Description and Analysis of Contemporary Standard Russian

THE SOUND PATTERN OF RUSSIAN

A Linguistic and Acoustical Investigation

by

MORRIS HALLE

M.I.T.

Mouton & Co 1971 The Hague

THE SOUND PATTERN
OF RUSSIAN

DESCRIPTION AND ANALYSIS OF CONTEMPORARY STANDARD RUSSIAN

Edited by

ROMAN JAKOBSON

Harvard University

AND

C. H. VAN SCHOONEVELD

Leiden University

I

1971

MOUTON

THE HAGUE · PARIS

THE SOUND PATTERN OF RUSSIAN

A Linguistic and Acoustical Investigation

by

MORRIS HALLE

Massachusetts Institute of Technology

with an Excursus on

THE CONTEXTUAL VARIANTS OF THE RUSSIAN VOWELS

by

LAWRENCE G. JONES

Harvard University

Second Printing

1971

MOUTON

THE HAGUE · PARIS

First Printing 1959

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

INTRODUCTORY NOTE

This series of monographs intends to contribute to an exhaustive description and consistent analysis of contemporary standard Russian (CSR).

Exhaustiveness means establishing phonological and grammatical rules on the basis of the total lexical inventory of CSR and of all those kinds of combinations which this inventory admits. If a rule has restricted application, all the exceptions are to be cited and interpreted. No vague 'etc.' is allowed. Oversimplified levelling of the standard must be avoided and the extant variations in the over-all code of CSR are to be noted and commented on, especially the coexistence and functional difference of older and newer variants in the course of linguistic changes. Thus the dynamic aspect of CSR cannot be overlooked by its comprehensive synchronic description.

This analysis strives both for a maximum scientific control of CSR and for testing new means and goals in linguistic theory and methodology. Consistency of analysis means that any divisible unit is first to be broken down into its immediate constituents (whether concurrent or successive) and, secondly, the laws of its composition are to be determined. More generally, on any level of language the laws underlying any given set of kindred linguistic units are to be sought. Thus an inventory of phonemes requires their analysis into distinctive features and an inquiry into the laws of their stratification. Any grammatical paradigm demands the elucidation of rules which underlie it and which connect it with all other kindred paradigms. The optimal rules disclosed by the analysis are those which offer the highest degree of predictability.

The autonomous structure of any given linguistic level does not mean its separate existence in language and does not justify the linguist's inattention to the intimate and lawful interconnections of the different levels. The analysis must bridge phonemics and morphology, morphology and syntax, grammatical and lexical patterning, grammar and stylistics. The linking up of these autonomous but correlated domains has been considered by our team a particularly timely and pertinent task. The communalities and differences in the phonemic make-up of the different classes of morphemes and of the different grammatical categories efface the boundary between phonemics and morphology. It is the confrontation of vocabulary with grammar which gives us an insight into the classification of words and into the structural laws of the lexical pattern.

The problem of invariance is fundamental for the description of any linguistic level: the phonemic invariants in contextual and stylistic sound variations; the semantic invariants of any given morphological category in syntactic and lexical variations. The search for linguistic invariants is inseparably connected with the extraction of redundancy, extraction of which is a device of primary importance on all linguistic levels. The elicitation of the redundancy-free models, the inquiry into the different linguistic roles and degrees of redundancy, and the comparative study of a scale of elliptic sub-codes in relation to the explicit sub-code of CSR belong to our pivotal aims.

The referential (designative, cognitive) function of verbal communication despite its primary importance is not to be promoted as the sole function of language and as the sole concern of linguistics. All the cardinal functions of language (referential, emotive, conative, phatic, poetic, and metalingual) in their interrelationship must be subjected to close examination. In particular, metalanguage, far from being a mere artificial tool of linguists and logicians, is a substantial constituent of our verbal communication in its various aspects. Metalingual statements of native informants are to be conceived on equal footing with their other utterances. This approach gives the linguist the necessary command of such essential aspects of verbal communication as intralingual and interlingual translation and enables him to cope with semantic information as the crux of language on all its levels – lexical, grammatical, and phonemic. The frequent occurrence and great extent of homonymity and synonymity is to be carefully examined but evidently cannot be used as an argument against the primarily semantic nature of grammatical or phonemic units.

In accordance with modern scientific methodology, the search for a set of rules underlying CSR or any other given linguistic system must rigorously take into account the interaction between the observer and the object under observation. The verbal code is viewed differently by a cryptanalyst than by the actual or virtual code-users, and among the latter it is differently treated by the decoder and the encoder. The rules of verbal input and output are complementary and are not to be mechanically intermingled. The phonemic cues are prior to the grammatical unit for the listener, whereas the phonemic operations of the speaker are based on grammatical prerequisites: the two opposite linguistic views of the hierarchical order between morphophonemics and phonemics proper simply mirror two different frames of reference, namely the standpoints of the two participants in the speech event.

Not only does linguistics tend to overcome mutual alienation of its own disciplines, but also its former isolationism in respect to neighboring sciences is to be relinquished. Of course, to secure fruitful cooperation, not only the common aims but also the basic differences among the concurrent sciences and the autonomous principles of each of them are to be watchfully respected.

In order to follow the verbal message, in particular its phonological components, from source to destination, we must be able to investigate the physiological data on their production, the physical properties of their transmission and the psychological traits of their perception. Also the aural and neurological stages of these signals will perhaps disclose their differential characteristics. The interrelation of these stages, particularly of the motor and acoustic phases, is a highly instructive lesson for the linguist. While in the first monograph of our series Morris Halle tackles the questions of the Russian phonemic pattern and its acoustic correlates, in the second monograph Gunnar Fant demonstrates the analytical ties between speech production and the stage of the speech wave, and by calculations based on X-ray studies of Russian articulations achieves an exact predictability of acoustic effect from sound production.

The scrutiny of the encoding aspect of a given language brings the linguist close to the constructs of language attempted by mathematical logic. On the other hand, the decoding aspect, necessarily probabilistic, confronts the linguist with the methods and concepts of the mathematical theory of communication. In both of these cases this contact opens new outlooks both for linguistics and for mathematics. In any thorough analysis of a given language, attention must be paid to these interdisciplinary quests. That part of phonology which deals with syntactic information, phrase stress, and pauses — as well as with their emotive modifications — has much in common with musicology which treats similar material although differently organized and used for quite other purposes. The third monograph of our series, prepared in common by a musicologist and linguist (Mrs. J. E. Buning and C. H. Van Schooneveld), tries to take into account for a phonological description of Russian intonations the methods of musical analysis with unfailing attention to differences in musical and linguistic patterning of the same elements.

The utilization of linguistic material in messages with a prevalently poetic function requires a systematic description and analysis of contemporary Russian poetic language through the joint efforts of poetics and linguistics proper.

The multiple and intricate relationship of language and culture calls for a far-reaching confrontation of CSR with the data of Russian cultural anthropology. The question of the ties between language and its social and cultural background is particularly beneficial in view of the multifarious and strenuous changes which Russian life has undergone since World War I.

The lexical, grammatical and stylistic richness of CSR and the opulent material collected by Russian philologists makes a thorough analysis of this language particularly welcome and fertile. It suffices to recall the numerous and invaluable inferences which a meticulous linguistic scrutiny has been able to draw from the four volumes of the famous Ušakov dictionary in the most intricate questions of Russian derivation, inflection, and word classes.

Our series of monographs begins with three groups of works: A) phonetic and phonological studies, B) morphological and syntactic studies, C) lexical and phraseological studies. All these groups are closely interlinked. These monographs arise from the research project on "Description and Analysis of Contemporary Standard Russian" directed by R. Jakobson at the Department of Slavic Languages and Literatures of Harvard University and sponsored by the Rockefeller Foundation 1950–1958. During this period the participants in the research work were: R. Abernathy (Massachusetts Institute of Technology), H. Arntz (University of North Carolina), B. Aroutunova-Tschirwa (Harvard University), J. Beebe (Indiana University), J. E. Buning (Leiden), C. Cherry (Imperial College, University of London), C. Dawson (Syracuse University), G. Fant (Royal Institute of Technology, Stockholm), J. Ferrell (University of Michigan), P. Garvin (Georgetown University, Institute of Languages and Linguistics), M. Halle (Massachusetts Institute of Technology), G. Hüttl-Worth (University of California, Los Angeles), A. Humecky (University of Michigan), R. Jakobson (Harvard University, Massachusetts Institute of Technology), L. G. Jones (Harvard University), G. Kelemen (Harvard University Medical School), H. Klagstad (Indiana University), E. Klima (Massachusetts Institute of Technology), H. Kučera (Brown University), E. Levin (Syracuse University), H. G. Lunt (Harvard University), A. S. Macmillan (Harvard University Medical School), L. Matejka (University of Michigan), L. Micklesen (University of Washington), I. Morozova-Lynch (Wellesley College), E. Pacaluyko (Wellesley College), S. Ragozin (Syracuse University), H. Rubenstein (Bolling Air Force Base, Operational Applications Laboratory, Washington, D. C.), G. Shevelev (Columbia University), E. Stankiewicz (Indiana University), V. Tumins (Brown University), C. H. van Schooneveld (University of Leiden), D. Worth (University of California, Los Angeles), M. Zarechniak (Georgetown University, Institute of Languages and Linguistics).

Among the books to appear next, besides the three cited contributions to sound analysis, there are monographs treating the structure of Russian roots; the vowel-zero alternations; nominal declension; nominal derivation; nominal composition; conjugation; verbal derivation; participles and gerunds.

The principles and tasks outlined have led our teamwork without, however, imposing a sectarian uniformity upon any single contribution. Each author remains responsible for his theoretical postulates and conclusions. This does not interfere with our common goal — to pose new questions and to strive toward the most constructive and complete answers to questions old and new.

Roman Jakobson
C. H. van Schooneveld

TABLE OF CONTENTS

Introductory Note 5
Foreword 11
A Note on the Sign Conventions 16

PART I PHONOLOGY

Chapter I SEGMENTS AND BOUNDARIES 19
1. Introduction: A Theory of Phonology 19
2. The Phonological System of Russian 44
3. Comments on Details of the Phonological System of Russian 50

Chapter II SEQUENTIAL CONSTRAINTS 55
1. Introduction 55
2. The Morpheme Structure (MS) Rules 56
3. The Phonological (P) Rules 62

PART II ACOUSTICS

Chapter III A NON-TECHNICAL ACCOUNT OF SOME ELEMENTARY CONCEPTS AND TECHNIQUES OF ACOUSTICS 79
1. Introduction 79
2. Elementary Exposition of Some Concepts and Techniques of Acoustics 79
3. Digression on the Relation Between the Descriptive Frameworks of Acoustics, of Psycho-Acoustics and of Phonetics 88

Chapter IV A CRITICAL SURVEY OF ACOUSTICAL INVESTIGATIONS OF SPEECH SOUNDS 91

Chapter V THE ACOUSTICAL CORRELATES OF THE DISTINCTIVE FEATURES OF RUSSIAN 111
1. Subjects and Material 111

2. Measurement Procedure 115
3. Vocalic vs. Non-Vocalic, and Consonantal vs. Non-Consonantal 117
4. Compact vs. Noncompact, Diffuse vs. Nondiffuse, and Low vs. High Tonality 126
5. Strident vs. Mellow 139
6. Nasal vs. Nonnasal 140
7. Continuant vs. Interrupted 147
8. Voiced vs. Unvoiced 149
9. Sharp vs. Plain 149

Excursus CONTEXTUAL VARIANTS OF THE RUSSIAN VOWELS. by LAWRENCE G. JONES 157
1. Introduction 157
2. Phonetic and Phonemic Analysis 157
3. The Russian Vowel Variants 160
4. General Remarks. 167

Appendix I Formant Frequencies of Russian Vowels 168

Appendix II Energy Density Spectra of Stop Bursts and Continuants 184

Bibliography 199

FOREWORD

The immediate stimulus for this book was provided by the acoustical techniques that became widely available in the years following World War II. It was my original plan to collect a body of data concerning an important yet relatively unexplored aspect of language, its acoustical properties. In particular, I wanted to establish the acoustical correlates of various entities that traditionally have been used in the description of the phonic aspect of language. To isolate these different entities is in part a linguistic problem, for they are organized in strikingly different ways in different languages. To the acoustician these differences are of little interest. From his point of view, a speech event is completely characterized by the pressure changes in the air at some point in front of the speaker, for a record of such pressure changes suffices to reconstitute the observed acoustical event with any desired degree of fidelity. To the student of language, on the contrary, the above differences are of paramount importance. Unless he can discover in the acoustical signal, reflexes of the complicated hierarchical organization of language, acoustical studies have little practical importance for linguistics.

In trying to apply the traditional methods of descriptive linguistics to the description of Russian I became aware of a number of difficulties. In too many instances traditional methods seemed to impose labored or counter-intuitive solutions, or to require an organization of the data that was not particularly enlightening.[1] Upon closer study it appeared that most of the difficulties stemmed from the methodological requirement, accepted almost universally by modern linguists, that the theoretical entities utilized in linguistic descriptions – i.e., the phonemes, junctures, etc. – be derivable from the raw data (from the corpus of utterances examined by the linguist) by following rigorous, deductive procedures, which at every step should be open to public scrutiny.[2] In the

[1] See, e.g., the long lists of phoneme sequences given in Chapter III of my unpublished Ph.D. dissertation *The Russian Consonants: A Phonemic and Acoustical Study* (Harvard, 1955).
[2] Perhaps the clearest statement of this can be found in the following passage from an article by Hockett: "The analytical process thus parallels what goes on in the nervous system of a language learner, particularly perhaps that of a child learning his first language... The essential difference between the process in the child and the procedure of the linguist is this: the linguist has to make his analysis overtly, in communicable form, in the shape of a set of statements which can be understood by

area of phonology this requirement has been interpreted even more narrowly as limiting the admissible analytical procedures to a specific set agreed upon in advance. Many linguists reject, therefore, phonological analyses in which different phonemes have identical (overlapping) allophones, because – as Bloch formulated it – "if we start from the facts of pronunciation as we meet them, there is never any clue in the utterance itself to tell us which kind of X (phoneme – M.H.) we are dealing with."[3] In other words, phonemes – as well as all other theoretical entities of phonology – must be discoverable from clues which are present in the utterance itself. This, however, is a consequence of the fundamental methodological principle that the theoretical entities are defined by means of the analytical procedures required for their discovery.

In contrast I take the position that the methods by which a scientific description is discovered are not of essential concern to the science in question. They belong to the subject matter of the philosophy of science rather than to that of any particular science. Whether Newton discovered the concept of gravitation as a consequence of his unfortunate experience with a falling apple or as a result of elaborate mathematical operations on some body of data is of only marginal interest to the science of physics. Similarly, linguistics is not so much interested in the particular sequence of steps that led (or could have led) Sir William Jones to the discovery of the genetic relationship between Sanskrit and various European languages than in how the introduction of this novel conception affects the description of the facts and in whether or not the new description is more general and simpler than were earlier ones.

I do not believe that it is desirable to require a rigorous deductive procedure which would lead from the physical data, from the speech events that constitute the linguist's corpus, directly to the theoretical entities – e.g., the phonemes and boundaries – in terms of which linguistic descriptions are normally framed. I feel that in linguistics, as in any other science, the main criteria that can reasonably be imposed on theoretical entities are that they make possible general and simple descriptions of all the facts. The theoretical entities must, furthermore, be related to the physical data in a statable – though not necessarily simple – fashion.[4]

I have assumed that an adequate description of a language can take the form of a set of rules – analogous perhaps to a program of an electronic computing machine – which when provided with further special instructions, could in principle produce all and only well-formed (grammatical) utterances in the language in question. This set of rules, which we shall call the grammar of the language and of which phonology forms a

---

any properly trained person..." "A Note on 'Structure'," *IJAL, 14*, 269–271 (1948). Though Hockett no longer holds this positon (cf. his *Manual of Phonology* (Bloomington, 1955), p. 147), it is probably held by most linguists if, e.g., the interventions in the discussion of phonological problems at the last International Congress of Linguists can be taken as representative; see *Proceedings of the Eighth International Congress of Linguists* (Oslo, 1957).

[3] B. Bloch, "Phonemic Overlapping," in M. Joos, ed., *Readings in Linguistics* (Washington, 1957), p. 96.

