be expected that *Eq. A.36-11* will provide values which are too large. The surface and shape dependencies are, however, not predictable and this simplified theory should be used with caution. Under the conditions of a very narrow constriction and especially if the glottis is open, thus increasing the effective volume, the first formant will be more than critically damped, and B_{D_1} of *Eq. A.36-11* is then merely the cutoff frequency of the constriction impedance, see *Section 2.64*.

The contribution of the F1-damping from the glottis termination $R_q = \rho v_q / A_q$ under the conditions of turbulent flow and a Helmholtz resonance character of F1 is

$$B_{q1} = \frac{1}{2\pi R_q C_2} = \frac{c^2 A_q}{2\pi v_q V}, \quad (\text{A.36-12})$$

where V_2 is the vocal tract volume and A_q and v_q the area and particle velocity at the glottis. The latter is approximately constant and the periodic variations are thus contained in A_q . Apart from the time delay and waveform change of the voice pulses, the variations of the energy dissipation at the glottis are thus approximately synchronous with the dissipation at the articulatory constriction. The bandwidth partial B_{q1} is of the order of 12 c/s only and thus fairly small under normal voice conditions. In the production of unvoiced turbulent continuants the open glottis will cause a free coupling down to the subglottal region, which gives rise to an appreciable energy dissipation of all formants which are appreciably dependent on the back cavity. Assuming the trachea impedance to be entirely resistive, $R_t = \rho c / A_t$ the F1-damping will be

$$B_{t1} = \frac{1}{2\pi R_t C_2} = \frac{A_t c}{2\pi V_2}, \quad (\text{A.36-13})$$

which is 240 c/s for $V_2 = 70 \text{ cm}^3$ and $A_t = 3 \text{ cm}^2$. In general, the subglottal impedance, as seen from the glottis, should first be transformed into a real part R_t and an imaginary part X_t paralleling the glottal end of the cavity system, see further *Section A.35*. The effect of the same resistive termination R_t on a standing wave resonance in a pharynx cavity of length l_2 is, according to *Eq. A.33-20*, *28*, and *Eq. A.36-1*,

$$B_{t2} = \frac{Z_2 c}{\pi R_t l_2} = \frac{\rho c^2}{\pi R_t V_2} = \frac{A_t c}{\pi V_2}, \quad (\text{A.36-14})$$

or twice the effect on the bandwidth of the first formant.

4. Heat Conduction and Cavity Wall Vibration Losses

The heat conduction losses at the cavity walls of a Helmholtz resonator of a neck area A_1 much smaller than the area A_2 of the main cavity, are small compared with the viscous neck losses. From *Eq. 1.21-13* and *A.36-2*, assuming the same shape factor for the cavity as the neck, it is found that the bandwidth contribution from the heat conduction losses within the cavity is

$$B_G = 0.45(A_1/A_2)^{\frac{1}{2}} B_R, \quad (\text{A.36-15})$$

where B_R is the contribution from the viscous losses within the neck. From Eq. 1.21-12 and A.36-2, the following fundamental expression is derived:

$$B_G = \frac{S(x-1)}{A} \left[\frac{K_h f c}{4\pi\rho C_p} \right]^{\frac{1}{2}} = 5.8(f/1000A_2)^{\frac{1}{2}} S_{A_2}, \quad (\text{A.36-16})$$

where S_{A_2} is the shape factor of the main cavity; see Eq. A.36-10 and compare Eq. A.36-7.

The heat conduction losses need to be considered only in a detailed analysis of the damping of hard-walled systems. The losses due to vocal tract cavity wall vibrations are of importance for the damping of the first formant according to van den Berg (1953). The parallel resistance $1/G' w$ of the wall impedance per unit length determines the bandwidth contribution

$$B_{G_w} = \frac{G' w}{2\pi C'_2} \quad (\text{A.36-17})$$

to the first formant of a front vowel, where C'_2 is the capacitance per unit length of the pharynx. Van den Berg (1953) estimates this bandwidth contribution to be of the order of 25-50 c/s at 300 c/s and proportional to $f^{-2.5}$. Theoretically, the muscular tension in the throat can have some influence on this damping but variations thus caused are probably imperceptible.

B. STANDING WAVES IN TUBES

1. Radiation Damping of an Open Tube

According to Eq. 1.21-16, A.33-19, A.33-28, and A.36-1, the bandwidth contribution from the radiation resistance is

$$B_{R0} = \frac{R_0 c}{Z l_e \pi} = \frac{f^2 A}{c l_e} K_s(f). \quad (\text{A.36-18})$$

$$R_0/Z < 0.5$$

The following expression is normalized with regard to an effective length of 17.6 cm and a cross-sectional area of 8 cm²:

$$B_{R0} = 29 \left(\frac{f}{1500} \right)^2 \frac{A}{8} \cdot \frac{17.6}{l_e} K_s(f). \quad (\text{A.36-19})$$

At the frequency of the second resonance, 1500 c/s, the radiation factor has reached the value $K_s(f) = 1.6$ and B_{R0} is thus 46 c/s. The fundamental resonance at 500 c/s is damped by the amount $B_{R0} = 3.9$ c/s. The same result is obtained by the formula for a simple resonator Eq. A.36-5 if the front half of the tube is regarded as the neck and the back half as the volume.

The frequency-squared dependency of the radiation resistance causes an increasing

damping of the higher formants. Thus $B_{3R_0} = 137$ c/s and $B_{4R_0} = 224$ c/s. The finite value of the radiation inductance will, however, prevent the perfect impedance matching of the tube; see *Section A.33*.

The damping is further proportional to A/l_e . Short, open-front resonators of the vocal cavity system are thus more susceptible to radiation damping than a completely open, unarticulated vocal tract.

2. Radiation Damping of Standing Waves in a Tube With a Narrow Opening

A tube of area A_2 and length l_2 is terminated by a short neck of effective length l_{1e} and a very small cross-sectional area A_1 . If $l_{1e}/c < 0.5$ and $l_1 A_2 / l_2 A_1 > 5$ the following formula may be used with a 10 per cent accuracy. It may be derived from *Eq. A.33-24B* or *A.35-12*.

$$B_{R_0} = \frac{cA_1^2}{4\pi^2 l_{1e}^2 l_2 A_2} K_s(f). \quad (\text{A.36-20})$$

The frequency dependency of R_0 is compensated for by the decreasing coupling between the radiation impedance and the main cavity. For numerical evaluations,

$$B_{R_0} = 90(A_1/l_{1e})^2(10/l_2 A_2) K_s(f). \quad (\text{A.36-21})$$

As was the case for the fundamental resonance *Eq. A.36-4*, the bandwidth contribution decreases with the square of the *lip-rounding* parameter l_{1e}/A_1 . This is the reason rounded vowels have narrower formant bandwidths than unrounded vowels.

3. Cavity Wall Damping of Standing Waves

Disregarding the effects of end corrections, *Eq. A.33-24*, i.e., assuming each end ideally closed or open,

$$Ba = \frac{c}{\pi} (\alpha_R + \alpha_G) = \frac{R}{2\pi L} + \frac{G}{2\pi C} = B_R + B_G. \quad (\text{A.36-22})$$

Here R , L , G , and C are distributed constants per unit length. The bandwidth contribution B_R from the series losses and the contribution B_G from the shunt losses for the case of an ideal hard-walled tube may be calculated by reference to *Eq. 1.21-13*. Neglecting the contribution of cavity wall vibrations to α_G , the normalized expression reduces to

$$Ba = 18.5 \left(\frac{f}{1000} \frac{1}{A} \right)^{\frac{1}{2}} S_A, \quad (\text{A.36-23})$$

and
 $B_G = 0.31 Ba;$
 $B_R = 0.69 Ba.$

It should be observed that the same shape factor S_A affects both α_R - and α_G -losses, irrespective of their physical origin. The bandwidth contributions from cavity wall losses are independent of tube length. The identities within the numerical expressions

A.36-23, A.36-16, and A.36-7 are due to the identical form of *Eq. A.36-2*, as compared with *A.36-22*.³

C. EXPERIMENTS AND CALCULATIONS ON THE PERFORMANCE OF SINGLE AND TWIN-TUBE HARD-WALLED RESONATORS

1. Single Tube, Open at one End

Frequency and bandwidth data of the resonance modes of an ideal hard-walled cylindrical tube of 8 cm^2 cross-sectional area and length 16.4 cm contained in a spherical baffle of 9 cm radius have been calculated and compared with data obtained from a brass tube model contained in a sphere of wood. The open end of the latter tube was terminated by a circular flange of 7.5 cm diameter, constituting a minor deviation from the ideal model.

A small condenser microphone placed in the bottom plane of the brass tube was utilized as a sound emitter. The complete vocal tract model was suspended freely in an anechoic chamber. A small pickup microphone was placed at a distance of 25 cm from the front end of the resonator. Measurements of resonance frequencies and the frequencies of the -3 dB points were performed with the aid of a decade frequency counter. At the temperature under consideration 21°C , the velocity of sound should be $c = 34400 \text{ cm/s}$.

The following tabulation summarizes the results:

Resonance frequency		Calculated bandwidth from			Total bandwidth	
Calculated	Measured	Friction plus heat conduction	Radiation		Calculated	Measured
F c/s	F c/s	B c/s	K_s	B_{Ro} c/s	B c/s	B c/s
486	485	4.5	1.1	3.4	7.9	8
1459	1459	7.9	1.6	43.6	51.5	44
2445	2434	10.2	1.7	133	143	128
3444	3442	12.1	1.45	225	237	228

The measured resonance frequencies agree closely with the calculated data, the differences being within the accuracy of mid-frequency evaluations. The measured

³ A recent calculation by House and Stevens (1958) on basis of a specific wall impedance $z_w = (r_w + j\omega l_w) \rho c$ for the vocal cavities (data from Franke, 1951) where $r_w = 200$ and $l_w = 0.02$ results in formant bandwidths of $B_{1G} = 70 \text{ c/s}$, $B_{2G} = 40 \text{ c/s}$, and $B_{3G} = 20 \text{ c/s}$ for a neutral vowel of $F_1 = 500 \text{ c/s}$, $F_2 = 1500 \text{ c/s}$, $F_3 = 2500 \text{ c/s}$. The contribution $B_{1G} = 70 \text{ c/s}$ above is twice that calculated by van den Berg (1953) for an [i]-model; see *Eq. A.36-17*. The discrepancy is not due to the difference in type of resonance, the first formant of [i] being a typical Helmholtz resonance and that of the neutral vowel being a quarter-wave standing wave resonance.

and calculated bandwidth data show an agreement close enough to verify the particular frequency characteristics $K_S(\omega)$ of the radiation resistance in excess of ω^2 calculated from the theory of a circular piston on the surface of a spherical baffle. The error is positive and maximally 10 per cent, but this is to be considered a small deviation in view of the departure of the model from the ideal shape.

2. Twin-Tube Resonator Model

The validity of the twin-tube resonator theory has been investigated by means of calculations and measurements on a model of the following configuration simulating the vocal cavities of a vowel [i].⁴ Front tube $A_1 = 1 \text{ cm}^2$, $l_1 = 6 \text{ cm}$, back tube $A_2 = 8 \text{ cm}^2$, $l_2 = 7.98 \text{ cm}$. A spherical baffle of 9 cm radius was utilized as in the previous example.

The end correction due to the radiation inductance is 0.45 cm. The end correction at the internal end of the front tube is 0.27 cm according to Eq. 1.21-17. Its effective length is thus $l_{1e} = 6.72 \text{ cm}$. The frequency of the fundamental resonance is $F_1 = 255 \text{ c/s}$ according to the complete twin-tube formula, Eq. A.35-1. When the simple Helmholtz resonator formula is used, Eq. A.36-3, the calculated value becomes somewhat higher, $F_1 = 264 \text{ c/s}$. If the inductance of half the back chamber is included as suggested by the low-frequency approximation of the transmission line network equivalent, the calculated value becomes too low, $F_1 = 247 \text{ c/s}$. It is obviously a better approximation to include one-third of the air in the back chamber as a contribution to the inductance which follows from the series expansion of $\cot\omega l_2/c$. The result is then $F_1 = 253 \text{ c/s}$.

The approximate value of $F_2 = c/2l_2 = 2150 \text{ c/s}$, i.e., a half-wavelength resonance of the back tube, is not far off the more exact calculated value 2045 c/s from Eq. A.35-1. The third resonance $F_3 = 2664 \text{ c/s}$ is not far off the half-wavelength resonance $c/2l_{1e} = 2560 \text{ c/s}$ of the front tube, and the fourth resonance at 4290 c/s is very close to the $c/l_2 = 4310 \text{ c/s}$ position.

Calculated and measured resonance frequencies and bandwidths can be studied in the tabulation on the next page.

The agreement between the measured and calculated data of resonance frequencies is quite good. The +1 per cent error in F_1 and the +0.4 per cent error in F_3 are attributable to the *skin effect* of the viscous boundary layer of the front tube, Eq. 1.21-11. Calculated corrections amount to -1.2 per cent and -0.4 per cent respectively. The bandwidths of those resonances that are primarily affected by damping in the front-section, i.e., B_1 and B_3 , are fairly close to calculated data, but the measured B_2 is 40 per cent larger than the calculated value. There are no comparable data in the literature on similar models. The general tendency is, however, for the damping due to viscous losses to be greater than predicted from the theory of ideal hard-walled rigid systems.

Phonetic quality slightly rounded.

Calculation of bandwidth contributions from														
	Resonance frequency Calc. Meas.		Radiation		Front tube viscosity heat condition				Back tube viscosity heat condition				Total bandwidth Calc. Meas.	
	F c/s	F c/s	B _{R0} c/s	k _{R0}	B _{R1} c/s	k _{R1}	B _{G1} c/s	k _{G1}	B _{R2} c/s	k _{R2}	B _{G2} c/s	k _{G2}	B c/s	B c/s
F ₁	255	253	0.3	0.96	6.2	0.96	0.1	0.03	0.1	0.045	1.0	0.97	7.6	8.5
F ₂	2045	2047	5.5	0.19	3.5	0.19	2.3	0.28	5.3	0.81	2.1	0.72	18.7	26
F ₃	2664	2653	40.1	0.80	16.9	0.80	7.3	0.74	1.5	0.20	0.9	0.26	70.7	74
F ₄	4290	4289	16.0	0.12	3.1	0.12	1.7	0.14	8.3	0.88	3.7	0.86	32.8	41

Any of the computed bandwidth partials is smaller than the value computed on the basis of one of the two tubes alone. The coefficients $k_v = \pi B_v / ca_v$, where $v = R1, G1, R2$, or $G2$, describe the bandwidth contributions associated with each of the four attenuation constants $a_{R1}, a_{G1}, a_{R2}, a_{G2}$. As previously mentioned they can be interpreted as indicative of the resonance-cavity dependency. The high k_{G1} of the third resonance reflects its property of standing wave in the front cavity. Similarly, the high k_{G2} of $F2$ is associated with the standing wave in the back cavity. As required by Eq. A.35-4, $k_{R1} + k_{R2} = k_{G1} + k_{G2} = 1$.

Standing waves have a pressure maximum and a particle velocity minimum at the closed end of a tube. At the open end of a tube there is a velocity maximum and a pressure minimum. If the wavelength is shorter than twice the tube length, other nodes and anti-nodes occur within the tube; see further Section 2.34.