[4] For a more detailed discussion of the above position, see N. Chomsky, *Syntactic Structures* ('s-Gravenhage, 1957), chapter 6.

separate chapter, embodies what one must know in order to communicate in the given language: it contains an essential part of what the child learns from his parents; or the language learner, from his teacher. As such it is neither a description of the language from the viewpoint of the speaker nor from that of the listener; it is rather an essential element in the simplest and most general account of the linguistic behavior of both speakers and listeners.

It is obvious that not every acoustical signal emanating from man's vocal organs is a linguistic utterance. Coughs, whistles, affective vocalizations must evidently be excluded. Similarly, we must exclude from our account utterances which would be regarded by native speakers of the language as improperly formed, ungrammatical; in short, as mistakes. We do this by requiring that every grammar produce only well-formed, i.e., grammatical utterances in the language in question. The notion of grammaticalness, therefore, emerges as a consequence of our description rather than being given in advance. A scientific account of a language can thus be said to be equivalent to an extended definition of grammaticalness or to a specification of the grammatical utterances in the language.

Like all other parts of the grammar a phonological description is formulated here as a set of deductive rules. It is distinguished from other parts of the grammar in that it utilizes theoretical entities of a special kind; e.g., segments, boundaries, etc. It is, therefore, necessary to state abstract conditions which define the theoretical entities of phonology and restrict their mode of operation so as to allow for simple and general descriptions of the relevant facts. These conditions, which are discussed at length in Chapter I, are in effect the theory of phonology that underlies the present work. The theory is further extended to include a possible measure of simplicity for phonological descriptions.

While theoretical issues have been a major concern in writing this book, they constitute but a fraction of its contents. The body of the book consists of a detailed description of the phonological system of contemporary Russian and a report on the results of acoustical measurements performed on various linguistic entities. The former illustrates the application of our phonological theory to concrete data, providing thereby an opportunity to judge the theory's effectiveness. The latter establishes the link between the theoretical entities of linguistic descriptions and the real world of sound, of which human speech is undeniably a part. It is this link which raises the theoretical entities of linguistics from the status of convenient fictions to be invented at will to that of terms in a scientific theory.

In a number of important respects in which it departs from procedures that enjoy almost universal acceptance among contemporary phonologists, the present description follows methods that are characteristic of the work of Edward Sapir. Thus, like the descriptions of Sapir, the present work does not recognize the need for a "phonemic" transcription in addition to a "morphophonemic" transcription. The present description also resembles those of Sapir in that the relation between the phonological representation and the phonetic facts is embodied in a set of rules which must be applied in a particular order, instead of being given – as in most contemporary work – by a list of allophones; i.e., by a set of one-step rules that can be applied in any order.[5] I believe that these parallels in the practice of phonology reflect closely related conceptions of the theory of phonology. It is for this reason that I have chosen as the title of my book a paraphrase of the title of one of Sapir's most important essays.

While working on this book I have been in close contact with many scientists and scholars, and it is a very pleasant duty to acknowledge my debt to them.

Foremost among these is my teacher Roman Jakobson, to whom I am indebted for much of my professional development. I hope that in this book he will find some tangible evidence that the many hours he has generously devoted to instructing me have not been altogether wasted.

To Noam Chomsky I am indebted for my conception of linguistic theory. He has also helped me with incisive discussions of detailed problems and has shown me the way out of more than one blind alley into which I was ready to plunge.

G. H. Matthews has read various versions of the book and has aided me by communicating to me results of parallel investigations that he has been conducting on American Indian languages.

Most of the acoustical measurements were carried out jointly with George W. Hughes at the Research Laboratory of Electronics, M.I.T. The design, construction and maintenance of the apparatus, much of which had to be developed especially for this investigation, was almost entirely in his hands. During the many years that we have worked together he has been a pleasant laboratory companion, willing to make the effort to translate my often poorly formulated ideas into physically testable hypotheses.

Gunnar Fant of the Royal Institute of Technology, Stockholm, Sweden, has introduced me to some of the mysteries of acoustics, and has given me much valuable advice.

For instruction and help in matters acoustical I am also deeply indebted to K. N. Stevens of M.I.T. He has always been available for consultation and has shared freely with me the results of his current investigations.

I am grateful for important advice and practical assistance to W. N. Locke, to Horace G. Lunt, to Eric Lenneberg, to Calvert Watkins, to R. Abernathy, to A. S. House, and to Robert Lees.

The illustrations are the work of my wife, Rosamond T. Halle, and the arduous task of typing a far from perfect manuscript was undertaken by Lorna Slocombe. I am grateful to them for the promptness, the competence and above all, the good humor with which they have accomplished their work.

The financial support for this work has come in part from the Signal Corps; the Office of Scientific Research, Air Research and Development Command; the Office of Naval Research; the Carnegie Foundation's grant to M.I.T. for the study of scientific

[5] Cf., C. F. Hockett, "Two Models of Grammatical Description," *Word*, *10*, 210–231 (1954), where these two methods, termed "item and process" and "item and arrangement", respectively, are described, and Sapir is singled out as the primary exponent of the former method.

aids to learning; the Rockefeller Foundation's grant to Harvard University for a description of contemporary standard Russian; and from the National Science Foundation's grants to M.I.T. for the study of the structure of language.

During the past eight years it has been my great and good fortune to be associated with the Research Laboratory of Electronics, M.I.T. This unique research organization has been an ideal environment in which to carry on investigations that overlap a number of traditional boundaries between disciplines. It is to the staff, past and present, of this laboratory that I dedicate this book as an expression of my gratitude for the numerous tangible and intangible benefits that I have derived from the privilege of working among them.

July 1959 *Morris Halle*

NOTE ON THE SIGN CONVENTIONS

The phonetic alphabet used in the book is a special modification of that of the International Phonetic Association. The following deviations from and modifications of IPA practices were made: As is common in Slavistic works, the palatal consonants are symbolized by *č š* and *ž* instead of the IPA *tʃ ʃ* and *ʒ*, and the dental affricate, by *c* instead of the IPA *ts*. Palatalization is indicated by a comma following the symbol for the unpalatalized sound; e.g., *t,* stands for the palatalized voiceless dental stop. Accent is indicated by an apostrophe preceding the symbol for the unaccented vowel; e.g., *'u* represents the accented close rounded back vowel. (The reader's particular attention is drawn to the last convention, for it may be confused with the sign for palatalization.)

Different types of representation are distinguished by being enclosed in different kinds of parentheses. Square brackets are employed for phonetic representations. The use of diagonals and braces is explained in fn. 13 of Chapter I. The significance of the asterisk is discussed at the end of sec. 1.53 of Chapter I, and that of italicized phonetic symbols, in fn. 23 of Chapter I.

PART I

PHONOLOGY

CHAPTER I

SEGMENTS AND BOUNDARIES

1. *Introduction: A Theory of Phonology*

The essence of the phonological theory underlying the present description of the sound pattern of Russian is contained in six formal conditions which phonological descriptions must satisfy. In the exposition I shall exhibit in some detail the consequences of these conditions – and hence of the theory – for the representation of specific phonetic facts, and compare these consequences with those following from requirements inherent in alternative theories of phonology. The proposed theory will be justified by the insightfulness, generality, and simplicity of these, its "practical" consequences.

1.1 Condition (1): In phonology, speech events are represented as sequences of entities of two kinds: *segments*, to which specific phonetic (articulatory as well as acoustical) properties are assigned, and *boundaries*, which are characterized solely by their effects on the former.[1]

1.2 Condition (2): The phonetic properties in terms of which segments are characterized belong to a specific, narrowly restricted set of such properties called the *distinctive features*. All distinctive features are binary.

In accepting Condition (2) one commits oneself to characterizing all segments in all languages in terms of a restricted check list of attributes like "nasality, voicing, palatalization, etc.", with regard to which the only relevant question is "does the segment possess the particular attribute?" It follows, therefore, that differences between segments can be expressed only as differences in their feature composition and that consequently segments (even in different languages) can differ from each other only in a restricted number of ways.

The view that all human languages can be characterized by a restricted list of phonetic properties has been accepted by most linguists and phoneticians. It is implicit in the numerous attempts at a general phonetics from Bell's *Visible Speech* of

[1] Boundaries are analogous to what some linguists have termed "junctures". Since the latter term has recently been used in a very special sense, the more neutral "boundary" has been adopted here. I find particularly unacceptable the identification of "juncture" with a slowing down in "tempo" – cf., R. P. Stockwell, J. D. Bowen, and I. Silva-Fuenzalida, "Spanish Juncture and Intonation," *Language*, *32*, 643 (1956).

1867 to Heffner's *General Phonetics* of 1949. The view has, however, been challenged by some scholars, who felt – as one of them tells us – "that languages could differ from each other without limit and in unpredictable ways".[2]

Condition (2) and the view just quoted are, therefore, mutually contradictory assertions about the nature of human language which must be subjected to empirical tests. If an examination of a large body of languages revealed that the number of different phonetic features needed increases with the number of languages examined, then Condition (2) would have to be rejected. If such an examination, however, revealed that the number of different phonetic features increases little or not at all beyond some small value as more and more languages are included in the examination, then Condition (2) should be accepted.

Although languages have been found which possess phonetic features unknown in the Western languages, the number of such features must not be exaggerated. In examining the phonetic frameworks tested on a large number of languages – e.g., that of Trubetzkoy's *Grundzüge*, Pike's *Phonetics*, or the modified IPA system successfully utilized by British students of African and Oriental languages – one cannot fail to be impressed by the small number (on the order of twenty or less) of features involved. Since the languages that have been described constitute a representative sample of all human languages, it is to be expected that the number of relevant phonetic properties will not radically increase as more languages are subjected to scientific study. There seems, therefore, no good reason for rejecting Condition (2) on these grounds.

Condition (2), however, entails an even heavier restriction. It requires that the segments be characterized in terms of a specific list of binary attributes: the distinctive features.[3] Systematic reviews of the available evidence from a great variety of languages have shown that the binary distinctive feature framework is adequate to the task.[4] No examples have been adduced by various critics that would seriously impair the validity of the binary scheme.[5] On the contrary, the imposition of the binary structure on all features has supplied a satisfactory explanation for a number of 'puzzling' phonetic changes[6] and made possible the formulation of an evaluation procedure for phonological descriptions.[7]

1.3 The segments and boundaries are theoretical constructs and must, therefore, be appropriately related to the observable data, i.e., to the actual speech events. The weakest condition that can be imposed and that is, in fact, accepted by all, is

[2] M. Joos, ed., *Readings in Linguistics* (Washington, 1957), p. 96.

[3] For a complete listing of the distinctive features, see R. Jakobson, C. G. M. Fant, and M. Halle, *Preliminaries to Speech Analysis* = *M.I.T. Acoustics Laboratory Technical Report No. 13* (1952), and R. Jakobson and M. Halle, *Fundamentals of Language* ('s-Gravenhage, 1956).

[4] Cf., the sections entitled "Occurrence" in *Preliminaries to Speech Analysis*, and M. Halle, "In Defense of the Number Two," *Studies Presented to J. Whatmough* ('s-Gravenhage, 1957), pp. 65–72.

[5] I have tried to answer certain theoretical objections of Martinet in "In Defense of the Number Two".

[6] E.g., the shift of velars to labials in Rumanian; cf. R. Jakobson, "Observations sur le classement phonologique des consonnes," *Proceedings of the Third International Congress of Phonetic Sciences* (Ghent, 1938), p. 37.

[7] See below, secs. 1.5–1.551 passim.

Condition (3): A phonological description must provide a method for inferring (deriving) from every phonological representation the utterance symbolized, without recourse to information not contained in the phonological representation.

In other words, it must be possible to read phonological representations regardless of whether or not their meaning, grammatical structure, etc., is known to the reader. It is obvious that this cannot be accomplished unless all distinctively different utterances are represented by different symbol sequences. It is, however, not necessary that the converse also be true, for instructions can be given which will result in several nonidentical symbol sequences being pronounced in the same way. E.g., the symbol sequences {m'ok bi} and {m'og bi} would be pronounced identically, if a rule were stated requiring unvoiced consonants to be voiced in position before voiced consonants. But in this case it will be impossible to determine from the utterance alone, which of the two (or more) symbol sequences is the proper representation of the particular utterance. Thus in the above example, the listener upon hearing the utterance [m'ogbɨ], will not be able to choose one or the other of the two representations unless he has access to meaning or other information which is not present in the signal. It follows, therefore, that only a single sequence of symbols can be allowed to represent a given sequence of sounds, if phonological descriptions are to satisfy also

Condition (3a): A phonological description must include instructions for inferring (deriving) the proper phonological representation of any speech event, without recourse to information not contained in the physical signal.[8]

1.31 Condition (3a) can be met most simply by establishing a set of symbols such that there is one sound per symbol and one symbol per sound. If this set is exhaustive, in the sense that it contains a symbol for every sound encountered, then anyone acquainted with the phonetic value of the symbols can not only read off correctly any symbol sequence, but can also write down the proper symbol sequence corresponding to the particular utterance. This is the way in which the phoneticians of the turn of the century sought to satisfy Condition (3a), as is evidenced by the famous slogan of the Association Internationale de Phonétique, "same sound, same symbol". As is well known, all attempts to implement this slogan failed because they invariably resulted in an apparently limitless proliferation of symbols, for strictly speaking, no two sounds are ever the same. The obvious escape from this difficulty seemed to be the imposition of some sort of limitation on the number of symbols to be employed.

1.32 The specific suggestion advanced was

Condition (3a-1): Only utterances which are different are to be represented by

[8] This requirement has played a particularly important role in the development of American linguistics. "For a notation to be phonemic we require a bi-unique, one-one relation rather than a many-one relation [between representation and utterance – M.H.]" C. F. Hockett, Review of A. Martinet's *Phonology as Functional Phonetics*, *Language*, *27*, 340 (1951).

different sequences of symbols.[9] The number of different symbols employed in all representations must be the minimum compatible with this objective.[10]

In short, the slogan "same sound, same symbol" was replaced by the requirement "same utterance, same representation", and a limitation was imposed on the number of symbols to be utilized in the representations. This limitation gave rise to a number of difficulties, however. E.g., in English [h] and [ŋ] do not occur in identical environments. Condition (3a-1) would require that these be considered positional variants of the same phoneme, but this runs strongly against our natural inclination. Even more perplexing is the fact that it is always possible to represent any number (of events, utterances, individuals) by a binary number; hence, Condition (3a-1) can always be satisfied in a trivial way by adopting an alphabet consisting of two symbols only. This, however, can be done quite regardless of the phonetic facts and would, therefore, lead to the absurd conclusion that all languages have the same number of phonemes, i.e., two.[11]

To overcome these difficulties it has been proposed to require that positional variants of a particular phoneme be "phonetically similar". Unfortunately this only serves to push the difficulty one step further, *viz.*, to the question of what is meant by "phonetically similar" – which is just another form of the unanswered question of what is meant by saying that two sounds are the same.

1.33 Consider next the consequences of Condition (3a) on the phonological representation of the following facts.[12] In Russian, voicing is distinctive for all obstruents except /c/, /č/ and /x/, which do not possess voiced cognates. These three obstruents are voiceless unless followed by a voiced obstruent, in which case they are voiced. At the end of the word, however, this is true of all Russian obstruents: they are voiceless, unless the following word begins with a voiced obstruent, in which case they are voiced. E.g., [m'ok l,i] "was (he) getting wet?", but [m'og bɨ] "were (he) getting wet"; [ž'eč l,i] "should one burn?", but [ž'eǯ bɨ] "were one to burn".

In a phonological representation which satisfies both Condition (3) and (3a), the quoted utterances would be symbolized as follows: /m'ok l,i/, /m'og bi/, /ž'eč l,i/,

[9] The question of what criterion is to be used to establish whether or not two utterances are different has been the subject of much argument. I accept the position outlined by N. Chomsky, "Semantic Considerations in Grammar," *Georgetown University Monographs on Languages and Linguistics*, No. 8 (Washington, 1955), pp. 141–158. This view is now held also by the leading Soviet phonetician R. I. Avanesov; cf. *Fonetika sovremennogo russkogo literaturnogo jazyka* (Moscow, 1956), pp. 14–15.

[10] This is in essence the condition advocated by Z. S. Harris, *Methods of Structural Linguistics* (Chicago, 1951), p. 43.

[11] M. Halle, "Why and How Do We Study the Sounds of Speech," *Georgetown University Monograph on Languages and Linguistics*, No. 7 (Washington, 1954), pp. 73–83.