At the frequency of a formant of a voiced sound the velocity is small but finite at the glottis, and large at the lip-opening. The velocity distribution within the vocal tract at the frequency of the second formant has one minimum besides the one at the glottal end. Two such spatial nodes occur for the third formant and three for the fourth formant and so on, see further Section 2.34.

If the length of a section or cavity within the vocal tract is large compared with the wavelength, it is apparent that the pressure dependent losses determining a_G and the velocity dependent losses determining a_R will have the normal proportion, i.e., $k_G \approx k_R$. At the frequency of a standing wave resonance both k_G and k_R are large and approach unity. At the frequency of the fundamental mode in a twin-tube resonator, on the other hand, k_G of the main cavity is large and nearly unity but k_R is small. The reverse is true of k_G and k_R of the neck section.

SELECTED BIBLIOGRAPHY

- Ayers, E. W. (1955):
“Address given at the S.R.D.E. colloquium, 1955, Ministry of Supply, Christchurch, England”,
S.R.D.E. Rep., No. 1100, 28-32 (1956).
- Barczinski, L., and Thienhaus, E. (1935):
“Klangspektren und Lautstärke deutscher Sprachlaute”, *Arch. Néerland. Phon. Exp.*, 11,
47-68 (1935).
- Bayston, T. E., and Campanella, S. J. (1957):
“Continuous analysis speech band-width compression system”, *J. Acoust. Soc. Am.*, 29,
1255 (A) (1957).
- von Békésy, G. (1929):
“Zur Theorie des Hörens. Über die eben merkbare Amplituden- und Frequenzänderung eines
Tones. Die Theorie der Schwebungen”, *Physik. Z.*, 30, 721-745 (1929).
- (1949):
“The structure of the middle ear and the hearing of one's own voice by bone conduction”,
J. Acoust. Soc. Am., 21, 217-232 (1949).
- Bell Telephone Laboratories (1946):
“Technical aspects of visible speech”, *Bell Telephone System, Monograph B-1415* (1946).—
J. Acoust. Soc. Am., 17, 1-89 (1946).
- Beranek, L. L. (1949):
Acoustic Measurements (New York, 1949).
- van den Berg, Jw. (1953):
Physica van de stemvorming, met toepassingen, diss., Rijksuniversiteit te Groningen ('s-Graven-
hage, 1953).
- (1954):
“Sur les théories myo-élastique et neuro-chronaxique de la phonation”, *Rev. de Laryng.*,
74, 494-511 (1954).
- (1954-1955):
“Über die Koppelung bei der Stimmbildung”, *Z. Phonet.*, 8, 5/6, 281-293 (1954-1955).
- (1955a):
“Calculations on a model of the vocal tract for vowel /i/ (meat) and on the larynx”, *J. Acoust.
Soc. Am.*, 27, 332-337 (1955).
- (1955b):
“On the rôle of the laryngeal ventricle in voice production”, *Folia Phoniatrica*, 7, 57-69 (1955).
- (1956):
“Direct and indirect determination of the mean subglottic pressure”, *Folia Phoniatrica*, 8,
1-24 (1956).

- (1957):
“Subglottic pressures and vibrations of the vocal folds”, *Folia Phoniatrica*, 9, 65-71 (1957).
- , Zantema, J. T., and Doornenbal Jr., P. (1957):
“On the air resistance and the Bernoulli effect of the human larynx”, *J. Acoust. Soc. Am.*, 29, 626-631 (1957).
- Boeryd, A. (1957):
“Undersökning av taleffekten (volymen) från en telefonapparat som funktion av telefonförbindelsens kvalitet.” Examensarbete i telegrafi och telefoni, Royal Institute of Technology, Stockholm (1957).
- Bogert, B. P. (1953):
“On the bandwidth of vowel formants”, *J. Acoust. Soc. Am.*, 25, 791-792 (1953).
- Boyanus, S. C. (1944):
A Manual of Russian Pronunciation, 2nd ed. (London, 1944).
- Broch, O. (1911):
Slavische Phonetik (Heidelberg, 1911).
- Chang, S.-H. (1956):
“Two schemes of speech compression system”, *J. Acoust. Soc. Am.*, 28, 565-572 (1956).
- , Stubbs, H. L., Dolanský, L. O., Wiren, J., Denes, P., Howard, C. R., and Carrabes, M. J. (1956):
“Visual message presentation”, *Northeastern Univ., Electronics Res. Lab., Scientific Report*, No. 5 (AFCRC-TN-56-582) (1956).
- Cherry, C. (1957):
On Human Communication (London, 1957).
- , Halle, M., and Jakobson, R. (1953):
“Toward a logical description of languages in their phonemic aspects”, *Language*, 29, 34-46 (1953).
- Chiba, T., and Kajiyama, M. (1941):
The Vowel—Its Nature and Structure (Tokyo, 1941).
- Cooper, F. S., Delattre, P. C., Liberman, A. M., Borst, J. M., and Gerstman, L. J. (1952):
“Some experiments on the perception of synthetic speech sounds”, *J. Acoust. Soc. Am.*, 24, 597-606 (1952).
- Crandall, I. B. (1925):
“Sounds of speech”, *Bell System Techn. J.*, 4, 586-626 (1925).
- (1927):
“Dynamical study of the vowel sounds”, *Bell System Techn. J.*, 6, 100-116 (1927).
- David Jr., E. E. (1956):
“Signal theory in speech transmission”, *Bell Telephone System, Monograph 2831* (1956).
- (1958):
“Artificial Auditory Recognition in Telephony”, *IBM J. of research and development*, 4, 294-309 (1958).
- Delattre, P. (1951):
“The physiological interpretation of sound spectrograms”, *PMLA*, LXVI, 864-875 (1951).
- (1954):
“Les attributs acoustiques de la nasalité vocalique et consonantique”, *Studia Linguistica*, VIII, 103-109 (1954).
- , Liberman, A. M., and Cooper, F. S. (1951):
“Voyelles synthétiques à deux formantes et voyelles cardinales”, *Le Maître Phonétique*, 96, 30-36 (1951).
- , —, — (1955):
“Acoustic loci and transitional cues for consonants”, *J. Acoust. Soc. Am.*, 27, 769-773 (1955).

- , Liberman, A. M., Cooper, F. S., and Gerstman, L. J. (1952):
 "An experimental study of the acoustic determinants of vowel color", *Word*, 8, 195-210 (1952).
- Dieth, E. (1950):
Vademekum der Phonetik (Bern, 1950).
- Dudley, H. (1939):
 "Remaking speech", *J. Acoust. Soc. Am.*, 11, 165-177 (1939).
- (1956):
 "Fundamentals of speech synthesis", *Bell Telephone System Monograph 2648* (1956).
- , Riesz, R. R., and Watkins, S. S. A. (1939):
 "A synthetic speaker", *J. Franklin Inst.*, 227, 739 (1939).
- Dunn, H. K. (1950):
 "The calculation of vowel resonances and an electrical vocal tract", *J. Acoust. Soc. Am.*, 22, 740-753 (1950).
- , and White, S. D. (1940):
 "Statistical measurements on conversational speech", *J. Acoust. Soc. Am.*, 11, 278 (1940).
- Eck, B. (1944):
Technische Strömungslehre (Berlin, 1944).
- Essner, C. (1947):
 "Recherches sur la structure des voyelles orales", *Arch. Néerland. Phon. Exp.*, 20, 40-77 (1947).
- Faaborg-Andersen, K. (1957):
Electromyographic Investigation of Intrinsic Laryngeal Muscles in Humans (Copenhagen, 1957).
- Fano, R. M. (1950):
 "The information theory point of view in speech communication", *J. Acoust. Soc. Am.*, 22, 691-696 (1950).
- Fant, C. G. M. (1948):
 "Analys av de svenska vokalljuden", *L M Ericsson protokoll H/P 1035* (1948).
- (1949):
 "Analys av de svenska konsonantljuden", *L M Ericsson protokoll H/P 1064* (1949).
- (1950a):
 "Transmission properties of the vocal tract", *M.I.T. Acoustics Lab. Quarterly Progress Rep.*, July-Sep., 20-23 (1950).
- (1950b):
 "Transmission properties of the vocal tract, Part II", *M.I.T. Acoustics Lab. Quarterly Progress Rep.*, Oct.-Dec., 14-19 (1950).
- (1952):
 "Transmission properties of the vocal tract with application to the acoustic specification of phonemes", *M.I.T. Acoustics Lab. Techn. Rep.*, No. 12 (1952).
- (1953a):
 "Speech communication research", *IVA*, 24 (Royal Swedish Academy of Engineering Sciences), 331-337 (1953).
- (1953b):
 "Discussion of paper read by G. E. Peterson at the 1952 Symposium on the Applications of Communication Theory", *Communication Theory*, ed. W. Jackson, 421-424 (London, 1953).
- (1956):
 "On the predictability of formant levels and spectrum envelopes from formant frequencies", *For Roman Jakobson*, 109-120 ('s-Gravenhage, 1956).
- (1957):
 "Modern instruments and methods for acoustic studies of speech", *Royal Inst. of Technology, Div. of Telegraphy-Telephony, Rep.*, No. 8 (1957).—*Proc. of the VIII Int. Congr. of Linguists 1957*, 282-358 (Oslo, 1958). Also published in *Acta Polytechnica Scandinavica*, No. 1, 1-81 (1958).

- (1958):
 “Acoustic theory of speech production”, *Royal Inst. of Technology, Div. of Telegraphy-Telephony, Rep.*, No. 10.—Presented in mimeographed form (1958).
- (1959):
 “Acoustic analysis and synthesis of speech with applications to Swedish”, *Ericsson Technics*, No. 1, 3-108 (1959).
- Farnsworth, D. W. (1940):
 “High speed motion pictures of the human vocal cords”, *Bell Telephone Lab., Record*, 18, 203 (1940).
- Fischer-Jørgensen, E. (1954):
 “Acoustic analysis of stop consonants”, *Miscellanea Phonetica*, II, 42-59 (1954).
- (1956):
 “The commutation test and its applicability to phonemic analysis”, *For Roman Jakobson*, 140-151 ('s-Gravenhage, 19' 6).
- (1957):
 “What can the new techniques of acoustic phonetics contribute to linguistics?”, *Proc. of VIII Int. Congr. of Linguists 1957*, 433-478 (Oslo, 1958).
- (1959):
 “Die Bedeutung der funktionellen Sprachbeschreibung für die Phonetik”, *Phonetica, Suppl. ad Vol. 4*, 7-28 (1959).
- , and Tybjærg Hansen, A. (1959):
 “An electrical manometer and its use in phonetic research”, *Phonetica*, 4, 43-53 (1959).
- Flanagan, J. L. (1955a):
 “A speech analyzer for a formant-coding compression system”, *M.I.T. Acoustics Lab. Scientific Rep.*, No. 4 (AFCRC-TN-55-793) (1955).
- (1955b):
 “Difference limen for vowel formant frequency”, *J. Acoust. Soc. Am.*, 27, 613-617 (1955).
- (1956):
 “Band width and channel capacity necessary to transmit the formant information of speech”, *J. Acoust. Soc. Am.*, 28, 592-596 (1956).
- (1957a):
 “Estimates of the maximum precision necessary in quantizing certain ‘dimensions’ of vowel sounds”, *J. Acoust. Soc. Am.*, 29, 533-534 (1957).
- (1957b):
 “Note on the design of ‘terminal-analog’ speech synthesizers”, *J. Acoust. Soc. Am.*, 29, 306-310 (1957).
- (1958):
 “Some properties of the glottal sound source”, *J. of Speech and Hearing Research*, 1, 99-116 (1958).
- (1959):
 “A resonance-vocoder and baseband complement: A hybrid system for speech transmission”, *IRE WESCON Convention Record*, 3, 5-16 (1959).
- , and House, A. S. (1956):
 “Development and testing of a formant-coding speech compression system”, *J. Acoust. Soc. Am.*, 28, 1099-1106 (1956).
- Fletcher, H. (1929):
Speech and Hearing (New York, 1929; later edition 1953).

- Forchhammer, J. (1942):
Die Sprachlaute in Wort und Bild (Heidelberg, 1942).
- (1954):
“Almindelig talelaere”, Chapter XII in *Nordisk Laerebog for Talepaedagoger, Almindelig Del*, 218-242 (1954).
- Fry, D. B. (1955):
“Duration and intensity as physical correlates to stress”, *J. Acoust. Soc. Am.*, 27, 765-768 (1955)
- Gabor, D. (1946):
“Theory of communication”, *J. Inst. Elect. Engrs.*, 93, Part III, 429-457 (1946).
- (1953):
“A summary of communication theory”, *Proc. of the 1952 Symposium on the Applications of Communication Theory. Communication Theory*, ed. W. Jackson, 1-24 (London, 1953).
- Ganeshsundaram, P. C. (1957):
“A cascade modulation theory of speech formants”, *Z. f. Phonet.*, 10/1, 1-7 (1957).
- Garner, W. R. (1949):
“The loudness and loudness matching of short tones”, *J. Acoust. Soc. Am.*, 21, 398-403 (1949).
- Grützmacher, M., and Lottermoser, W. (1937):
“Über ein Verfahren zur trägeheitsfreien Aufzeichnung von Melodiekurven”, *Akust. Z.*, 242-248 (1937).
- Haase, K. H., and Vilbig, F. (1956):
“Errors in spectrum analysis by a set of narrow band selecting filters”, *AFCRC, Bedford, Communications Lab.* (AFCRC-TR-56-121) (1956).
- Hála, B. (1956):
“Nature Acoustique des Voyelles”, *Acta Universitatis Carolinae* (Prague, 1956).
- Halle, M. (1954a):
“The strategy of phonemics”, *Word*, 10, 197-209 (1954).
- (1954b):
The Russian Consonants. A Phonemic and Acoustical Investigation, Dr. Phil. thesis, Harvard University (1954).
- (1956):
Review of *Manual of Phonology* by C. F. Hockett, in *J. Acoust. Soc. Am.*, 28, 509-511 (1956).
- (1957):
“In defense of the number two”, *Studies Presented to Joshua Whatmough on His Sixtieth Birthday*, 65-72 ('s-Gravenhage, 1957).
- (1959):
The Sound Pattern of Russian ('s-Gravenhage, 1959).
- , Hughes, G. W., and Radley, J. P. (1957):
“Acoustic properties of stop consonants”, *J. Acoust. Soc. Am.*, 29, 107-116 (1957).
- Halsey, R. J., and Swaffield, J. (1948):
“Analysis-synthesis telephony, with special reference to the vocoder”, *Proc. Inst. Elect. Engrs.* 95, 391-411 (1948).
- Hardy, H. C., and others (1957):
“Symposium on sound level meters”, *J. Acoust. Soc. Am.*, 29, 1330-1341 (1957).
- Harris, C. M. (1953):
“A study of the building blocks in speech”, *J. Acoust. Soc. Am.*, 25, 962-969 (1953).
- (1953):
“A speech synthesizer”, *J. Acoust. Soc. Am.*, 25, 970-975 (1953).
- Harris, Z. S. (1951):
Methods in Structural Linguistics (Chicago, 1951).