[12] The facts are somewhat simplified here. A complete statement is given in Chapter II, sec. 3. – Analogous examples can be cited from many languages. An especially interesting example is discussed by G. H. Matthews, "A Phonemic Analysis of a Dakota Dialect," *International Journal of American Linguistics, 21*, 56–59 (1955), who shows that the labial and dental nasal consonants are automatic alternants of the cognate stops as well as of /m/ and /n/, while the velar nasal is an alternant of the velar stop exclusively. See R. B. Lees's review of N. Chomsky, *Syntactic Structures*, *Language, 33*, 389–390 (1957), for a Turkish example.

/ž'eč bi/.[13] Moreover, a rule would be required stating that obstruents lacking voiced cognates – i.e., /c/ /č/ and /x/ – are voiced in position before voiced obstruents. Since this, however, is true of all obstruents, the net effect of the attempt to meet both Condition (3) and (3a) would be a splitting up of the obstruents into two classes and the addition of a special rule. If Condition (3a) is dropped, the four utterances would be symbolized as follows: {m'ok l,i} {m'ok bi} {ž'eč l,i} {ž'eč bi},[13] and the above rule could be generalized to cover all obstruents, instead of only {č} {c} and {x}. It is evident that Condition (3a) involves a significant increase in the complexity of the representation.

Traditionally linguistic descriptions have contained both representations satisfying Condition (3) alone, and representations satisfying Conditions (3) and (3a). The former are usually called 'morphophonemic' to distinguish them from the latter, which are called 'phonemic'.[14] Morphophonemic representations cannot be dispensed with in a linguistic description since they are the means for accounting for ambiguities due to homophony. E.g., the fact that English [tʰ'æks] ('tacks' and 'tax') is ambiguous is normally explained by saying that these "phonemically identical" utterances differ morphophonemically.

Note, however, that in the Russian example discussed the morphophonemic representation and the rule concerning the distribution of voicing suffice to account for all observed facts. Phonemic representations, therefore, constitute an additional level of representation made necessary only by the attempt to satisfy Condition (3a). If Condition (3a) can be dispensed with, then there is also no need for the 'phonemic' representation.

1.34 Condition (3a) is concerned with procedures that are essentially analytical. Analytical procedures of this kind are well known in all sciences. Qualitative and quantitative chemistry, electrical circuit analysis, botanical and zoological taxonomy, medical diagnosis are examples of disciplines concerned with discovering the appropriate theoretical representations (i.e., chemical formula, configuration of circuit elements, classification within the taxonomical framework, name of disease, respectively) of different complexes of observable data. The theoretical constructs which make up the representations discovered by the different types of analysis are, however, postulated within the framework of the individual sciences without regard for the procedures whereby they can be discovered in the data. Theoretical constructs are never introduced because of considerations that have to do with analytic procedures. Thus, for instance, it is inconceivable that chemistry would establish substances that can be identified by visual inspection as a category distinct from substances that

[13] Representations satisfying both Condition (3) and (3a) will be enclosed in diagonals (/), in order to distinguish them from representations satisfying only Condition (3), which will be enclosed in braces ({}).

[14] "The native listener may be said to perceive – to somehow exploit for message-understanding ends – items in what he hears. Insofar as this process does not depend on understanding, the items are phonemic; insofar as items cannot be perceived without understanding, morphophonemics at least (perhaps more) is involved." Joos, *op. cit.*, p. 92.

require more elaborate techniques for their identification. Yet this is precisely the import of Condition (3a), for it sets up a distinction between phonemes and morphophonemes for the sole reason that the former can be identified on the basis of acoustic information alone, whereas the latter require other information as well.

So important a deviation from standard scientific practice can only be justified if it were shown that phonology differs from other sciences in such a way as to warrant the departure. This, however, has never been demonstrated. Quite to the contrary, it has been common to stress the essential similarity between the problems of phonology and those of other sciences. The conclusion, therefore, imposes itself that Condition (3a) is an unwarranted complication which has no place in a scientific description of language.

The abolition of Condition (3a) is not as much at variance with traditional practice as might at first appear. It is hardly an accident that in the phonological descriptions of E. Sapir,[15] and to some extent also in those of L. Bloomfield,[16] Condition (3a) played no role.

1.4 Condition (4): The phonological description must be appropriately integrated into the grammar of the language. Particularly, in selecting phonological representations of individual morphemes, these must be chosen so as to yield simple statements of all grammatical operations – like inflection and derivation – in which they may be involved.

In the present work a grammar will be viewed as a device for specifying all sentences of a language.[17] It may, therefore, be thought of as an extended definition of the term "sentence in language L". In structure a grammar resembles a postulational system from which theorems are derived by the application of definite rules of inference. Each sentence in the language can be considered a theorem of the postulational system constituting the grammar.

The process of specification begins with the symbol "Sentence", since it is this term that is explicated by the grammar. In the process of specification this symbol is translated into various representations which are connected with one another by appropriate rules; i.e. at each stage in the process of specification the sentence is represented by a particular arrangement (not necessarily one-dimensional sequences) of symbols which is the consequence of applying the rules of the grammar. In order to delimit the individual symbols from each other and to join them to their neighbors, each symbol is preceded and followed by a special marker – & (the ampersand). It will be shown later that these markers play an important role in the phonological

[15] See his "Sound Patterns in Language," and "The Psychological Reality of Phonemes," in D. G. Mandelbaum, ed., *Selected Writings of Edward Sapir* (Berkeley and Los Angeles, 1949), pp. 46–61.

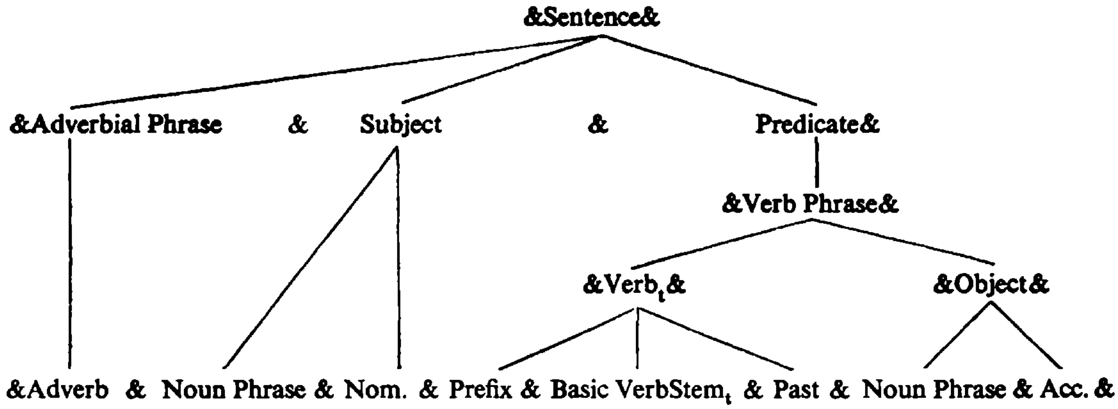
[16] Significant with regard to Bloomfield's practice is the following comment by M. Joos: "When we look back at Bloomfield's work we are disturbed at this and that, but more than anything else *Bloomfield's confusion between phonemes and morphophonemes* disturbs us." (Italics supplied.) *Op. cit.*, p. 92.

[17] The concept of grammar is taken from the recent work of N. Chomsky. For more details, see his *Syntactic Structures* ('s-Gravenhage, 1957), and R. B. Lees's review in *Language*, *33*, 375–407 (1957).

representation of the sentence, for some of them are ultimately translated into phonological boundaries. The last step in the process of specification of the sentence is the translation of the abstract representation into sound.

The rules of translation which make up the grammar can all be subsumed under the formula "replace x by y under condition z" They differ, however, in the type of representation that results from their application. The different types of representation are the consequence of restrictions placed on the values over which the variables x, y and z may range. A set of rules yielding representations of a particular type is called a *linguistic level.*

The purpose of the rules of the highest level, the so-called *Phrase Structure* level, is to eventuate in tree-like representations which embody the phrase structure of the sentence. Such a tree is illustrated by the following partial phrase structure of a Russian sentence:



The rules yielding this tree are:

Replace	&Sentence&	by	&Adverbial Phrase&Subject&Predicate&	(1)
	&Adverbial Phrase&	by	&Adverb&	(2)
	&Subject&	by	&Noun Phrase&Nom.&	(3)
	&Predicate&	by	&Verb Phrase&	(4)
	&Verb Phrase&	by	&Verb$_t$&Object&	(5)
	&Verb$_t$&	by	&Prefix&Basic Verb Stem$_t$&Past&	(6)
	&Object&	by	&Noun Phrase & Acc. &	(7)

The phrase structure of a sentence is completely specified when there are no more symbols that can be replaced by any of the rules listed. (E.g., &Noun Phrase& cannot be replaced by any other symbol in the above rules.) These "irreplaceable" symbols are called *terminal* symbols, and a sequence of such symbols is a *terminal string*. Since in an actual grammar, however, there are many more rules than those given here, the "irreplaceable" symbols of the illustrative example are not in fact terminal symbols of the Phrase Structure rules of Russian.

The different branching points in the tree correspond to different Immediate Constituents of the sentence.[18] The tree represents, therefore, the Immediate Constituent structure of the sentence, and the Phrase Structure rules are the formal analog of Immediate Constituent analysis (parsing). In order that trees of this type result from the rules, it is necessary to restrict the rules so that not more than one symbol can be replaced by a single rule. This restriction also insures that there will be a Phrase Structure tree for every terminal string. It also makes it possible to trace a unique path from the initial symbol &Sentence& to any other symbol (Immediate Constituent) in the tree. This path is called the *derivational history* of the symbol.

The trees are then subjected to the rules of the *transformational* level. In the transformational level more than one symbol may be replaced in a single rule, which opens the way to effecting changes in the representation that are beyond the power of the Phrase Structure rules. E.g., symbols may be re-ordered in the sequence or may be eliminated. Moreover, in the transformational rules account is taken of the derivational history of individual symbols. It is, therefore, possible, e.g., to give a different rule for a &Noun Phrase& derived from a &Subject& than for a &Noun Phrase& derived from an &Object&. Because of this reference to the derivational history of individual symbols, transformational rules are said to operate on Phrase Structure trees, rather than on terminal strings.

The final set of rules, the *phonological* rules, takes the transformed terminal strings, which consist entirely of special kinds of segments and of boundaries, and completes the assignment of phonetic features to these symbols. Unlike a Phrase Structure rule, a single phonological rule can replace more than one symbol. Phonological rules, however, do not take account of the derivational history of the symbols to which they apply.

1.41 Up to this point in the discussion, sentences have been represented exclusively in terms of morpheme class symbols, like &Subject&, &Adverb&, &Nom.&, etc. It is evident that at some stage in the course of the specification these morpheme class symbols must be replaced by specific morphemes; e.g. &Adverb& must be replaced by a specific Russian adverb. The replacement can be included within the Phrase Structure level, since it can be done by means of rules like

"replace &Adverb& by A
B
C, etc."

where A, B, C, stand for specific Russian adverbs like *tam* "there", *bystro* "rapidly", *včera* "yesterday", etc. Rules of this sort comprise the *dictionary* of the language.

The selection of some morphemes is governed by the context in which they occur. For instance, in Russian there is an intimate connection between the phonological

[18] For a discussion of Immediate Constituent analysis, see R. S. Wells, "Immediate Constituents," *Language*, 23, 81–117 (1947).

composition of the morpheme replacing &Basic verb stem& and the choice of the present tense suffix. It is possible, in principle, to argue either that the latter determines the former, or that the former determines the latter. In all cases that I have studied, however, fairly elementary considerations of economy require that the selection of the suffix be made dependent on that of the stem, and not vice versa.[19]

Considerations of this nature have always been operative in linguistic descriptions and have been instrumental in establishing the distinction between *lexical* and *grammatical* morphemes.[20] It is not possible here to go into the question of which morpheme classes are lexical and which are grammatical. For present purposes it suffices to have established that there is a need for this distinction and that lexical morphemes must be introduced into the representation before grammatical morphemes.

1.42 It is now necessary to consider the manner in which individual grammatical morphemes are introduced into the representation. The Phrase Structure rules, which have served well up to this point, lead to difficulties in cases like the following. A number of Russian &Noun& are homophonous with &Adjective&; e.g. {s,'in,} as &Noun& meaning "blueness" and {s,'in,} as &Adjective& meaning "blue". Furthermore, both &Noun& and &Adjective& occur before the same grammatical morpheme classes, e.g. before &Pl.&Nom.&. Consequently by the Phrase Structure rules, both &Adjective&Pl.&Nom.& and &Noun&Pl.&Nom.&[21] are converted into &{s,'in,}& &Pl.&Nom.&. This presents a real difficulty: &Pl.&Nom.& is translated into a different suffix after &Noun& than after &Adjective&. Yet by the conventions of the Phrase Structure rules it is not possible to treat a given sequence of symbols in different ways, depending on their derivational history. It is consequently impossible to obtain from &{s,'in,}&Pl.&Nom.& two representations, i.e. {s,'in, -i} in the case of the &Noun&, and {s,'in, -iji} in the case of the &Adjective&.

It is possible to remedy this by establishing additional Phrase Structure rules of the following type:

Replace &Adjective&Pl.&Nom.& by &Adjective&Pl.&Nom. $_{\text{adj.}}$&
Replace &Noun&Pl.&Nom.& by &Noun&Pl.&Nom. $_{\text{noun}}$&

[19] This can be seen in the following example. Consider a language in which a) the past tense suffix is [t] or [d] depending on whether or not the verbal stem ends in a voiceless consonant, and b) the present stem suffix is zero. If the past stem suffix is selected first, it is not necessary to specify the feature of voicing in the last consonant of the stem, since it can be inferred from the suffix which has been chosen. This economy, however, cannot be realized in all representations of the stem, since before the present tense (zero) suffix, it is not possible to infer whether or not the last segment of the stem is voiceless. On the other hand, if the stem is selected first, the feature of voicing need not be specified in the suffix since it can be inferred from the last segment of the stem. Since, moreover, the suffix never appears without a stem, there is no need for a multiple representation of the suffix, as would be necessary in the case of the stem. It follows, therefore, that it is more economical to select the stem before the suffix, which has, of course, been the traditional procedure.

[20] Cf., N. S. Troubetzkoy, "Le rapport entre le déterminé, le déterminant et le défini," *Mélanges de linguistique offerts à Charles Bally* (Geneva, 1939), pp. 75–82.

[21] Although Ušakov's dictionary does not admit a plural of noun *sin'*, plurals are attested in poetry.

These rules eliminate the ambiguity noted within the constraints of the Phrase Structure level. The price for this move, however, is high: it is an increase in the number of grammatical morpheme classes. Instead of dealing with a single grammatical morpheme class &Nom.&, it will now be necessary to deal with a large number of these, because not only &Noun& and &Adjective& present instances of homophony, but many other classes as well.

If the fact that in many languages more than one suffix corresponds to a single grammatical morpheme class causes difficulties for the Phrase Structure rules, a different type of difficulty is posed by the widespread phenomenon of "syncretism".[22] "Syncretism" in linguistic terminology refers to the situation where more than one grammatical category is expressed by a single mark; e.g., the case endings of Russian nouns normally represent number or gender as well as case. Rules of the Phrase Structure level, however, expressly require that a single rule replace no more than one symbol. Hence within Phrase Structure it is impossible to give a rule like "replace &Pl.&Nom.& by &{i}&" where two symbols – i.e., &Pl.& and &Nom.& – are replaced in one step. In sum, the morphological process of inflection cannot be incorporated into the Phrase Structure level.

The natural solution of these two difficulties is to make morphology – i.e. the part of the grammar dealing with the replacement of grammatical morpheme class symbols by individual grammatical morphemes – a part of the transformational level, where the two restrictions do not apply. This solution is particularly attractive, since it parallels closely the traditional manner of handling morphological processes, where individuals are treated differently depending upon the morpheme class from which they derive, and where the replacement of more than one symbol in a single rule is a standard procedure.

1.5 As was noted in section 1.41 above the Phrase Structure level must contain rules of the type

Replace &Adverb&	by *tam*	(8a)
Replace &Adverb&	by *včera*	(8b)
Replace &Adverb&	by *tak*	(8c) etc.,

i.e., lists of morphemes. In a scientific description of a language it is, however, not possible to be satisfied with giving simple lists of all existing morphemes. Just as a syntax of a language is more than an exhaustive inventory of a corpus of sentences, so a phonological description of a language cannot be identical with a list of morphemes. It must contain a statement of the structural principles of which all actual morphemes are special instances.