- Hartley, R. V. L. (1928):
 "Transmission of information", *Bell System Techn. J.*, 7, 535 (1928).
- Hattori, S., Yamamoto, K., and Fujimura, O. (1956):
 "Nasalization of vowels and nasals", *Bull. of the Kobayashi Inst. of Phys. Res.*, 6, 226-235 (1956).
- Heffner, R. M. S. (1949):
General Phonetics (Madison, 1949).
- Heinz, J. M. (1956):
 "Fricative consonants", *M.I.T. Acoustics Lab. Quarterly Rep.*, Oct.-Dec., 5-7 (1956).
- (1957a):
 "Fricative consonants", *M.I.T. Acoustics Lab. Quarterly Rep.*, April-June, 1 (1957).
- (1957b):
 "A terminal analog of fricative consonant articulation", *M.I.T. Acoustics Lab. Quarterly Rep.*, July-Sep., 1-3 (1957).
- Hockett, C. F. (1955):
Manual of Phonology, Indiana Univ. Publications in Anthropology and Linguistics, No. 11 (Bloomington, 1955).
- House, A. S. (1957):
 "Analog studies of nasal consonants", *J. of Speech and Hearing Disorders*, 22, 190-204 (1957).
- , and Stevens, K. N. (1956):
 "Analog studies of the nasalization of vowels", *J. of Speech and Hearing Disorders*, 21, 218-232 (1956).
- , — (1957):
 "Measurements of the transient response of the vocal tract", *M.I.T. Acoustics Lab. Quarterly Rep.*, July-Sep., 3-5 (1957).
- , — (1958):
 "Estimation of formant band widths from measurements of transient response of the vocal tract", *J. of Speech and Hearing Research*, 1, 309-315 (1958).
- Howard, C. R. (1956):
 "Speech analysis-synthesis scheme using continuous parameters", *J. Acoust. Soc. Am.*, 28, 1091-1098 (1956).
- Huggins, W. H. (1952):
 "A phase principle for complex-frequency analysis and its implications in auditory theory", *J. Acoust. Soc. Am.*, 24, 582-589 (1952).
- Hughes, G. W., and Halle, M. (1956):
 "Spectral properties of fricative consonants", *J. Acoust. Soc. Am.*, 28, 303-310 (1956).
- Husson, R. (1950):
Étude des phénomènes physiologiques et acoustiques fondamentaux de la voix chantée, thèse, l'Univ. de Paris (Paris, 1950).
- (1956):
 "Stemmebandsvibrationernes fysiologi" (translation by M. Kloster-Jensen), *Nord. Tidskrift for Tale og Stemme*, 16, 49-73 (1956).
- Ingård, U. (1953):
 "On the theory and design of acoustic resonators", *J. Acoust. Soc. Am.*, 25, 1037-1067 (1953).
- (1956):
 "On the spectra of explosive speech sounds", *M.I.T. Acoustics Lab. Quarterly Rep.*, July-Sep., 13-15 (1956).
- Jacobson, H. (1931):
 "Information and the human ear", *J. Acoust. Soc. Am.*, 23, 463-471 (1951).
- Jakobson, R. (1939):
 "Observations sur le classement phonologique des consonnes", *Proc. of the III Int. Congr. of Phonetic Sciences, Ghent 1939*, 34 (1939).

- (1940):
Kindersprache, Aphasie und allgemeine Lautgesetze. Språkvetenskapl. Sällskapets i Uppsala Förhandl. (1940-1942).
- (1956):
“Die Verteilung der stimmhaften und stimmlosen Geräuschlaute im Russischen”, *Festschrift für Max Vasmer*, 199-202 (Berlin, 1956).
- , Fant, C. G. M., and Halle, M. (1952):
“Preliminaries to speech analysis. The distinctive features and their correlates”, *M.I.T. Acoustics Lab. Techn. Rep.*, No. 13 (1952); 3rd printing.
- , and Halle, M. (1956):
Fundamentals of Language ('s-Gravenhage, 1956).
- , — (1957):
“Phonology in relation to phonetics,” *Manual of Phonetics*, 215-251 (Amsterdam, 1957).
- Jassem, W. (1959):
“The phonology of Polish stress”, *Word*, 15, 252-269 (1959).
- Jones, D. (1934):
An Outline of English Phonetics (Leipzig, 1934).
- Jones, L. G. (1952):
Acoustic Patterns of the Russian Vowels, Dr. Phil. thesis, Harvard University (1952).
- (1953):
“The Vowels of English and Russian: An Acoustic Comparison”, *Word*, 9, 354-361 (1953).
- Joos, M. (1948):
“Acoustic Phonetics”, *Language*, 24, 1-136 (1948).
- Koenig, W. (1949):
“A new frequency scale for acoustic measurements”, *Bell Telephone Lab., Record*, 27, 299-301 (1949).
- Koneczna, H., and Zawadowski, W. (1956):
Obrazy Rentgenograficzne Głosek Rosyjskich (Warszawa, 1956).
- Küpfmüller, K., and Warns, O. (1956):
“Sprachsynthese aus Lauten”, *Nachrichtentechn. Fachber.*, 3, 28-31 (1956).
- Ladefoged, P., and Broadbent, D. E. (1957):
“Information conveyed by vowels”, *J. Acoust. Soc. Am.*, 29, 98-104 (1957).
- Laurent, T. (1940):
“Matematisk behandling av kontinuerligt inhomogena ledningar medelst ekvivalenter samt exempel på metodens användning för olika praktiska problem”, *Tekn. Medd. K. Telegrafstyrelsen*, 113-133 (1940).
- (1940):
“Om kontinuerligt inhomogena ledningar”, *Tekn. Medd. K. Telegrafstyrelsen*, 186 (1940).
- (1953):
“Delay time and transient time in electrical filters with phase distortion”, *Proc. of the 1952 Symposium on the Applications of Communication Theory. Communication Theory*, ed. W. Jackson, 310-313 (London, 1953).
- (1956):
Vierpoltheorie und Frequenztransformation (Berlin, 1956).
- Lawrence, W. (1953):
“The synthesis of speech from signals which have a low information rate”, *Proc. of the 1952 Symposium on the Applications of Communication Theory. Communication Theory*, ed. W. Jackson, 460-469 (London, 1953).
- Lewis, D. (1936):
“Vocal resonance”, *J. Acoust. Soc. Am.*, 8, 91 (1936).