[22] These two features are particularly characteristic of the Indo-European languages; cf., the following remarks by Meillet: "Un trait caractéristique de l'indo-européen est que les catégories grammaticales n'y ont pas chacune une expression isolée; il n'y a pas, comme en turc, une marque de pluriel à laquelle s'ajouterait la marque du cas (et du genre) pour les noms, de la personne et des autres catégories pour les verbes ... en même temps les éléments grammaticaux expriment à la fois plusieurs catégories...," *Introduction à l'étude comparative des langues indo-européennes* (Paris, 1937), p. 191.

In generating a specific sentence it will be necessary to select from the alternatives available in the language – i.e., from lists like rules (8a) to (8c) – the particular morphemes which are to appear in the sentence. The process of selection must be under the control of extra-grammatical factors. The instructions to select a particular morpheme in the list must be provided to the grammar from the outside (presumably by the speaker). These instructions might be given in the form "select rule (8a)" which the grammar would then interpret as an instruction to replace the symbol &Adverb& by *tam*.

Instead of framing the instructions in an arbitrary numerical code which contains no information about the phonic structure of the morphemes involved, it is possible – and also more consonant with the aims of a linguistic description – to utilize for this purpose the distinctive feature representation of the morphemes directly. Thus, for instance, instead of instructing the grammar to "select rule (8a)", the grammar will be instructed to "replace &Adverb& by a segment sequence in which the first segment contains the features: nonvocalic, consonantal, noncompact, high tonality, mellow, nonnasal, etc.; the second segment, the features: vocalic, nonconsonantal, nondiffuse, compact, etc.; and the third segment, the features: nonvocalic, consonantal, noncompact, low tonality, mellow, nasal, etc." Instructions of this type can conveniently be represented by matrices in which each column stands for a particular segment, and each row stands for a particular feature. Since features are binary, the symbol plus (+) will be used to indicate that the particular segment possesses the feature in question, and the symbol minus (—), that it does not possess the feature. Such a representation is illustrated in Tab. I-1.

Since the purpose of the instructions is to select one morpheme from a list, an important role in the instructions will be played by those features and feature complexes which serve to distinguish one morpheme from another. Features and feature complexes which fulfil this function will be called *phonemic*; features and feature complexes which are distributed in accordance with a general rule of the language and hence cannot serve to distinguish one morpheme from another, will be called *nonphonemic*.

Each phonemic feature in a segment represents a piece of information that must be supplied from the outside. If the grammar presented here is to be taken as a realistic picture of the functioning of a language, then the instructions for selecting a particular morpheme represent a conscious effort by the speaker, in contradistinction to the operation of various obligatory rules of the language, which the speaker follows automatically when speaking a particular language. Since we speak at a rapid rate – perhaps at a rate requiring the specification of as many as 30 segments per second – it is reasonable to assume that all languages are so designed that the number of features that must be specified in selecting individual morphemes is consistently kept at a minimum. This assumption is embodied in the following formal requirement:

Condition (5): In phonological representations the number of specified features is consistently reduced to a minimum compatible with satisfying Conditions (3) and (4).

It will be necessary in the course of the following discussion to refer to nonphonemic features which have not been specified in a phonological representation. The convention will be adopted here of indicating such unspecified features by writing zeros in the appropriate rows and columns of the matrices.[23] The zeros are auxiliary symbols utilized for purposes of exposition only; they have no function in the phonological system of the language.

1.51 Certain features are nonphonemic because they can be predicted from certain other features in the same segment. Thus, for instance, in Russian the feature diffuse-nondiffuse is nonphonemic everywhere except in the vowels; i.e., its distribution can be predicted in all segments which are nonvocalic and/or consonantal. Similarly, the feature of sharping (palatalization) can be predicted in the segment {*c*}, regardless of context.

In addition to cases of features that are nonphonemic without regard to the context in which they occur, there are well-known cases in all languages of features that are nonphonemic in particular segments in specific contexts only. Since the application of Condition (5) is not restricted to single segments, a feature must remain unspecified in the phonological representation whenever the feature is nonphonemic by virtue of its occurrence in a particular context. Contextual restrictions of this type are called *distributional constraints.* Condition (5), therefore, is the device by means of which the distributional constraints are built into the grammar of the language. This is an important result of the present descriptive framework, since the treatment of distributional constraints has presented serious difficulties to linguistic theory.

The following examples illustrate how distributional constraints are handled in the present theory.

Example 1. Though very common at the junction of two morphemes, only two vowel sequences are admitted within a Russian morpheme; they are {**i*u*} or {**a*u*}; e.g., {pa'uk} "spider", {kl,'auz+a} "intrigue", {t,i'un} "feudal governor". Consequently if it is known that a segment sequence within a morpheme consists of two vowels, it is also possible to infer all other distinctive features of the second vowel except for the accent, and all other features of the first vowel except for diffuseness and the accent. In the dictionary representation of a lexical morpheme containing such a sequence, it is, therefore, necessary to specify only the features vocalic vs. nonvocalic, consonantal vs. nonconsonantal, accented vs. unaccented and, in the first vowel only, also the feature diffuse vs. nondiffuse. All other features can be inferred; hence by Condition (5) they must remain unspecified, e.g.,

[23] In order to save space and aid the reader, the practice will be adopted of transcribing distinctive feature complexes by the appropriate phonetic symbols enclosed in braces ({ }). Since a distinction is made between segments in general and morphonemes, which are the analogs in the present theory of "phonemes", "archi-phonemes", and/or "morphophonemes", italicized letters will be used to transcribe the former and Roman letters, to transcribe the latter. See sec. 1.53 below.

	{*'*a*	**u*}
Vocalic	+	+
Consonantal		
Diffuse		○
Compact	○	○
Low tonality	○	○
Accented	+	○

Example 2. Within a morpheme the feature of voicing is not distinctive before obstruents, except for {**v*} followed by a vowel or by a sonorant, i.e. by a nasal consonant, liquid or glide. The presence or absence of voicing in a sequence of obstruents is uniquely determined by the final obstruent in the sequence. If it is voiced, so are the remaining obstruents in the sequence; if it is voiceless, the remaining obstruents are voiceless, too. In such sequences, therefore, it is unnecessary to specify the feature of voicing in all but the final obstruent.

	{**p*	**s*	*k*}
voicing	○	○	—

1.512 Not strictly part of the distributional limitations are those instances where a feature can be inferred from the grammatical context, rather than from purely phonological factors. For instance, some Russian nouns have accented vowels in one part of their forms and unaccented vowels in another. E.g., {*v*al*} "billow" has the accent on the root vowel in all forms of the singular and on the case endings in all forms of the plural. In representing the lexical morpheme {*v*al*} in the dictionary, it would be patently incorrect to state that the root vowel is accented. Nor would it be correct to say that the root vowel is unaccented. As a matter of fact, the feature of accent cannot be specified until the grammatical context is known in which {*v*al*} is to be used. Once this is known, however, the accent is automatically assigned by the rules of the nominal declension. Since the accent in this case can be inferred from other symbols that must appear in the representation anyway, Condition (5) requires that the feature of accent not be specified here.

In those cases where a feature can be inferred from certain grammatical contexts only, and not from others, this procedure is not followed. Thus, for instance, in Russian the feature of voicing in word final obstruents is governed by the rule that the latter are voiced if followed by a voiced obstruent other than {**v*}, and voiceless otherwise. This rule makes it possible to infer the feature voicing in the final segment of {*r*og*} "horn" in the nom.sg. and acc.sg., but not in other cases. It is, therefore, necessary to represent this lexical morpheme with a final voiced obstruent.

1.52 Russian possesses a series of stems which have forms with and without vowels. Wherever these alternations are not predictable from other – i.e., grammatical or phonological – factors, it is necessary to indicate them in the dictionary representation of the morpheme. This will be done by writing the symbol $\#$ in the position where the vowel is inserted – e.g., $\{t'ur\#k\}$ "Turk", but $\{p'ark\}$ "park"; cf. the respective nom. sg. {t'urok} and {p'ark} and the gen. sg. {t'urk+a} and {p'ark+a}.

It has been shown by Klagstad that with a few exceptions which must be given in a list, the vowel features of $\#$ can be predicted from the context.[24] $\#$ will, therefore, be characterized as vocalic and nonconsonantal with zeros for all other features, i.e., as a vowel without reference to any other vowel feature.

In sum, lexical morphemes are represented in the dictionary by means of two-dimensional tables (matrices) in which each column corresponds to a particular segment and each row to a particular feature. Since all features are binary, they are specified by a plus or minus. Whenever a feature can be inferred from the context, this fact is reflected in the representation by leaving the appropriate row and column unspecified.

The sentence whose derivation was begun in sec. 1.4 is represented, therefore, at this stage as shown in Table I-1.[25]

1.53 It is now necessary to investigate more closely the types of segment to be found in the matrices representing the different morphemes. We define the following relation between segment-types: Segment-type $\{A\}$ will be said to be *different from* segment-type $\{B\}$, if and only if at least one feature which is phonemic in both, has a different value in $\{A\}$ than in $\{B\}$; i.e., plus in the former and minus in the latter, or vice versa.

Examples.

	$\{A\}$	$\{B\}$	$\{C\}$	
Feature 1	+	−	+	$\{A\}$ is not "different from" $\{C\}$
Feature 2	O	+	−	

	$\{A\}$	$\{B\}$	$\{C\}$	
Feature 1	+	−	−	All three segment-types are "different".
Feature 2	O	+		

The set containing all and only "different" segment-types that can be found in the matrices representing the morphemes of the language is called the set of *fully specified morphonemes.* Since fully specified morphonemes serve to distinguish one morpheme from another, they are the analogs of "phonemes" and "morphophonemes" in other linguistic theories. Fully specified morphonemes will be represented by Roman letters enclosed in braces ({ }).

[24] H. L. Klagstad, Jr., *Vowel-Zero Alternations in Modern Standard Russian* (Ph.D. Dissertation, Harvard University, 1954).
[25] Note that many segments are incompletely specified. This is the direct result of Condition (5), which formalizes the notion of descriptive economy.

1. vocalic		−		+	+	+		−		○	−		−	+	○	−	
2. consonantal		+	+		+	−		+		○	+		+	+	○	+	
3. diffuse		○	○	+	○	−		○	○		○		○	○		○	
4. compact		−	+	○	○	+		−	○	+	−		−	○		−	
5. low tonality		+			○	○		+	○	○	−		+	○	○	−	
6. strident	&	+	○	○	○	○	&	−	○	○	−	&Masc.&Nom.&	−	○	○	−	&Nominalizing suffix (masc.)&Sg.&
7. nasal		○	○	○	○	○		−	○	○	+		−	○	○	−	
8. continuant		○		○		○		○	○	○	○		○		○	○	
9. voiced		○	○	○	○	○		−	○	○	○		+	○	○	+	
10. sharped		○	○	○		○		○	○	○	−		○		○	−	
11. accented		○	○	○	○	+		○	○	+	○		○	○	○	○	
		*f	č	*i	*r	'a		*p	j	'a	n		*b	r	*o	d	

1. vocalic		−	+	−		−	+	+		+	−	
2. consonantal		+	−	+		+	−	+	+		+	
3. diffuse		○	○	○		○		○	○	○	○	
4. compact		+	○	+		−		○	+	○	−	
5. low tonality		−	○	+		−		○	+	○	+	
6. strident	Verbal prefix&	○	○	○	&Masc.&Past&	+	○	○	○	○	+	&Sg.&Acc.&
7. nasal		○	○	○		○	○	○	○	○	○	
8. continuant		+	○	−		−	○			○	○	
9. voiced		+	○	+		○	○	○		○	+	
10. sharped		○	○	○		○	○	○		○	+	
11. accented		○	○	○		○	○	○	○	○	○	
		ž	#	g		c	*e	*r	*k	#	v,	

Table I-1. Representation of sample sentence given in sec. 1.4 after lexical morphemes have been selected.

Like all segment-types occurring in phonological representations, fully specified morphonemes are subject to Condition (5), which requires that the number of specified features be kept at a minimum.[26] It can be shown that imposing this condition on a set of fully specified morphonemes is tantamount to requiring that the matrix consisting of the set of fully specified morphonemes be mappable into a branching diagram in such a way that if to each node a particular feature is assigned and the two branches emanating from each node are made to represent the values plus and minus that the feature can assume, then each path through the branching diagram beginning at the initial node and ending at the end points of the branching diagram will uniquely define a fully specified morphoneme. Since in such a branching diagram only phonemic – i.e., specified – features are taken into account, it follows that fully specified morphonemes are uniquely defined by the pluses and minuses without regard for unspecified features.

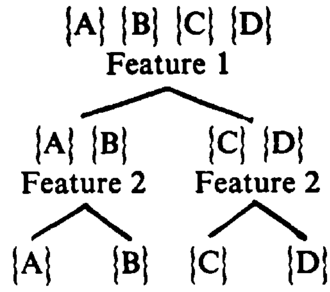
The possibility of mapping a distinctive feature matrix into a branching diagram hinges upon the existence in the matrix of at least one feature for which there are no zeros. This feature, which must be assigned to the first node, subdivides the segment-types into two classes. The next two nodes must be assigned to features which have no zeros for any other segments in the two sub-classes. These may be the same or different features. The same procedure must again be possible with regard to the segment-types in each of the four sub-classes established by the former features; etc. When a sub-class contains a single segment-type, the latter is fully specified, and the path through the branching diagram represents its distinctive feature composition.

Mapping a matrix into a branching diagram is thus equivalent to establishing a hierarchy among the features. This hierarchy, however, need not be complete. For instance, when in a phonological system – cf., Tab. I-3 – there are two features without zeros, there is no reason to put one feature before the other; any order will be satisfactory. A partial ordering of features for different reasons is illustrated in the last example but one below. The hierarchy of features seems to provide an explanation for the intuition that not all features are equally central to a given phonological system; e.g., the distinction between vowels and consonants is more fundamental to phonological systems than the distinction between nasal and oral vowels, or voiced and voiceless consonants.

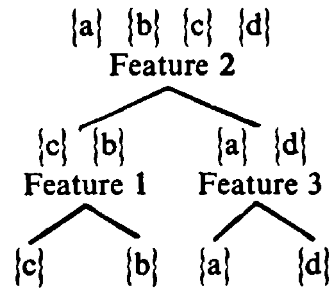
The following examples illustrate the mapping of matrices into branching diagrams, and some of the conditions under which this can and cannot be achieved.

The following discussion is based in part on an unpublished paper by N. Chomsky and M. Halle "On the Logic of Phonemic Description," presented at the M.I.T. Conference on Speech Communication, June, 16, 1956.

	{A}	{B}	{C}	{D}
Feature 1	−	−	+	+
Feature 2	−	+	−	+
Feature 3	O	O	O	O



	{a}	{b}	{c}	{d}
Feature 1	O	+	−	O
Feature 2	+	−	−	+
Feature 3	−	O	O	+



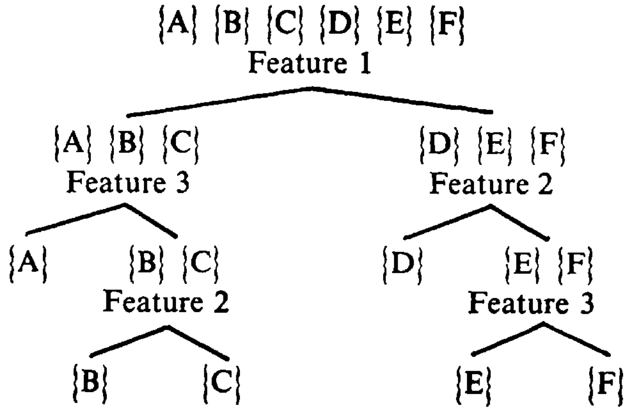
It is not possible to convert every segment-type matrix into a branching diagram. For instance, the following matrix cannot be converted because it lacks a feature without a zero.