- Liberman, A. M. (1957):
 "Some results of research on speech perception", *J. Acoust. Soc. Am.*, 29, 117-123 (1957).
- , Delattre, P. C., Gerstman, L. J., and Cooper, F. S. (1956):
 "Tempo of frequency change as a cue for distinguishing classes of speech sounds", *J. of Experimental Psychology*, 52, 127-137 (1956).
- Licklider, J. C. R. (1951):
 "Basic correlates of the auditory stimulus", *Handbook of Experimental Psychology*, 985-1039 (New York, 1951).
- (1952):
 "On the process of speech perception", *J. Acoust. Soc. Am.*, 24, 590-594 (1952).
- , and Miller, G. A. (1951):
 "The perception of speech", *Handbook of Experimental Psychology*, 1040-1074 (New York, 1951).
- Lisker, L. (1957):
 "Closure duration and the intervocalic voiced-voiceless distinction in English", *Language*, 33, 42-49 (1957).
- Lotz, J. (1950):
 "Speech and language", *J. Acoust. Soc. Am.*, 22, 712-717 (1950).
- (1954):
 "The structure of human speech", *Transactions of the New York Academy of Sciences*, Ser. II, 16, No. 7, 373-384 (1954).
- Lundell, J. A. (1890):
Étude sur la prononciation russe (Stockholm, 1890).
- MacMillan, A. S., and Kelemen, G. (1952):
 "Radiography of the supraglottic speech organs", *A.M.A. Archives of Otolaryngology*, 55, 681-682 (1952).
- Malécot, A. (1955):
 "An experimental study of force of articulation", *Studia Linguistica*, IX, 35-44 (1955).
- (1956):
 "Acoustic cues for nasal consonants", *Language*, 32, 274-284 (1956)..
- Malmberg, B. (1952):
 "Le problème du classement des sons du langage et quelques questions connexes", *Studia Linguistica*, VI, 1-56 (1952).
- (1956):
 "Distinctive features of Swedish vowels; some instrumental and structural data", *For Roman Jakobson*, 316-321 ('s-Gravenhage, 1956).
- Mandelbrot, B. (1953):
 "An informational theory of the statistical structure of language", *Proc. of the 1952 Symposium on the Applications of Communication Theory. Communication Theory*, ed. W. Jackson, 486-500 (London, 1953).
- Mason, W. P. (1948):
Electromechanical transducers and wave filters (New York, 1948).
- Menzerath, P., and de Lacerda, A. (1933):
Koartikulation, Steuerung und Lautabgrenzung (Berlin-Bonn, 1933).
- Meyer-Eppler, W. (1950):
 "Die Schwingungsanalyse nach dem Suchton-Verfahren", *Archiv der Elektr. Übertragung*, 4, 331-338 (1950).
- (1953):
 "Zum Erzeugungsmechanismus der Geräuschlaute", *Z. für Phonetik*, 7, Nr. 3/4, 196-212 (1953).

- Miller, G. A. (1948):
"The perception of short bursts of noise", *J. Acoust. Soc. Am.*, 20, 160-170 (1948).
- (1951):
Language and Communication (New York, 1951).
- , and Nicely, P. E. (1955):
"An analysis of perceptual confusions among some English consonants", *J. Acoust. Soc. Am.*, 27, 338-352 (1955).
- Miller, R. L. (1956):
"Nature of the vocal cord wave", *J. Acoust. Soc. Am.*, 28, 159 (1956).
- Morse, Ph. M. (1948):
Vibration and Sound (New York, 1948).
- Munson, W. A. (1947):
"The growth of auditory sensation", *J. Acoust. Soc. Am.*, 19, 584-591 (1947).
- Nielsen, A. K. (1949):
"Acoustic resonators of circular cross-section and with axial symmetry", *Trans. Dan. Acad. Techn. Sci.*, 10, 9-70 (1949).
- Ochiai, Y., Fukumura, T., and Nakatani, K. (1957):
"Timbre study of nasalics, Part II", *Memoirs of the Faculty of Engineering, Nagoya University*, 9, 160-173 (1957).
- O'Connor, J. D., Gerstman, L. J., Liberman, A. M., Delattre, P. C., and Cooper, F. S. (1957):
"Acoustic cues for the perception of initial /w, j, r, l/ in English", *Word*, 13, 24-43 (1957).
- Paget, Sir R. (1930):
Human Speech (London, 1930).
- Parmenter, C. E., and Treviño, S. N. (1932):
"Vowel positions as shown by X-rays", *The Quarterly J. of Speech*, XVIII, 351-369 (1932).
- Peterson, G. E. (1951):
"The phonetic value of vowels", *Language*, 27, 541-553 (1951).
- (1952a):
"Application of information theory to research in experimental phonetics", *J. Speech and Hearing Disorders*, 17, 175 (1952).
- (1952b):
"The information bearing elements of speech", *J. Acoust. Soc. Am.*, 24, 629-637 (1952).
- (1955):
"An oral communication model", *Language*, 31, 414-427 (1955).
- (1957):
"Fundamental problems in speech analysis and synthesis", *Proc. of the VIII Int. Congr. of Linguists 1957*, 267-281 (Oslo, 1958).
- , and Barney, H. L. (1952):
"Control methods used in a study of the vowels", *J. Acoust. Soc. Am.*, 24, 175-184 (1952).
- Peterson, E., and Cooper, F. S. (1957):
"Peakpicker: A band-width compression device", *J. Acoust. Soc. Am.*, 29, 777 (1957).
- Pierce, J. R., and David Jr., E. E. (1958):
Man's world of sound (New York, 1958).
- Pollack, I. (1952):
"The information of elementary auditory displays", *J. Acoust. Soc. Am.*, 24, 745-749 (1952).
- (1953):
"The information of elementary auditory displays, II", *J. Acoust. Soc. Am.*, 25, 765-769 (1953).
- , and Ficks, L. (1954):
"Information of elementary multidimensional auditory displays", *J. Acoust. Soc. Am.*, 26, 155-158 (1954).

- Polland, B., and Hála, B. (1926):
Les radiographies de l'articulation des sons tchèques (Prague, 1926).
- Potter, R. K., Kopp, A. G., and Green, H. C. (1947):
Visible Speech (New York, 1947).
- , and Steinberg, J. C. (1950):
 "Toward the specification of speech", *J. Acoust. Soc. Am.*, 22, 807-820 (1950).
- Raleigh, Lord (1896):
Theory of Sound (London, 1896).
- Rösler, G. (1957):
 "Über die Vibrationsempfindung. Literaturdurchsicht und Untersuchungen im Tonfrequenzbereich", *Z. f. exper. u. angew. Psych.*, 4, 549-602 (1957).
- Russel, G. O. (1928):
The Vowel (Columbus, 1928).
- Schatz, C. D. (1954):
 "The role of context in the perception of stops", *Language*, 30, 47-56 (1954).
- Schlichting, H. (1951):
Grenzschicht-Theorie (Karlsruhe, 1951).
- Shannon, C. E. (1951):
 "Prediction and entropy of printed English", *Bell System Techn. J.*, 30, 50-64 (1951).
- , and Weaver, W. (1949):
The Mathematical Theory of Communication (Urbana, 1949).
- Smith, S. (1947):
 "Analysis of vowel sounds by ear", *Arch. Nederl. Phon. Exp.*, XX, 78-96 (1947).
- (1951):
 "Vocalization and added nasal resonance", *Folia Phoniatrica*, 3, 165-169 (1951).
- (1954):
 "Remarks on the physiology of the vibrations of the vocal cords", *Folia Phoniatrica*, 6, 166-178 (1954).
- Snow, W. B. (1957):
 "Rectification in the sound level meter", *J. Acoust. Soc. Am.*, 29, 1338 (1957).
- Sovijärvi, A. (1938a):
Die gehaltenen, geflüsterten und gesungenen Vokale und Nasale der finnischen Sprache (Helsinki, 1938).
- (1938b):
 "Die wechselnden und festen Formanten der Vokale erklärt durch Spektrogramme und Röntgengramme der finnischen Vokale", *Proc. of the III Int. Congr. of Phonetic Sciences, Ghent, 1938*.
- Steinberg, J. C. (1934):
 "Application of sound measuring instruments to the study of phonetic problems", *J. Acoust. Soc. Am.*, 6, 16-24 (1934).
- Stetson, R. H. (1951):
Motor Phonetics (Amsterdam, 1951).
- Stevens, K. N. (1956):
 "Stop consonants", *M.I.T. Acoustics Lab. Quarterly Rep.*, Oct.-Dec., 7-8 (1956).
- (1958):
 "Research on speech synthesis", *M.I.T. Acoustics Lab. Scientific Rep.*, No. 17 (AFCRC-TN-58-140) (1958).
- (1960):
 "Toward a model for speech recognition", *J. Acoust. Soc. Am.*, 32, 47-55 (1960).
- , and House, A. S. (1955):
 "Development of a quantitative description of vowel articulation", *J. Acoust. Soc. Am.*, 27, 484-493 (1955).

- , — (1956):
“Studies of formant transitions using a vocal tract analog”, *J. Acoust. Soc. Am.*, 28, 578-585 (1956).
- , Kasowski, S., and Fant, C. G. M. (1953):
“An electrical analog of the vocal tract”, *J. Acoust. Soc. Am.*, 25, 734-742 (1953).
- Stevens, S. S. (1956):
“Calculation of the loudness of complex noise”, *J. Acoust. Soc. Am.*, 28, 807-832 (1956).
- , and Davis, H. (1938):
Hearing (New York, 1938; 1947).
- , and Volkmann, J. (1940):
“The relation of pitch to frequency”, *Amer. J. Psychology*, 53, 329-353 (1940).
- Sund, H. (1957):
“A sound spectrometer for speech analysis”, *Transactions of the R.I.T.*, No. 112 (1957).
- Tarnóczy, T. (1948):
“Resonance data concerning nasals, laterals and trills”, *Word*, 4, 71-77 (1948).
- Trendelenburg, F. (1950):
Einführung in die Akustik, Zweite Auflage, “Die menschliche Stimme”, 138-150; “Physikalische Eigenschaften natürlicher Schallvorgänge”, 359-362 (Berlin, 1950).
- Truby, H. M. (1957):
“A note on visible and indivisible speech”, *Proc. of the VIII Congr. of Linguists 1957*, 393-400 (Oslo, 1958).
- (1959):
“Acoustico-cineradiographic analysis considerations with especial reference to certain consonantal complexes”, *Acta Radiologica, Suppl.* 182, 1-227 (1959).
- Tuller, W. G. (1949):
“Theoretical limits on the rate of transmission of information”, *Proc. I.R.E.*, 37, 468 (1949).
- Vilbig, F., and Haase, K. H. (1956):
“Some systems for speech-band compression”, *J. Acoust. Soc. Am.*, 28, 573-577 (1956).
- Weibel, E. S. (1955):
“Vowel synthesis by means of resonant circuits”, *J. Acoust. Soc. Am.*, 22, 858-865 (1955).
- Westervelt, P. J., and Sieck, P. W. (1950):
“The correlation of nonlinear flow and differential resistance for sharp-edged circular orifices”, *M.I.T. Acoustics Lab. Quarterly Progress Rep.*, Apr.- June, 24-28 (1950).
- Witting, C. (1959):
“Physical and functional aspects of speech sounds with special application to standard Swedish”, *Uppsala Universitets Årsskrift 1959:7*, 1-151 (1959).

AUTHOR INDEX

- Barczinski, L., 238
Barney, H. L., 237
von Békésy, G., 131, 132, 233
Beranek, L. L., 230, 231
van den Berg, Jw., 33, 35, 38, 45, 105, 131, 132, 136, 138, 242, 265, 266, 267, 268, 269, 272, 274, 305, 307, 309
Boeryd, A., 233
Bogert, B. P., 242
Borst, J. M., 148
Boyanus, S. C., 110, 216
- Chang, S-H., 272
Chiba, T., 35, 85, 86, 87, 97, 105, 113, 127, 131, 237, 269, 271, 284
Cooper, F. S., 25, 123, 148, 218
Crandall, I. B., 235, 284
- Delattre, P. C., 25, 26, 112, 123, 148, 160, 161, 218
Dieth, E., 184
Doornenbal Jr., P., 132, 265, 266, 267, 268, 274
Dudley, H., 18
Dunn, H. K., 35, 38, 113, 230, 232, 242
- Eck, B., 273
Essner, C., 28
- Faaborg-Andersen, K., 265
Fant, G., 18, 22, 23, 25, 26, 42, 46, 47, 48, 50, 52, 56, 58, 59, 99, 123, 124, 127, 188, 200, 208, 212, 213, 214, 215, 219, 230, 233, 234, 237, 238, 241, 242, 271, 272, 300
Farnsworth, D. W., 265, 266
Fischer-Jørgensen, E., 19, 23, 179, 208, 214, 218
Flanagan, J. L., 47, 268
Fletcher, H., 113, 233
Forchhammer, J., 97
Fry, D. B., 214
Fujimura, O., 149, 160
Fukumura, T., 146
- Gabor, D., 236
Ganeshsundaram, P. C., 113
Garner, W. R., 233
Gerstman, L. J., 123, 148
Green, H. C., 22, 25
Grützmacher, M., 241
- Haase, K. H., 231
Hála, B., 97, 113
Halle, M., 18, 26, 58, 59, 124, 170, 178, 188, 191, 203, 208, 212, 215, 216, 219, 225
Hardy, H. C., 230
Hattori, S., 149, 160
Heffner, R. M. S., 113
Heinz, J. M., 173, 203, 274, 280, 298
House, A. S., 25, 26, 64, 72, 81, 87, 111, 124, 138, 141, 146, 149, 161, 210, 218, 242, 309
Huggins, W. H., 39, 235
Hughes, G. W., 26, 178, 188, 191, 203, 208, 216
Husson, R., 265
- Ingård, U., 34, 36, 280, 281, 283, 298
Jakobson, R., 18, 58, 59, 124, 208, 212, 215, 216, 219
Jones, D., 111, 113
Jones, L. G., 110
Joos, M., 111, 148, 232, 284
- Kajiyama, M., 35, 85, 86, 87, 97, 105, 113, 127, 131, 237, 269, 271, 284
Kasowski, S., 52, 99, 271
Kelemen, G., 93, 97
Koenig, W., 241
Koneczna, H., 97, 184
Kopp, A. G., 22, 25
- Laurent, T., 29, 30, 232
Lewis, D., 235
Liberman, A. M., 25, 123, 148, 218
Licklider, J. C. R., 236

- Lisker, L., 225
Lottermoser, W., 241
- MacMillan, A. S. 93, 97
Malécot, A., 23, 148, 280
Malmberg, B., 113
Mason, W. P., 33
Meyer-Eppler, W., 184, 238, 273, 274, 275, 279
Miller, G. A., 233
Miller, R. L., 272
Morse, Ph. M., 29, 35, 44, 293
Munson, W. A., 233
- Nakatani, K., 146
Nielsen, A. K., 281
- Ochiai, Y., 146
- Paget, Sir R., 284
Parmenter, C. E., 97
Peterson, G. E., 237, 242
Pollard, B., 97
Potter, R. K., 22, 25, 237
- Radley, J. P., 26, 188, 191, 203, 216
Raleigh, Lord, 237, 284
Riesz, R. R., 18
Rösler, G., 131
Russel, G. O., 97, 113
- Schatz, C. D., 19, 208
- Schlichting, H., 273
Sieck, P. W., 269, 274
Smith, S., 148, 160, 265, 266
Snow, W. B., 231
Sovijärvi, A., 97, 105, 113, 235, 238
Steinberg, J. C., 235, 237
Stetson, R. H., 280
Stevens, K. N., 25, 26, 34, 52, 64, 72, 81, 87, 99,
111, 124, 125, 138, 141, 149, 161, 210, 218, 242,
271, 276, 279, 309
Stevens, S. S., 234, 238
Sund, H., 236
- Tarnóczy, T., 235
Thienhaus, E., 238
Trendelenburg, F., 33, 237
Treviño, S. N., 97
Truby, H. M., 207
- Vilbig, F., 231
Volkmann, J., 238
- Watkins, S. S. A., 18
Weibel, E. S., 47
Westervelt, P. J., 269, 274
White, S. D., 230, 232, 242
- Yamamoto, K., 149, 160
- Zantema, J. T., 132, 265, 266, 267, 268, 274
Zawadowski, W., 97, 184