	{A}	{B}	{C}
Feature 1	O	+	
Feature 2	+	O	
Feature 3	+		O

In transforming a matrix into a branching diagram it is not necessary that the feature-questions be ordered in the same way in different parts of the branching diagram. Consider the following segment type matrix:

	{A}	{B}	{C}	{D}	{E}	{F}
Feature 1		−		+	+	+
Feature 2	O	−	+		+	+
Feature 3		+	+	O	−	+

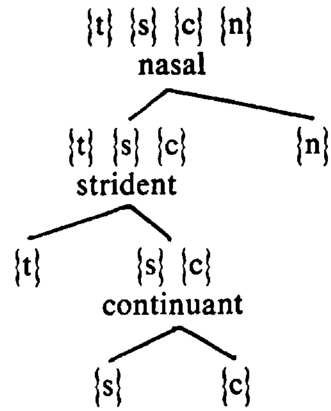
which is converted into a branching diagram where Feature 2 precedes Feature 3 in the left part, and follows Feature 3 in the right part:



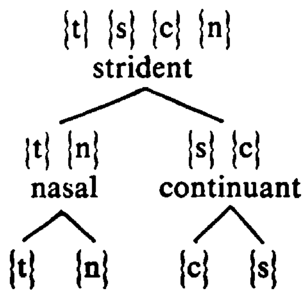
I have been unable to ascertain whether cases of this type arise in natural languages.

The freedom in ordering feature-questions may result in several branching diagrams compatible with the above requirements. In such cases the choice may be dictated by Condition (5) which, in terms of the branching diagram, means that preference be given to the more symmetrical diagram. This can be seen in the following illustration of a partial system (quite similar to that of Russian) where alternate specifications are possible:

	{t}	{s}	{c}	{n}
1. nasal	–	–	–	+
2. strident	–	+	+	O
3. continuant	O	+	–	O



	{t}	{s}	{c}	{n}
1. strident	–	+	+	–
2. nasal	–	O	O	+
3. continuant	O	+	–	O



It is evident that the second ordering is the more economical since it yields the greater number of zeros – a fact which is reflected in the greater symmetry of its associated branching diagram.

The phonological system of a language will be presented by means of a branching diagram; cf., Fig. I-1. Paths through the branching diagram starting at the initial node and terminating in one of its end points define different fully specified morphonemes. It will be shown later that segment types defined by paths starting at the initial node and ending at intermediate nodes – i.e., segment-types that are "not different" from several fully specified morphonemes – play an important role in the functioning of language. These segment-types shall be called *incompletely specified morphonemes* and shall be symbolized by starring the cognate fully specified morphonemes. It is to be noted that a feature specified in a fully specified morphoneme can remain unspecified in an incompletely specified morphoneme, only if all features below it in the hierarchy of the branching diagram also remain unspecified.

1.54 As a consequence of Condition (5) only phonemic features are specified in the phonological representation. In an actual utterance, however, there can be no unspecified features.

Languages differ in the way in which they treat nonphonemic features. For some nonphonemic features there are definite rules, for others the decision as to how the feature is to be implemented is left to the discretion of the speaker. It is this difference which underlies the distinction between what are called *allophonic* and *free variants* of phonemes.

Nonphonemic features in free variation do not properly fit into a linguistic description. The fact that they are in free variation is the only fact of interest insofar as such a description is concerned. This information, however, can be conveyed simply by omitting all mention of such features. Consequently, if in the following description no statement is forthcoming regarding the actualization of a particular feature in a particular context, this feature can be assumed to be in free variation.

1.55 The rules of the grammar are a partially ordered set. It is, therefore, appropriate to inquire what place in this hierarchy should be occupied by the rules governing the nonphonemic distribution of features, which in the present exposition shall be termed the *F rules*. It will be recalled that at the end of the Phrase Structure level the lexical morpheme class symbols will have been replaced by sequences of distinctive feature segments (matrices), but the grammatical morpheme class symbols will remain in the representation unchanged (cf., the representation in Tab. I-1). Only after the (transformational) rules of inflection and derivation have been applied, will grammatical morpheme class symbols like "Past", "Sg.", etc., be replaced by their phonological consequences. Since the transformational rules introduce into the representation additional distinctive feature segments as well as modify segments previously introduced, the decision to place the F rules before the transformations may force us to apply certain F rules twice, once before the last transformational rules, and again after the last transformational rule. Thus, for example, the (transformational) rules of Russian substantival inflection replace &{*iv'an*}&Sg.&Dat.& by {*iv'anu*}. If the rules assigning nonphonemic features to unaccented vowels had been applied before this transformation, it would be necessary either to apply the same rules again at this point, or to specify all the nonphonemic features in $\{u\}$ in some other way. It may, therefore, seem most advantageous to put all rules governing the distribution of nonphonemic features after the transformational rules. There are, however, reasons which make it desirable to apply some of the F rules before the transformations, even if that entails complications of the kind just outlined.

In Russian and in many other languages – though perhaps not universally – there are transformational rules, in particular rules of inflection and derivation, which require for their proper operation that certain features be specified in the representation regardless of whether or not these features are phonemic. For instance, the rules of the Russian conjugation require for their operation the information of whether or not the verbal stem ends in a vowel.[27] In the third segment of the verbal stem $\{rv'a\}$ "tear" the features vocalic-nonvocalic and consonantal-nonconsonantal are nonphonemic, since in Russian morphemes beginning with a sequence of a liquid followed by a consonant, the third segment must be a vowel (see MS rule 1c, Chapter II, sec. 2.161). Condition (5) would, therefore, require that the phonological representation of this morpheme be in part as follows:

	r	*v*	*'a*
vocalic-nonvocalic	+		O
consonantal-nonconsonantal	+	+	O

Since, however, the features of the third segment are not specified it cannot be ascertained whether or not it is a vowel, and hence the stem cannot be properly conjugated. If, however, the F rule specifying these nonphonemic features – rule MS 1c – is applied before the transformation, the difficulty is avoided. Since there are many instances of this type we concluded that at least some F rules must be applied before the transformational rules in spite of the complications that this may entail.

1.56 The considerations just reviewed have established the necessity of splitting the F rules into two parts; one part, which shall henceforth be called the *morpheme structure* or *MS rules*, must be applied before the transformations, and the other part, to be called the *phonological* or *P rules*, must be applied after the transformations. The next question which arises naturally is on what basis can the decision be made to class a particular F rule with the MS rules or with the P rules. In the case of Russian the following criterion was found to yield the appropriate results:[28]

The MS rules must insure that all distinctive feature segments appearing in the representation be either fully or incompletely specified morphonemes.

In other words, the set of segment-types occurring after the application of the MS rules is defined by all possible paths through the branching diagram which begin at its initial node. As noted at the end of sec. 1.53 this places a restriction on the features that can remain unspecified: certain nonphonemic features must be specified at this

[27] See R. Jakobson, "Russian Conjugation," *Word*, 4, 155–167 (1948).

[28] I believe that the criterion will hold for other languages as well; I have, however, not been able to check this.

point. This result, however, is precisely the one desired, for, as was shown in the preceding section, unless some such limitation on the occurrence of unspecified features is imposed at this point, it will not be possible to apply properly the (transformational) rules of inflection and derivation.

It will have been noticed that incompletely specified morphonemes are analogous to the Prague school's "archi-phonemes".[29] Although Trubetzkoy defined the latter as "the set of distinctive features shared by two phonemes",[30] in his practice – as exemplified, for instance, in his *Das morphonologische System der russischen Sprache* – he operated with "archi-phonemes" in which there was more than one neutralized (unspecified) feature. As a matter of fact, the MS rules given in Chapter II insure that the (transformational) rules of Russian morphology will operate with incompletely specified morphonemes that are substantially identical with the "archi-phonemes" postulated by Trubetzkoy in the last mentioned work.

1.57 The decision to split the F rules into two parts and to apply the MS rules before the transformations is further supported by the fact that in many languages there is a striking difference between the constraints that hold for segment sequence within single morphemes only, and those that hold for segment sequences in general, without regard for morpheme junctions. Thus, for instance, in Russian, vowel sequences are severely restricted within single morphemes – cf., MS rule 11 a, Chapter II, sec. 2.21 – while across morpheme junctions practically all two-vowel sequences are admitted. Put differently, in vowel sequences within morphemes many features are nonphonemic and hence must remain unspecified in the representation.

Many of the rules which specify these nonphonemic features can be applied only as long as the individual morphemes are delimited.[31] The transformations, however, may re-arrange symbols in such a way that the individual morphemes are no longer separate. There has already been occasion to mention the phenomenon of "syncretism", which is an example of such a process. Another instance is the so-called "discontinuous morphemes", which are particularly common in Semitic; "discontinuous morphemes" can also be found in many Indo-European languages, including Russian. For example, in the neuter adjective {p'ust+o} "empty" the neuter gender is reflected by the placement of the accent on the stem *and* the ending {+o}. Since the transformations may obliterate the separation among morphemes, those F rules which require for their operation the information of where a morpheme begins and ends must obviously be applied before the transformations.

1.58 The MS rules of Russian are given in Chapter II, sec. 2. After the MS rules have been applied all segments appearing in the representation are either fully or

[29] On "archi-phonemes" see, e.g., N. S. Trubetzkoy, "Die Aufhebung der phonologischen Gegensätze," *Travaux du cercle linguistique de Prague, 6*, 29–45 (1936): A. Martinet, "Neutralisation et archiphonème," *ibid.*, 46–57 and recently R. I. Avanesov, "O trex tipax naučno-lingvističeskix transkripcij," *Slavia, 25*, 347–371 (1956).

[30] N. S. Troubetzkoy, *Principes de phonologie* (Paris, 1949), p. 81.

[31] Cf., the MS rules in Chapter II, secs. 2.161–2.163, which require reference to the beginning and end of the morpheme.

incompletely specified morphonemes. Since all morphonemes are uniquely defined by paths through the branching diagram representing the phonological system of the language, it is possible to replace the matrices representing different lexical morphemes by linear sequences of pluses and minuses, provided that a special symbol – an asterisk in the present instance – is introduced to signal the end of the specification of incompletely specified morphonemes. No symbol is required to signal the end of the specification of fully specified morphonemes, since this can be automatically determined. In the illustrative example a space has been introduced at these places in order to aid the reader. Unlike the asterisk, however, the space is a redundant symbol and need not appear in the representation.

The results of applying the MS rules to the sentence illustrated in Tab. I-1 can, therefore, be written as follows:

```
&{−+−++*−++−− +−+−− ++−− +−++}&{−+−+−−−*−−
   *f     č              r    'a        *p    j
+−−++ −+−−−+−}&Masc.&Nom.&{−+−+−−+*++−−
  'a                            *b     r
+−−−+*−+−−−+−}&Nominalizing suffix (masc.)&Sg.&Nom.&Verbal
  *o      d
prefix&{−++−++ +−*−+++−+}&Masc.&Past&{−+−−+−
           ž     #     g                       c
+−−−−*++−*−+++−      +−*−+−++++}&Sg.&Acc.&
  *e     *r      k     #      v,
```

The pluses and minuses of this representation must be interpreted with the help of the branching diagram representing the phonological system of Russian; see Fig. I-1, sec. 2.1 below. They are to be taken as instructions for running through the diagram always beginning at the initial node, a plus indicating that the right branch is to be selected, a minus, that the left branch is to be selected. When in this manner an end point of the diagram or an asterisk in the representation is reached, the whole process is repeated beginning again at the initial node. Following this procedure it can be readily determined, e.g., that the first segment in the above representation is the incompletely specified morphoneme defined by the features nonvocalic, consonantal, noncompact, low tonality, strident.

1.581 Allowing incompletely specified morphonemes in the representation entails the following important consequence. Consider the noun {*l*es} "forest",[32] which has the accent on the case endings in the plural and in the loc. II sg., and the accent on the stem vowel in all other cases of the singular. In the light of the discussion in sec. 1.512, the gen.sg. will be represented as {*l'es+a} and the nom.pl. as {*les+'a}.

[32] The feature of sharping (palatalization) is not specified in {*l}, because before {*e} all liquids and noncompact consonants except {c} (i.e. all paired morphonemes, cf., Chapter II, sec. 3, comment before rule P 6a) are automatically sharped; see Chapter II, sec. 3, rule P 1a.

Since, however, {*les+'a} and {l,is+'a} "fox" – as well as all unaccented {e} and {i} – are homophonous, it is necessary to add a rule stating that unaccented {e} becomes [i], or the equivalent in terms of distinctive features.[33] Thereby, however, we have admitted unaccented {e} (and also unaccented {o}) into the phonological system of Russian in spite of the fact that these feature bundles do not serve to distinguish utterances. This is a direct violation of Condition (3a-1),[34] which specifically rules out this step. Since Condition (3a-1) has been rejected as a requirement on phonological representations, this violation is not unexpected. It is, however, to be noted that the alternative to violating Condition (3a-1) is to set up multiple representations for all lexical morphemes containing the vowel {*e}. For instance, it would be necessary to list {*l*es} as both /l,'es/ and /l,is-/, which is evidently an undesirable complication.

1.6 It was remarked in sec. 1.42 above that after the application of the transformational rules, which include the rules of derivation and inflection, the representation of the sentence will contain only phonological symbols; i.e., morphonemes and boundaries. The grammatical morpheme class symbols will be replaced by their phonological consequences, and the symbol # (vowel alternating with zero) will be converted either into a vowel or eliminated from the representation. This leaves only the & marker to be accounted for.

Condition (6): The & markers are translated by the rules of morphology into phonological boundaries or altogether eliminated.

The exact description of this process of translation is part of the morphology of the language and can, therefore, not be given here in detail. In the present study the boundaries will only be listed and the contexts in which they are found will be enumerated.

Russian possesses five (phonological) boundaries, which are denoted by the following marks:

1) The phonemic phrase boundary is denoted by a vertical bar |.

2) The word boundary is denoted by a space, or, in cases where confusion could arise, by a % (percent) sign.

3) Prefix and preposition boundaries are denoted by an = ("equals") sign.

4) Certain word final suffixes are preceded by a special boundary denoted by a + (plus) sign, or, in cases where confusion could arise, by a § (section) sign.

5) Morpheme boundaries in abbreviations of the type {p'art–b,i*l'et} are denoted by a – (dash).

Since no phonological consequences are connected with any other & markers, all remaining & markers are eliminated from the representation. When in the course of the exposition, the need arises to refer to these morpheme junctions, a hyphen (-) will be used. The hyphen is, however, not a symbol in the phonological representation.

1.7 The specification of the sample sentence can now be continued. The application of the transformational rules of the language yields the following representation:

[33] See Chapter II, sec. 3, rules P 7g-9a.

[34] Cf., sec. 1.32 above.

```
{ | −+−++*−++−−   +−+−   ++−−   +−−++ | −+−+−−−*
     *f          č        i       r          'a '         *p
    +−−++ −+−−−+− +−+−−       % −+−+−−+* ++−−
 j      'a         n          i       j            *b         r
+−−−+−   −+−−−−++   +−−++   −+++−+  §  +−−+− |
     o           d,             'a          g              a
−+−−++−− = −++−++  +−−−++  −+++−+ % −+−−+−
      s                ž            'o          g             c
+−−−−+ ++−*−+++−−−  +−−−+−  −+−++++ | }
     'e        *r            k             o            v,
```

This is the phonological representation of the sentence, for it consists entirely of morphonemes and boundaries, and all rules that are needed to turn this representation into sound deal exclusively with effects of specific configurations of distinctive features and/or boundaries, on particular distinctive feature complexes.[36]

The phonological rules can be formulated so as to contain no reference to the derivational history of morphonemes and boundaries, provided that an order is imposed on the application of the rules. If the rules are not ordered, their structure will have to be more complex, for they will require reference to the derivational history of the symbols.

As an illustration, consider the following example. In Russian all liquids and paired consonants are sharped (palatalized) before {*e}. Moreover, when unaccented, {e} becomes diffuse; i.e., [i]. The simplest statement of these facts is the following:

Rule A. Before {*e}, liquids and paired consonants are sharped.

Rule B. Unaccented {e} becomes diffuse.

If, however, Rule B is applied first, Rule A will have to be replaced by the following:

Rule A′. Before {'e} and before [i] which derives from {e}, liquids and paired non-compact consonants are sharped.

Rule A is evidently simpler than Rule A′; it can, however, be used only if an order is established in the application of the rules.

Table I-2 illustrates the operation of the P rules of Russian in the case of the sample sentence. In the initial stage each morphoneme is represented by its set of distinctive features, which are interpreted with the help of the branching diagram (Fig. I-1) representing the phonological system of Russian. The morphonemes are then modified as a result of the application of the individual P rules. Since only some of the P rules are relevant for the illustrative example, not all P rules appear in Tab. I-2. The first rule to be applied is P 1b, which assigns the feature of voicing to certain morphonemes in which this feature is unspecified. Next, rule P 2 is applied with the consequences indicated in Table I-2. Then succeeding rules are applied in order until the list of rules is exhausted, resulting ultimately in the following "narrow" transcription of the sentence, which can be directly converted into sound.