SUBJECT INDEX

- Acoustic capacitance, 27, 29
impedance, 16, 27
inductance, 27, 29
resistance, 27, 32, 34
termination, 294
- Acuteness, 218
- Admittance, 145
- Aeolian tones, 273
- Affricate, 169
- Air consumption, 269
- Airflow, 186, 275
- Amplitude-frequency section, 243
- Analytical constraint, 48
- Angular frequency, 39
- Area function, 106
- Articulation, 17
- Aspiration, 19
- Attenuation constant, 28
- Back cavity, 15, 43, 113
- Baffle effect, 44
- BARK computer, 42
- Bernouilli effect, 265
- Bernouilli force, 179
- BESK computer, 99
- Boyle's law, 276
- Burst, 208
- Calculations (of formant frequencies), 107
- Capacitance, 27
- Cavity formant relations, 71, 113, 120
- Cavity models, 63
- Cavity wall vibration, 43, 131, 138
- Center of gravity, 60
- Centralization, 81
- Characteristic impedance, 28
- Compactness, 58, 217
- Compensatory articulation, 66
- Complex frequency, 16, 39
- Computer calculations, 36, 107
- Computer calculation of vocal response, 107
- Conductance, 27
- Conjugate poles, 42
- Consonants
affricate, 169
fricative, 169
lateral, 163
nasal, 139
r-sound, 162
stop, 185
- Constriction, 72
- Continuous, 18
- Coupling, 61
- Critical damping, 282
- Cross-sectional area, 16, 27
- Cues, 213
- Current, 16
- Current transfer ratio, 38, 39
- Damping, 300
- Damping constant, 39
- Delay time, 232
- Density of air, ρ , 27
- Dentals, 193
- Determinant calculations, 37
- Diffuseness, 58, 217
- Distinctive features, 212
- Distributed elements, 28
- Double Helmholtz resonator, 117, 282
- Dyne/cm², 16, 230
- Effective length, 292
- Efficiency of voice, 272
- Electrical analog (of vocal tract), 27
- End correction, 36, 292
- Energy, 229
- Energy spread, 207
- Equivalent circuits, 27
- False vocal cords, 102
- Female F-patterns, 242
- Filter, 15, 17

- Filter function, 16
 Flatness, 219
 Formant, 20, 47, 241
 amplitude, 47
 bandwidth, 47
 density, 62
 frequency, 47
 level, 56, 241
 spread, 60
 Fortis, 24, 224, 279
 Fourier analysis, 235
 Four-tube models, 74, 297
 Four-tube systems, 297
 F-pattern, 24, 209, 221
 F-pattern envelope relations, 56
 Frequency of voice fundamental, 17, 241
 Frication, 19
 Fricational losses, 32
 Fricative noise, 18
 Front cavity, 15, 113

 Glottal
 area, 271
 flow, 268
 impedance, 269
 shunt, 305
 source, 265
 waveform, 268
 Glottis, 266
 Gravity, 58, 218

 Hard-soft distinction, 171
 Harmonic, 18
 Harmonic analysis, 235
 Heat conduction losses, 33, 311
 Helmholtz resonator, 281
 Higher poles correction, 42, see Fant (1959), 50
 Highest point of the tongue, 67, 112
 Horn model of vocal tract, 30
 Hub, 25, 209

 Impedance, 16, 28
 acoustical and electrical, 16, 27
 of connecting passages, 34, 36
 very narrow passages, 267
 with superimposed flow, 34, 273
 Impulse index, 234
 Impulse response, 235
 Inductance, 27
 Inharmonic, 18
 Initial amplitude, 47
 Instantaneous amplitude, 234
 Integration time, 232
 Intensity, 229

 Interrupted, 18
 Inverse Laplace transform, 46

 Kinetic pressure, 267

 Labial transition, 199
 Laminar flow, 267
 Laplace transform, 16, 42-47
 Larynx, 102
 Larynx microphone, 128
 Lax, 223
 LEA, 100
 Lenis, 24, 224
 Lip parameter, 64, 72
 Lip Rounding, 64
 Liquids, 162
 Locus, 25, 209
 Losses (in cavity structures), 33, 135
 Loudness, 238

 Mean speech power, 233
 Mel scale, 238
 Mingograph, 243

 Nasal cavity, 140
 Nasal consonants, 142
 Nasalization, 43, 148
 Network representation of acoustic
 Neutral vowel, 56
 Noise source, 18, 272
 Numerical calculations, methods, 36

 Open aspiration, 19
 Oral cavity, 142
 Orthogonality, 59
 Oscillograms, 158, 229, 240
 Over-pressure, fricatives, 275

 Palatals, 193
 Palatalization, 171, 220
 Palatograms, 184
 Particle displacement, 34
 Particle velocity, 16, 34
 Periodic, 18
 Pharyngealization, 219
 Pharynx, 69
 Phase constant, 28
 information, 235
 of spectral component, 20
 of transfer function, 39
 Phonation, 17, 265
 Pitch voice, 17
 Place of articulation, 192, 199
 Pole, 39

- Pole-zero specification, 45, 60, 150, 194, 221
 Power, 233
 Pressure, 16
 Pressure decay, stops, 280
 Propagation constant, 28
- Q, 40
 Quality, 212, 236
 Quasi-periodic, 18
- Radiation, 44
 impedance, 35
 inductance, 35
 losses, 136, 307, 308
 reactance, 36
 resistance, 32, 35, 304
- Resistance, 27
 flow dependent, 34, 202, 268, 303
 non-linear, 34
- Resonance bandwidth, 39
 curve, 53
 frequency, 39
- Response, 15
 Retroflexion, 219
 Reynold's number, 273
 Rounding, 64
- Sagittal plane, 93
 Schneidetöne, 273
 Section (spectrum), 243
 Segmentation, 21, 185
 Sensation, 17
 Sharp/plain, 219
 Single tube, 290
 Sinus piriformis, 45, 102
 Smear, 232
 Smoothed, 25
 Sonagraph, 242
 Sound distribution, 44
 Sound level meter, 243
 Sound pressure, 16, 229
 distribution, inside the vocal tract, 125
 level, 230
 Sound spectrograph, 242
 Sound velocity, 16
 Source, 15, 17, 18, 19
 spectrum, 19, 45, 49, 265
 voiced sounds, 265
 unvoiced sounds, 201, 272
 Spatial distribution of sound pressure, 125
 Specific heat, 33
 Spectral spread, 213
 Spectrum (short time), 235
- envelope, 17, 54
 level, 56
 Standing waves, 25, 131, 307
 Stop sounds, 185, 276
 Stress, 234
 Subglottal pressure, 269
 Supraglottal pressure (CVC syllables), 276
 Sustainable, 18
 Sweep-frequency analysis, 238
- Tense/lax, 223, 279
 Tensioness, 223
 Thevinin's equivalent generator, 280
 Three-parameter vocal tract model, 71
 Timbre, 236
 Time function, 19
 T-network, 28, 36
 Tongue constriction, 72
 Transfer constant, 28
 Transfer function, 16
 Transform equations for speech production, 16, 42
 Transient (initial), 185
 Transmission line analogue, 27
 Turbulent, 18
 flow, 272
 source, 272
 Twin-tube model, 63
- Velars, 193
 Velocity of sound, c, 27
 Velocity (particle), 16
 Velum, 114
 Viscosity (of air), 32, 311
 Visible Speech, 22
 Vocal cords, 265
 Vocal tract
 models, 63
 transmission, 42
 Voice, 18
 bar, 23
 fundamental frequency, 17
 source, 265
 Voiced/voiceless, 220, 224
 Voicing (mechanism), 265
 Voltage, 16
 Volume velocity, 16
 Vowel articulation, traditional concepts, 113
 Vowels, 107
- Wavelength, 28
- X-ray analysis of the vocal tract, 93
- Zeros, spectral, 43