[35] These rules are the phonological or P rules; cf., sec. 1.56 above and Chapter II, sec. 3.

	*f	č	*i	r	'a	*p	j	'a	n		j	b*	r	*o	d,	'a	g	+	a	s	ž	'o	g	c	'e	*r	k	o	v,
vocalic				+	+								+	+											+	+			
consonantal	+	+		+		+			+			+	+		+		+			+	+		+	+		+	+		+
diffuse	O	O	+	O		O	O		O	+	O	O	O		O		O		O	O	O		O	O		O	O		O
nasal	O	O	O	O	O		O	O	+	O	O		O	O		O	O		O	O	O	O	O	O	O	O	O	O	O
accented	O	O	−	O	+	O	O	+	O		O	O	O	−	O	+	O			O	O	+	O	O	+	O	O		O
voiced	O	O	O	O	O		O	O	O	O	O		O	O		O			O	−	+	O	+	O	O	O		O	+
P 1b		−																						−					
P 2																													−
P 3a	−																			+			−						
compact		+	O	O	+		O	+		O	O		O	−	−	+				−					−	O	+	−	−
P 3b																				+									
P 8														+														+	
P 10b																									+				
continuant	O	−	O		O	O	O	O	O	O	O	O	−	O	O	O						O			O			O	O
P 5a	+					−						−			−														+
strident	+	O	O	O	O		O	O		O	O			O		O	O		O		O	O	O	+	O	O	O	O	+
P 5a																	−										−		
sharped	O	O	O		O	O	O	O		O	O	O	−	O	+	O	O		O		O	O	O	O	O	O	−	O	+
P 5b		+					+				+										−			−					
P 6a						+																							
P 6c												−																	
P 7b	−																									−			
P 7f																	−						−						
low tonality				O	O	+	O	O	−	−	O	+	O	+	−	O	+		O						−	O	+	+	+
P 7e					−			−								−			−										
P 8														−														−	

	*f	č	*i	r	'a	*p	j	'a	n		j	b*	r	*o	d,	'a	g	+	a	s	ž	'o	g	c	'e	*r	k	o	v,
grave (back)																													
P 10c																													
P 11																													
P 12																						+						+	
"[u]-transition"																													
P 13a																								+					
"[i]-transition"																													
P 13b											+				+														+
tense-lax																													
P 14a+14b																+						+			+				
reduced																													
P 14a+14b										+																		+	
prominence																													
P 15					1			2								1						2			1				
P 16			2							3				3					3									3	
P 17																			4									4	

Table I-2. Applications of P rules to sample sentence. See sec. 1.4 and Table I-1.

3 1   2 3 3 1 4   2 1 4
|fčira|p,janijbrad,agə|žžokcerkəf,|

The numerals above the vowel symbols indicate the degree of prominence, 1 being the highest and 4 the lowest.

In principle the P rules should be extended to the point where all distinctive features of all segments are specified, including indications where a given feature is in free variation. It would, thus, have been necessary to include, e.g., a rule stating that all sonorants in Russian are voiced, except in special cases like {o*kt,'abr, *skoj} "pertaining to October" where {r,} is frequently unvoiced. Such rules have not been included here. Since the facts are frequently in dispute, it was felt that additional details would contribute little towards attaining the objectives of the present study.

2. *The Phonological System of Russian*

When a phonological analysis is presented, the question always arises as to what extent the proposed analysis covers the pertinent data. It is clearly impossible in a description to account for all phonological manifestations in the speech of even a single speaker, since the latter may (and commonly does) use features that are characteristic of different dialects and even foreign languages. (E.g., a speaker of Russian may distinguish between nasalized and nonnasalized vowels in certain [French] phrases which form an integral part of his habitual conversational repertoire.) If such facts were to be included, all hopes for a systematic description would have to be abandoned. It is, therefore, better to regard such instances as deviations to be treated in a separate section and to restrict the main body of the grammar to those manifestations which can be systematically described.

The variety of Russian described here is substantially identical with the one described in such standard works of Russian as the recent Academy Grammar and the dictionary of Russian pronunciation edited by Avanesov and Ožegov.[36] In some of its features this "literary" type of Russian admits several alternatives. An effort has been made to include these vacillations in the description.[37] It is interesting that these vacillations do not affect the phonological representation of utterances, but rather influence the order and content of the phonological rules which translate the phonological representations into sound.

2.1 *The Morphonemes.* The morphonemes of Russian are presented in a branching diagram (Fig. I-1) from which a distinctive feature matrix (Tab. 1-3) has been constructed. The system contains 43 morphonemes which are specified by means of 271 distinctive feature statements (pluses or minuses in Tab. I-3, or branches in Fig. I-1),

[36] V. Vinogradov, E.·S. Istrina, S. G. Barxudarov, eds., *Grammatika russkogo jazyka*, I (Moscow, 1952), and R. I. Avanesov and S. I. Ožegov, *Russkoe literaturnoe udarenie i proiznošenie – Opyt slovarja-spravočnika* (Moscow, 1955).

[37] Cf., e.g., Chapter II, sec. 3, rules P 7g, P 7g', P 8, P 9a, P 9a'.

or 6.3 distinctive feature statements per morphoneme. Condition (5) requires that the number of distinctive feature statements utilized in the representations be reduced to a minimum. In order to obtain an indication of how well Condition (5) is satisfied by the above branching diagram, one might compare the quoted figure with $\log_2 43 = 5.26$, which is the lower limit that may be obtained by the procedure of reducing the number of feature statements to a minimum. It is to be emphasized that this comparison must be treated with a great deal of caution: its only purpose is to show that the minimization process has achieved results of the type that might reasonably be expected.

	j	t	d	t,	d,	n	n,	c	s	z	s,	z,	p	b	p,	b,	m	m,	f	v	f,	v,
vocalic	–																					
consonantal	–	+	+	+	+	+	+	+	+	+		+		+	+	+	+		+	+	+	+
diffuse	O	O	O	O	O	O	O	O	O	O		O		O	O	O	O		O	O	O	O
compact	O																					
low tonality	O																					+
strident	O							+		+		+									+	+
nasal	O					+		O	O	O	O	O					+		O	O	O	O
continuant	O	O	O	O	O	O	O			+	+	+	O	O	O	O	O	O	O	O	O	O
voiced	O		+		+	O	O	O		+		+		+		+	O	O		+		+
sharped	O			+	+		+	O			+	+			+	+		+			+	+
accented	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O

	č	š	ž	k	k,	g	x	e	'e	o	'o	a	'a			u	'u	r	r,	l	l,
vocalic										+	+	+	+			+	+	+	+	+	+
consonantal	+	+	+	+	+	+	+											+	+	+	+
diffuse	O	O	O	O	O	O	O							+	+	+	+	O	O	O	O
compact	+	+	+	+	+	+	+					+	+	O	O	O	O	O	O	O	O
low tonality				+	+	+	+			+	+	O	O			+	+	O	O	O	O
strident	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O
nasal	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O
continuant		+	+				+	O	O	O	O	O	O	O	O	O	O			+	+
voiced	O		+			+	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O
sharped	O	O	O		+	O	O	O	O	O	O	O	O	O	O	O	O		+		+
accented	O	O	O	O	O	O	O		+		+		+		+		+	O	O	O	O

Table I-3. Matrix representing the morphonemes of Russian.

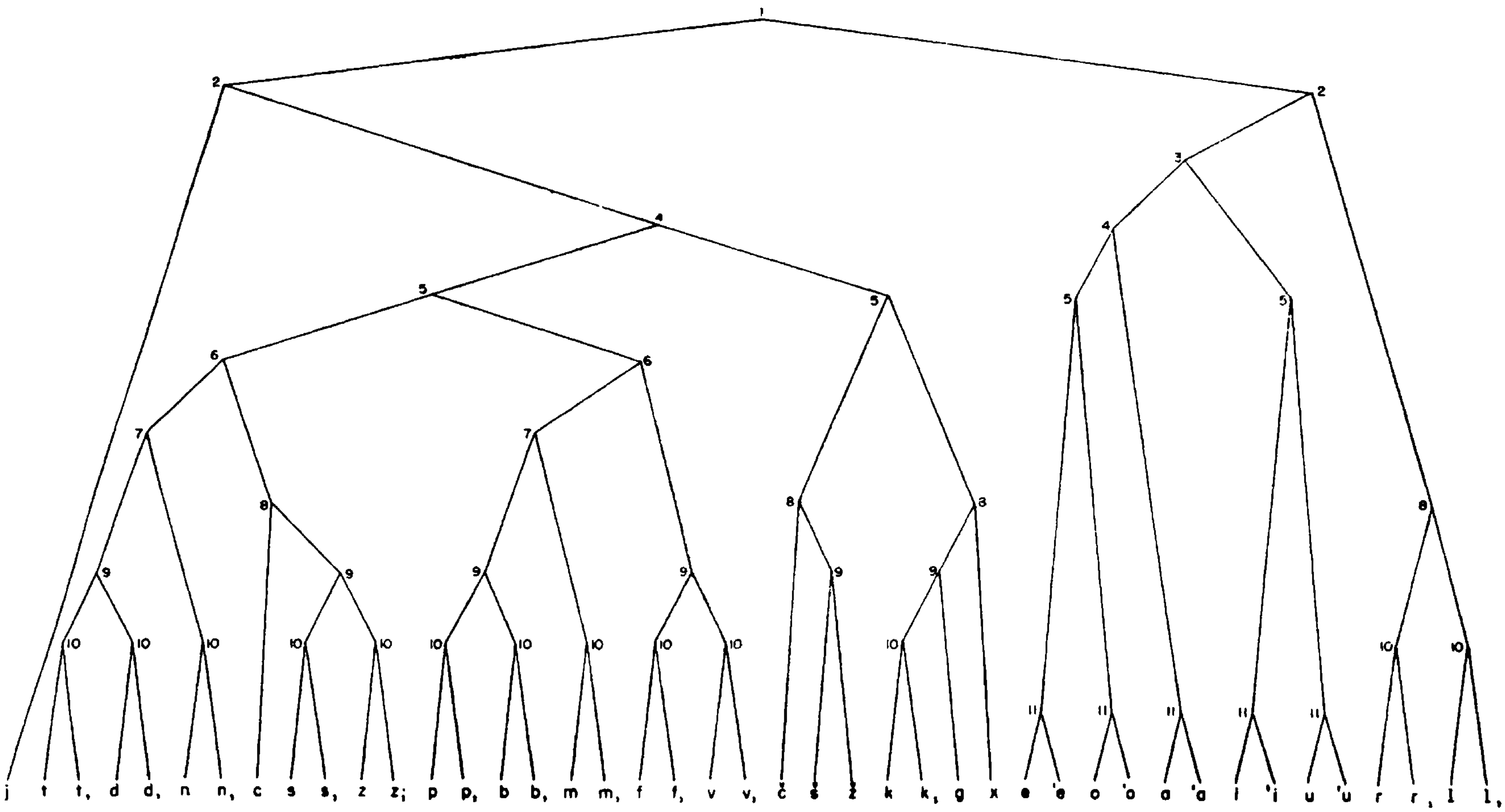


Fig. I-1. Branching diagram representing the morphonemes of Russian. The numbers with which each node is labelled refer to the different features, as follows: 1. vocalic vs. nonvocalic; 2. consonantal vs. nonconsonantal; 3. diffuse vs. nondiffuse; 4. compact vs. noncompact; 5. low tonality vs. high tonality; 6. strident vs. mellow; 7. nasal vs. nonnasal; 8. continuant vs. interrupted; 9. voiced vs. voiceless; 10. sharped vs. plain; 11. accented vs. unaccented. Left branches represent minus values, and right branches, plus values for the particular feature.

2.11 Examples of phonemic oppositions in monosyllabic words ending in {l}:

	{'i}	{'e}	{'a}	{'o}	{'u}
{p}	{p'il} "ardor"		{p'al} "fell" (masc.)	{p'ol} "floor"	{p'ul} "business pool"
{p,}	{p,'il} "drank" (masc.)	{*p'el} "sang" (masc.)	{p,'al} "scaffold" (gen.pl.)		
{b}	{b'il} "was" (masc.)				{b'ul} "papal bull" (gen.pl.)
{b,}	{b,'il} "beat" (masc.)	{*b'el} "white" (masc.)			
{f}					
{f,}					
{v}	{v'il} "howled" (masc.)		{v'al} "big wave"	{v'ol} "ox"	
{v,}	{v,'il} "fork" (gen.pl.)		{v,'al} "limp" (masc.)	{v,'ol} "led" (masc.)	
{m}	{m'il} "washed" (masc.)		{m'al} "small" (masc.)	{m'ol} "pier"	{m'ul} "mule"
{m,}	{m,il} "nice" (masc.)	{*m'el} "chalk"	{m,'al} "crumpled" (masc.)	{m,ol} "swept" (masc.)	
{t}	{t'il} "rear"		{t'al} "willow"		{t'ul} "quiver" (gen.pl.)
{t,}		{*t'el} "body" (gen. pl.)			
{d}			{d'al} "gave" (masc.)	{d'ol} "dale"	{d'ul} "bore" (gen.pl.)
{d,}		{*d'el} "deed" (gen. pl.)			
{s}					
{s,}	{s,'il} "force" (gen.pl.)	{*s'el} "sat" (masc.)		{s,'ol} "village" (gen.pl.)	
{z}			{z'al} "hall" (gen.pl.)	{z'ol} "angry"	
{z,}					
{c}		{c'el} "whole" (masc.)			
{n}	{n'il} "whined" (masc.)				
{n,}	{n,'il} "Nil" (name)				
{k}			{k'al} "feces"	{k'ol} "stake"	
{k,}	{*k'il} "hernia" (gen.pl.)				
{g}				{g'ol} "naked"	{g'ul} "rumble"
{x}	{x'il} "puny" (masc.)		{x'al} "white bread" (gen.pl.)		{x'ul} "blame" (gen.pl.)
{č}			{č'al} "roan"		
{š}	{š'il} "sewed" (masc.)		{š'al} "mad" (masc.)	{š'ol} "went" (masc.)	{š'ul} "testicle" (gen.pl.)
{ž}	{ž'il} "lived" (masc.)		{ž'al} "squeezed" (masc.)		
{r}	{r'il} "dug" (masc.)		{r'al} "plow" (gen.pl.)	{r'ol} "cylinder"	
{r,}					
{l}			{l'al} "ruby"		
{l,}	{l,'il} "poured" (masc.)				
{j}		{j'el} "ate" (masc.)	{j'al} "yawl"		{j'ul} "whirligig" (gen.pl.)

Not all oppositions of the language are exemplified above. The lacunae are supplied herewith:

The opposition sharp vs. plain (palatalized vs. nonpalatalized):

{f} vs. {f,}: {f'irkat,} "to neigh" {f,'ink+a} "Finnish woman"
{t} vs. {t,}: {t'ikat,} "to address someone by the familiar form" {t,'ikat,} "to tick"
{s} vs. {s,}: {*v'es} "weight" {*v'es,} "entire"
{z} vs. {z,}: {z'ib,} "ripple" {z,'im} "winters" (gen.pl.)
{r} vs. {r,}: {r'ad} "glad" {r,'ad} "row, series"
{l} vs. {l,}: {l'is} "bald" {l,'is} "foxes" (gen.pl.)
{k} vs. {k,}: {sadk+'om} "animal nursery" (instr.sg.) {so=tk, 'om} "let us weave it together"

2.2 *The Boundaries.* Condition (6) requires that the (transformational) rules of morphology translate the & markers into phonological boundaries or eliminate the & markers from the representation. Since & markers were postulated in the Phrase Structure level in order to delimit the individual morpheme class symbols and to join them to their neighbours, Condition (6) assures that boundaries will appear only at junctions between morphemes.

It is impossible in the present work to give a systematic account of when and how the different boundaries are introduced into the representation, since this would demand a full exposition of the morphology of the language. Instead we shall list the contexts in which each of the different classes of boundary are found, and state the phonological consequences of each class of boundary.

2.21 Phonemic Phrase Boundaries. Boundaries, symbolized by a vertical bar |, delimit the phonemic phrase. In normal speech a pause may be made at the end of a phonemic phrase, but may not be made in the middle of a phonemic phrase. Accented vowels within a phonemic phrase have different degrees of prominence relative to each other, the degrees of prominence being determined by the P rules. The phonemic phrase is, therefore, the domain over which the rules of prominence apply.[38]

It was noted above in sec. 1.4 that the branching points in the Phrase Structure trees represent different Immediate Constituents of the sentence. Only elements in the sentence that can be traced back to a common branching point are Constituents of the sentence. It can thus be seen from the tree given in sec. 1.4 that in the sample sentence the sequence {s=ž'og c'erkov,} is an Immediate Constituent, for it can be traced back to the common branching point &Verb Phrase&, while the sequence {*fč*ir'a *pj'anij} is not a Constituent, since it lacks a common branching point.

Phonemic phrase boundaries are introduced

a) at the beginning and end of sentences;

b) before and after the longest Immediate Constituent (i.e., the one associated with the highest node in the Phrase Structure tree) containing not more than two and not less than one accented vowel.

[38] Cf., Chapter II, sec. 3.2.

As a consequence of b), pauses are normally not admitted between the adjective and the noun it qualifies. If, however, more than one adjective qualifies the same noun, a pause is admitted. Moreover, unaccented words must adjoin an accented word without pause.

2.22 Word Boundaries. Boundaries, symbolized either by a space or by a % (percent) sign, delimit the "phonological" word. Word boundaries are introduced at the following places:

a) before and after unaccented proclitics, enclitics, conjunctions and adverbs;

b) after the grammatical morpheme class symbol &Imperative&[39]; e.g., {žg'i%t,i} "burn!" (imperative pl.);

c) at all phonemic phrase boundaries;

d) before and after the longest Immediate Constituent containing a single accented vowel.

As a result of d) the "phonological" word does not necessarily coincide with the "grammatical" word, which is normally considered to be the longest Immediate Constituent associated with a single sequence of inflectional morphemes. Divergences occur in the following three types of *composita:*

1) where the first member has an inflectional ending of its own; e.g., {*tr,'ox% gol'ovij} "three headed", but cf., {*dvu*sm'i*sl,onnij} "ambiguous";

2) where the first member contains two lexical morphemes; e.g., {para-x'oda% *s*troj'it,il,nij} "ship building" (adj.);

3) where the two members form a *dvandva* compositum – i.e., are a transformation of a coordinate construction; e.g., {*s'e*vera%z'apad} "north west".[40]

The above defines the "phonological" word, which is constituted by one or more morphemes preceded and followed by a word boundary. As has just been shown, the postulation of the word boundary is determined by various grammatical considerations. The word boundary, like all other phonological boundaries, has certain phonetic consequences; its introduction into the representation, however, is not determined by these consequences, but by the set of conditions outlined above.

Some of the phonetic properties of the "phonological" word might be mentioned here (others will be discussed in the appropriate places below): All words containing more than one vowel are organized about a single accent. There are, therefore, in Russian, no polysyllabic words without an accent, nor are there "phonological" words with more than one accented vowel. A monosyllabic word may or may not contain an accented vowel; e.g., {*s*t'ep,} "steppe", {*s*kv'oz,} "through" (autonomous preposition) have accented vowels; conjunctions like {da} "and" and enclitics like {ži} "indeed" contain an unaccented vowel.

2.23 Prefix and Preposition Boundaries. Prefixes and nonautonomous (un-

[39] This solution was first proposed by R. Jakobson; see his "Russian Conjugation," *Word, 4*, 159 (1948).

[40] The above rules are part of the morphological rules of Russian. See R. Jakobson's review of G. Trager, *Introduction to Russian, Slavonic and East European Review, 22*, 122 (1944).

accented) prepositions are not considered separate words in Russian. Prefixes and nonautonomous prepositions are always followed by a boundary symbolized by an "equals" sign (=); e.g., {s=iv'an+om} "with Ivan", {pro=igr'at,} "to lose". This boundary plays an important role in the distribution of the sharping feature in sequences of acute compact (palatal) consonants.[41]

2.24 Suffix Boundaries. A special boundary is symbolized by a + (plus) or by a § (section) sign. This boundary is postulated before all final inflectional suffixes consisting of the vowel {*a} or {*o}, and before substantival case endings beginning in {*o} or {*a} and not followed by {j}. After the + boundary, unaccented {a} and {o} are not changed into [i] when preceded by sharped (palatalized) morphonemes. Examples: {p,at'ak} [p,it'ak] "five copeck coin" and {p,it'at,} [p,it'at,] "to nourish" have identical vowels, but {v=m'or,e} [vm'or,i] "in the sea" is not homophonous with {v=m'or,+o} [vm'or,ə] "into the sea". Note also: {d,n,+'om} [d,n,'om] "in the day time" (instr. sg.), and {p'old,n,+om} [p'old,n,əm] "at noon" (instr. sg.); {il,jič+'om} [il,jič,'om] "Ilyich" (instr. sg.) and {s'av,ič+om} [s'av,ič,əm] "Savič" (instr. sg.).[42] No + boundary is postulated where the suffix ends with {j}; e.g. {žon'oj} [žɨn'oj] "wife" (instr. sg.) and {b'an,oj} [b'an,ij] "bath house" (instr. sg.), or before polyphonemic suffixes other than substantival case endings; e.g. {moj'om} [maj'om] "my" (loc. masc.) and {k'ojom} [k'ojim] "which" (loc. masc.); {p,oč'om} [p,ič,'om] "we bake" and {v'i=p,očom} [v'ɨp,ič,im] "we shall bake" (perf.) and {v=čj'om d'ome} [fč,j'om-d'om,i] "in whose house", {v=čila*v'ečjom} [fč,ilav,'eč,jim] "in a human (house)".

2.25 Boundaries in Abbreviations. A boundary, symbolized by a dash (-), is introduced between the lexical morphemes of abbreviations; e.g. {p'art-b,i*l'et} "party card" This boundary is needed because in such abbreviations the voicing feature is distributed differently than in other contexts. In the example just cited, e.g., the sequence {t-b} is pronounced [tb], whereas in all other contexts sequences of voiced and voiceless consonants are not admitted.[43]

All other & markers are eliminated, for no phonological consequences are connected with these markers.

3. *Comments on Details of the Phonological System of Russian*

3.1 *The Affricates.* In some analyses of Russian the affricates {c} and {č} are treated as sequences, i.e., as {ts} and {tš} respectively. In the present study they are treated as single segments. The solution adopted here must now be justified.

In Russian there are many noun stems in which a zero alternates with a vowel, which is always inserted before the last morphoneme of the stem, e.g., {kot,'ol} {kot,l+'a}

[41] See Chapter II, sec. 3, rule P 4.

[42] Vacillations are, however, quite common, particularly where the vowel is followed by a sharped morphoneme, i.e., in instr. pl. Cf., R. Avanesov, *Očerki russkoj dialektologii*, I (Moscow, 1949), pp. 118-119.

[43] Cf., below, Chapter II, sec. 3, rules P 1b, 2, and 3a.

"pot" (nom. and gen. sg.). Stems ending in {c} are treated in the same way – e.g., {ov,'ec} {ov,c+'a} "sheep" (fem.) (gen. pl. and nom. sg.) – if {c} is assumed to be a single segment. If {c} is considered a sequence, this would constitute an exception to an otherwise perfectly general rule. It is, therefore, preferable to consider {c} a single morphoneme.

The case of {č} is more complicated. {č} differs from the sequence {tš} by being sharped. {tš}, however, is found only across morpheme junctions, while {č} is found within single morphemes. If, in view of their complementary distribution, {č} is to be considered a positional variant of {tš}, it will, however, be necessary to include an additional boundary into the phonological representation in order to account for the different phonetic effects in examples like {za=*pl,'ot-ši} [zapl,'otšɨ] "having braided" as opposed to {*dvojot'očij} [dvəjit'oč,ij] "colon" (gen. pl.). The choice thus lies between considering {č} a single morphoneme or admitting an additional phonological boundary.[44] Since the latter solution would also require that the phonological rules be further complicated to account for the phonetic effects just noted, it is more economical to consider {č} a single morphoneme.

3.2 *The Compact Acute* (*Palatal*) *Consonants*. The compact acute (palatal) consonants raise special problems only in the Moscow literary standard. Since the facts are well known, they shall be reviewed only briefly.

(1) On the morpheme boundary after a prefix or preposition we have [šš], [žž], and [š,č,]: [šš'it] "sewn together", [žž'ok] "burned down", and [š,č,'as,t,ju] "with a part".

(2) Within a morpheme and at the junction of a root and a derivational suffix, we have [šš] [ž,ž,] [š,š,]: [n,'iššɨj] "lowest"; [daž,ž,'a] "rain" (gen. sg.); [n,'iš,š,ij] "poor", [r,'eš,š,i] "sharper".

In each context there are three distinct sequences; there is, therefore, need for only three compact aċute (palatal) morphonemes, if rules for the distribution of the other features can be stated.

Let the utterances be represented as follows: {s=š'it,} {s=ž'og} {s=č'a*st,ju} {n,'izšij} {dožž+'a} {n,'iščij} {*r'ezči}.

If these are indeed the appropriate representations, the correct phonetic consequences must result from the application of the relevant phonological rules given in Chapter II, sec. 3. These rules operate as follows:

Rule P 1a turns {*r'ezči} into [r,'ezči].

Rule P 1b specifies {č} in [r,'ezči], {s=č'a*st,ju} and {n,'iščij} as voiceless.

Rule P 2 turns {s=ž'og} into [s=ž'ok].

Rule P 3a turns [s=ž'ok] into [z=ž'ok]; [r,'ezči] into [r,'esči]; {n,'izšij} into [n,'isšij].

Rule P 3b turns {s=š'it,} into [š=š'it,]; [z=ž'ok] into [ž=ž'ok]; [s=č'a*st,ju] into [š=č'a*st,ju]; [n,'isšij] into [n,'iššij]; and [r,'esči] into [r,'ešči].

[44] As already indicated in sec. 1.6 above, the – (hyphen) is not an element in the phonemic representation of utterances.

Rule P 4 turns {dožž+'a} into [dož,ž,+'a]; [n,'iščij] into [n,'iš,š,ij]; and [r,'ešči] into [r,'eš,š,i].

Rule P 5b specifies {č} in [š=č'a*st,ju] as sharped.

Rule P 6b turns [š=č,'a*st,ju] into [š,=č,'a*st,ju].

Rule P 7a turns [š,=č,'a*st,ju] into [š,=č,'as,t,ju].

Rule P 8 turns [dož,ž,+'a] into [daž,ž,+'a].

Rule P 10c turns [n,'iššij] into [n,'iššij], thereby completing the process of deriving the phonetic consequences – given in a "narrow" transcription at the beginning of this section – from the phonological representation by the application of the phonological rules. This demonstrates that the representations chosen were indeed appropriate.

3.3 *Comments on Individual Distinctive Features.* The practical application of a theory to a large body of data always brings with it more or less minor modifications in the theory. Certain concepts may have to be redefined in a manner differing somewhat from the original theory; special terminology may have to be created, etc. The following section is devoted to a discussion of these modifications as well as of some other points bearing on the use of the distinctive features.

3.31 Vocalic vs. Nonvocalic and Consonantal vs. Nonconsonantal. These two features are treated together because they differ in a number of respects from all other features. Since these are the only features that have no zeros – i.e., they are distinctive for all morphonemes – all identification procedures must begin with them (cf., section 1.55 and Fig. I-1 above). These features divide the morphonemes of Russian into four classes: vowels, liquids, consonants, and the glide {j}. Within three of these four classes, any given morphoneme differs from at least one other morphoneme by a single feature,[45] but except for pairs of which one is {j}, there are no pairs of morphonemes that are distinguished solely by the features vocalic vs. nonvocalic or consonantal vs. nonconsonantal. In other words, unlike all other features, the features vocalic vs. nonvocalic and consonantal vs. nonconsonantal do not generally function as distinctive marks *ceteris paribus.* This appears very clearly in a comparison of the feature bundles of {p} and {t} with those of {t} and {r}. {p} and {t} share all features except one, namely the feature of gravity, which differentiates them, whereas {t} and {r} share but few features.

The two features also show a high degree of negative correlation. For all but five morphonemes a plus for one feature is paired with a minus for the other feature. From the point of view of coding efficiency – or of Condition (5) of sec. 1.5 above – this is by no means optimal. It might, therefore, be suggested that the features vocalic vs. nonvocalic and consonantal vs. nonconsonantal be replaced by less redundant features. This suggestion overlooks the fact that Condition (5) gives precedence to other requirements over the reduction of distinctive feature statements in a representation. The four classes of morphonemes established by the features vocalic vs. nonvocalic and consonantal vs. nonconsonantal enable us to characterize

[45] The class of glides is somewhat anomalous in this respect because Russian has a single glide {j} and consequently in this class no further distinctions are made.

in a simple manner the phonological constraints on the construction of Russian morphemes, and to state concisely the phonetic changes due to the operation of the morphological and phonological rules of the language. They are, thus, of crucial importance in satisfying Condition (4). Moreover, since the two features are part of the theoretical framework of phonology (cf., sec. 1.2), the suggested changes are in effect a proposal to modify the underlying theory itself – a step to be undertaken only for the most cogent reasons. It would appear, therefore, that the gain to be realized by a more efficient coding hardly outweighs the disadvantages which such a step would entail.

3.32 Compact vs. Noncompact and Diffuse vs. Nondiffuse. In previous analyses of Russian (as well as of other languages) the features compact vs. noncompact and diffuse vs. nondiffuse were formulated so as to constitute a single feature. This, however, destroyed the uniformity of the distinctive feature framework, which otherwise contained only binary features. It also caused difficulties for the application of Condition (5), for there seemed to be no convincing answer to the question of how many binary feature-statements were the equivalent of a ternary feature. The situation was further complicated by the fact that the feature was ternary only for the vowels, and binary for the consonants. It was, therefore, decided to replace the feature compact vs. diffuse by two binary features: compact vs. noncompact and diffuse vs. nondiffuse, of which the latter is distinctive only for the vowels.[46] This decision later gained support from the structuring of the acoustical data: it turned out that the most effective organization of the latter – i.e., the organization with the fewest exceptions – demands two binary features instead of a single ternary one.[47]

3.33 The Primary Tonality Features: Grave vs. Acute and Flat vs. Plain. Tonality features involve the comparison of a high frequency with a low frequency region in the spectrum. In *Preliminaries* three tonality features were described: grave vs. acute (front vs. back), flat vs. plain (rounded vs. unrounded) and sharp vs. plain (palatalized vs. unpalatalized). In order to avoid confusion the feature flat vs. plain is termed in the present study flat vs. natural. The tonalities of these three features are as follows:

High:	Acute	Natural	Sharp
Low:	Grave	Flat	Plain

In the Russian consonants grave vs. acute and sharp vs. plain function distinctively, while the vowels have only a single distinctive feature, flat vs. natural (rounded vs. unrounded). It was suggested in *Preliminaries*, and the suggestion is followed here, that flat vs. natural, the only distinctive tonality feature for the vowels, be considered in complementary distribution with grave vs. acute, the primary tonality feature of the consonants.[48] In the present work, therefore, the features grave vs. acute and flat vs. natural have been included into a single feature: low tonality vs. high tonality. Since,

[46] See Halle, "In Defense of the Number Two," and Chapter V, sec. 4.1 below.
[47] The acoustical data are presented by L. G. Jones in the Excursus "Contextual Variants of the Russian Vowels", pp. 157–183. See also Chapter V, sec. 4.1 below.
[48] *Loc. cit.*, pp. 33-34.

however, the terms "low tonality" and "high tonality" are somewhat awkward, they have not been employed consistently throughout the book. In various places they have been replaced with the phonetically correct terms "flat vs. natural" for the vowels, and "grave vs. acute" for the consonants and the nondistinctive variations in the vowels.

3.34 Strident vs. Mellow. Stridency is distinctive for all noncompact consonants. Continuant consonants except {x} are strident while most stops are mellow. The feature of stridency is, therefore, in complementary distribution with the feature continuant vs. interrupted in the large majority of morphonemes. In the present work stridency has been chosen as distinctive for the noncompact consonants, and the feature continuant vs. interrupted, as distinctive for the compact consonants. This decision was made because it leads to a decrease in the number of distinctive feature statements required. Consider, e.g., the labial consonants of Russian. If stridency is postulated as distinctive, we obtain the following matrix:

	{p}	{p,}	{b}	{b,}	{f}	{f,}	{v}	{v,}	{m}	{m,}
strident vs. mellow	−				+	+	+	+		
nasal vs. nonnasal	−				O	O	O	O	+	+
sharped vs. plain		+		+		+		+		+
voiced vs. voiceless			+	+			+	+	O	O

If instead continuant vs. interrupted is considered distinctive, we obtain the following matrix:

	{p}	{p,}	{b}	{b,}	{f}	{f,}	{v}	{v,}	{m}	{m,}
nasal vs. nonnasal	−		−				−		+	+
continuant vs. interrupted					+	+	+	+	O	O
sharped vs. plain		+		+		+		+		+
voiced vs. voiceless	−		+	+			+	+	O	O

As can be seen, there are 34 distinctive feature statements (pluses and minuses) in the first matrix and 36 distinctive feature statements in the second matrix. By Condition (5) the former analysis is to be preferred.

3.35 In addition to those discussed above, the following features function distinctively in Russian: nasal vs. nonnasal, continuant vs. interrupted, sharp vs. plain, and accented vs. unaccented.

CHAPTER II

SEQUENTIAL CONSTRAINTS

1. *Introduction*

The subject of the following chapter is the limitations which the language places on the occurrence of distinctive feature complexes in the sequence. Ordinarily such restrictions are given in the form of statements describing the occurrence of individual segments or features in specific contexts. It is evident that if the occurrence of a segment or of a feature can be inferred from the context in which it occurs, it is not necessary to specify the segment or feature in the representation. In the framework of the present work, the natural way of handling this situation is by omitting from the representation the feature(s) in question. This satisfies Condition (5), which, as will be recalled, requires that phonological representations contain a minimum of specified features.

The automatic distribution of features is governed by the following three types of rule:

The morpheme structure or MS rules deal exclusively with the feature composition of individual morphemes. Their only function is to assign values to unspecified non-phonemic features.

The morphological rules, which are part of the transformational level, require reference not only to the feature composition of morphemes but also to the morpheme class to which the latter belong. For instance, in Russian the so-called "transitive softening" (*perexodnoe smjagčenie*) plays a role in the conjugation but not in the declension. The decision of whether or not this phonological process applies depends, therefore, materially on factors other than the feature composition of the particular segment.[1] It is for this reason that the morphological rules – i.e., rules of inflection and derivation – have been included among the transformations.[2]

[1] It is clearly impossible to include in the present work a discussion of the morphological rules of Russian, since they are of such complexity as to require an entire book to themselves. Treatments of the morphological rules of Russian can be found in: N. S. Trubetzkoy, *Das morphonologische System der russischen Sprache* (= *Travaux du cercle linguistique de Prague*), 5-2 (1934); R. Jakobson, "Russian Conjugation," *Word*, *4*, 155-167 (1948); H. Rubenstein, *A Comparative Study of Morphophonemic Alternants in Standard Serbo-Croatian, Czech and Russian* (Ann Arbor, 1950); H. L. Klagstad, *Vowel Zero Alternations in Contemporary Standard Russian* (Ph.D. Dissertation, Harvard, 1954); E. Stankiewicz, *Declension and Gradation of Substantives in Contemporary Standard Russian* (Ph.D. Dissertation, Harvard, 1955).

[2] See Chapter I, secs. 1.4, 1.41, 1.42.

The phonological or P rules assign features on the basis of purely phonological criteria; they require reference only to features and phonological boundaries. They differ from the MS rules in that they may reassign values to specified features in addition to specifying nonphonemic features.

2. *The Morpheme Structure (MS) Rules*

As noted in Chapter I, individual morphemes are represented in the dictionary as sequences of segments. In conformity with Condition (5), the representation must be such that the number of distinctive feature statements used for the specification of the morpheme is the smallest possible. In terms of the conventions adopted in the present study, this means that a feature must be left unspecified whenever it can be inferred from the context. The purpose of the MS rules is to specify some of the unspecified features. In order to give these rules in the simplest form possible, it is necessary to establish a partial order in the specification of the features. Thus, for instance, the regularities in the distribution of the feature compact-noncompact in a vowel sequence differ from those in a consonant sequence. Hence, the features establishing the differences between vowels and consonants – i.e., the features vocalic vs. nonvocalic and consonantal vs. nonconsonantal – must be specified before these regularities can be stated. There is, however, no complete ordering of the rules. This partial ordering of the rules will be reflected in their numbering. Each rule will be denoted by a number and a letter, of which only the number refers to the order in which the rule must be applied.

The MS rules apply to morphemes in their most explicit form, i.e., in the form from which all other forms of the same morpheme can be derived in the simplest fashion. Morphemes which have forms with and without vowels will contain a special symbol (#) to indicate that this vowel segment may be dropped as a consequence of certain morphological rules.

A sequence of segments composed entirely of vowels is called a *chain.* A sequence of segments containing no vowels (or #) is called a *cluster.* In certain formulaic representations, V will stand for any vowel; R, for any liquid; C, for any consonant; and J, for the glide {*j*}.

2.1 *Vocalic vs. Nonvocalic and Consonantal vs. Nonconsonantal.* The skeleton of the Russian morpheme is formed by alternations of vowels and nonvowels in the sequence. These alternations are reflected in the distribution of the features vocalic vs. nonvocalic and consonantal vs. nonconsonantal. Constraints on the composition of the sequence hold only within chains or within clusters: the composition of a chain or of a cluster is not affected by the composition of other chains or clusters.

2.11 Although severely restricted in number and variety, vowel chains are admitted in Russian morphemes. E.g., {pa'uk} "spider", {ka'ur} "pertaining to a chestnut-colored horse", {kl,'auz+a} "intrigue", {t,i'un} "feudal governor", {apl,i 'ux+a} "slap".

2.12 The longest clusters in Russian morphemes consist of four segments. There are only two such clusters in the language, one initial, in the morpheme {*f*s*t*r'et,i} "to encounter, meet", and one final, in the morpheme {č'o*r*s*tv} "stale"

2.13 The following three-segment cluster types are attested:

	Initial	Medial	Final
CCR	{*skr,*ip} "squeak"	{k*a*s*tr,'ul,} "sauce pan"	{z'a*txl} "musty"
CCC	{*s*tv*ol} "gun barrel"	not attested	{'o*p*šč} "common"
RCR	not attested	{v,*i*r*bl,'ud} "camel"	not attested
RCC	not attested	not attested	{t*ol*st} "fat"

2.14 The following two-segment cluster types are possible:

	Initial	Medial	Final
CC	{*s*v*et} "light"	{'a*sp,*id} "slate"	{k*o*st,} "bone"
RC	{*rt'ut,} "mercury"	{*alm'az} "diamond"	{*s*m'e*rč} "cedar"
JC	not attested	{b*ajb'ak} "marmot"	{*ajv} "quince"
CR	{*s*l*ep} "blind"	{*u*tr'ob} "womb"	{ž*e*zl} "staff"
RR	not attested	{j*A*rl'ik} "label"[2a]	{g*o*rl} "throat"
JR	not attested	not attested	not attested[3]
CJ	{*dj*ak} "clerk"	{*ab,*i*zj'an} "monkey"	not attested[3]
RJ	{*rj*an} "zealous"	{b*u*rj'an} "tall weeds"	not attested[3]
JJ	not attested since geminates are admitted only of {*n}.		

2.15 The constraints just surveyed have to be taken into consideration in representing the individual morphemes in the dictionary since they make it possible to leave unspecified various features in the representation of certain morphemes, thereby satisfying Condition (5). It is not possible to give a *complete* procedure for discovering the most economical representation in every case. The best that can be done is to formulate the sequential constraints as rules specifying certain contexts. The representation of every morpheme then has to be chosen in such a way as to take maximum advantage of these rules, while at the same time leading to the correct phonetic consequences. Condition (5), however, provides a means for choosing between alternative representations in those cases where several representations are possible.

To gain some insight into why no procedure for discovering the individual representations can be given, consider the following example.

In Russian a glide in pre-final position followed by a consonantal segment must be preceded by a vowel. (Cf., rule MS 5d below.) It might seem, therefore, that in this position the features vocalic vs. nonvocalic and consonantal vs. nonconsonantal do not have to be specified in the vowel. This, however, is not true, e.g., in representing the morpheme {'ajv} "quince", for if the two features are not specified, it is impossible to distinguish the morpheme under discussion from one beginning with a glide, since

[2a] *A stands for a further unspecified diffuse vowel.

[3] All attested sequences of this type contain #; e.g., {*st'at,#j} "article".

unspecified features can not serve to distinguish morphemes from each other. In that case it would be necessary to apply rule MS 1a, which turns the second segment of the morpheme into a vowel. This, however, is in open conflict with the phonetic facts.

2.16 The following rules assign the proper values to all segments in which the features vocalic vs. nonvocalic and consonantal vs. nonconsonantal are unspecified.

2.161 The following rules hold in segment sequences immediately following the & marker (morpheme initial).

Rule MS 1a. If the segment following the & marker is a glide, the next segment is vocalic and nonconsonantal; i.e. a vowel.

Rule MS 1b. If the segment following the & marker is a liquid and the next segment is vocalic, the latter segment is also nonconsonantal.

Rule MS 1c. The segment after &RC is vocalic and nonconsonantal.

Rule MS 1d. If in an initial cluster the first two segments are consonants and the third segment is nonvocalic, the third segment is also consonantal.

Rule MS 1e. If in an initial cluster a consonantal segment precedes a glide, the glide is followed by a segment which is vocalic and nonconsonantal.

Rule MS 2. If in an initial sequence the first three segments are consonants, the fourth segment is vocalic.

Rule MS 3. If in an initial cluster a liquid is not in an initial position, the liquid is followed by a segment which is vocalic and nonconsonantal.

Rule MS 4. If the morpheme contains no vowel or if the vowel terminating the initial cluster is followed by a & sign rules MS 5 to 10 do not apply.

2.162 The following rules hold for segment sequences immediately preceding the & marker (morpheme final).

Rule MS 5a. If the last segment before the & marker is a glide, the penultimate segment is vocalic and nonconsonantal.

Rule MS 5b. If the last segment is a liquid and the penultimate segment is nonvocalic, the latter is also consonantal.

Rule MS 5c. If the last two segments are consonants and the antepenultimate segment is nonvocalic, the antepenultimate segment is also consonantal.

Rule MS 5d. If the last segment is a consonant and the penultimate segment is a glide, the glide is preceded by a segment which is vocalic and nonconsonantal.

Rule MS 6a. If the segment preceding the sequence CR& is vocalic, the segment is also nonconsonantal.

Rule MS 6b. If the segment preceding the sequence CR& is nonvocalic, it is also consonantal.

Rule MS 6c. The segment preceding the sequence CCC& is vocalic.

Rule MS 7a. If in a final cluster, the liquid is not in final position, the segment preceding the liquid, is vocalic and nonconsonantal.

Rule MS 7b. The segment preceding the sequence CCR& is vocalic and nonconsonantal.

Rule MS 8. If the vowel preceding the final cluster is also the vowel terminating the initial cluster or is the initial segment in the morpheme, rules MS 9 and 10 do not apply.

2.163 The MS rules for medial clusters have to be applied after the MS rules for initial and final sequences. In applying the following rules it is necessary to begin with the vowel adjacent to the initial cluster and work towards the next vowel. The same rules are then re-applied beginning with the second vowel, etc. The application of these rules ceases when the final vowel of the morpheme is reached.

Rule MS 9a. If the sequence VJ is followed by a nonvocalic segment, the latter is also consonantal.

Rule MS 9b. The segment following the sequence VCC is vocalic.

Rule MS 9c. The segment following the sequences VCR, VCJ, VRR, or VRJ is a vowel.

Rule MS 9d. The sequence VRC is followed by a vocalic segment.

Rule MS 10a. If the sequence VJ is followed by a consonantal segment, the next segment is vocalic and nonconsonantal.

Rule MS 10b. The sequence VCCR is followed by a vowel.

Rule MS 10c. The sequence VRCR is followed by a vowel.

2.17 Examples.

In the morphemes {p,'o*s*tr} "variegated", {j'a*s*t*reb} "falcon" and {gr'u*st,} "sorrow" the features vocalic vs. nonvocalic and consonantal vs. nonconsonantal are represented as follows:

$$\begin{array}{l}\text{vocalic}\\ \text{consonantal}\end{array} \quad \&\begin{Bmatrix} - & + & - & - & + \\ + & - & \bigcirc & \bigcirc & + \end{Bmatrix}\& \quad \&\begin{Bmatrix} - & \bigcirc & & - & \bigcirc & + & - \\ - & \bigcirc & + & + & + & - & + \end{Bmatrix}\& \quad \&\begin{Bmatrix} - & + & \bigcirc & - & - \\ + & + & \bigcirc & + & + \end{Bmatrix}\&$$

By applying rules MS 5b, 6b and 8 to the first morpheme, MS 1a and 9b to the second morpheme, and MS 3 and 8 to the third morpheme, the unspecified segments are fully specified as shown here:

$$\begin{array}{l}\text{vocalic}\\ \text{consonantal}\end{array} \quad \&\begin{Bmatrix} - & + & - & - & + \\ + & - & + & + & + \end{Bmatrix}\& \quad \&\begin{Bmatrix} \cdot & + & - & - & + & + & - \\ - & - & + & + & + & - & + \end{Bmatrix}\& \quad \&\begin{Bmatrix} - & + & + & - & - \\ + & + & - & + & + \end{Bmatrix}\&$$

2.2 *Compact vs. Noncompact, Diffuse vs. Nondiffuse, Low vs. High Tonality.* These features are not distinctive for the liquids and the glide.

2.21 Most regularities connected with the distribution of these features in the vowels are subject to morphological rules – e.g., vowel alternation in different stems – and will not be treated here. Within individual morphemes the only constraint affects vowel sequences, which always consist of {*u}, preceded by either {*a} or {*i}.[4] Hence, if a sequence of segments is known to consist of two vowels, the features remaining to be specified are diffuse vs. nondiffuse in the first segment and accented vs. unaccented in the accented segment. All other features are assigned by the following rule:

[4] Examples are in sec. 2.11.

Rule MS 11a. In vowel sequences, the second segment is diffuse and of low tonality; the first segment is either diffuse and of high tonality, or nondiffuse and compact.

2.22 The occurrence of the above features in the consonants is subject to a number of constraints.[5]

Compact consonants are not admitted after grave compact (velar) consonants.[6] Except after noninitial {*m} there can be no sequences of grave noncompact (labial) consonants.[7]

2.23 These constraints are reflected in the following rules:

Rule MS 11b. In position after a grave noncompact (labial) consonant other than noninitial {*m}, a consonant is compact.

Rule MS 11c. A consonant in position after a grave compact (velar) consonant is noncompact.

2.24 To illustrate the operation of these rules we have chosen the following sequences (X stands for compact grave consonants, S, for noncompact acute consonants, F, for noncompact grave consonants other than noninitial {*m}):

	F	S	X	S	X	F
compact vs. noncompact			+	O	+	O
grave vs. acute	+	O	+		+	+

Then by Rules MS 11b and MS 11c the correct values are assigned to the unspecified features:

	F	S	X	S	X	F
compact vs. noncompact			+		+	
grave vs. acute	+		+		+	+

2.3 *Strident vs. Mellow.* In sequences of nonnasal acute noncompact (dental) consonants, strident continuant consonants precede nonstrident consonants; i.e., {*st*} and {*zd*} are admitted, but not {*ts*} or {*tc*}. Hence in representing {*st*} and {*zd*} there is no need to specify the feature of stridency. It is given by the following rule:

Rule MS 11d. In sequences of nonnasal acute noncompact (dental) consonants, the first segment is strident and continuant and the second segment is nonstrident.

2.4 *Nasal vs. Nonnasal.* There are no restrictions on the distribution of the feature of nasal vs. nonnasal.

2.5 *Continuant vs. Interrupted.* The sequence {*lr*} does not occur, but {*rl*} does

To simplify the terminology when speaking of consonants, the terms "low tonality" and "high tonality" will be replaced by "grave" and "acute" respectively. See Chapter I, sec. 3.33.

6 All counter-examples are either of foreign origin – e.g., {v'a*kx} "Bacchus", {ba*kš'iš} "backsheesh" – or occur on morpheme junctions – e.g. {*v'ek-š+a} "squirrel", cf., M. Vasmer, *Russisches etymologisches Wörterbuch* (Heidelberg, 1953).

7 Examples like the gen.sg. {l,ubv,+'i} "love" do not contradict this observation, since the lexical morpheme is represented as {*l,ub # v,*} in view of the nom.sg. {l,ub'ov,}.

exist. Hence the feature of continuant vs. interrupted need not be specified when representing sequences of liquids. It is given by the following rule:

Rule MS 11e. In sequences consisting of two liquids the first segment is interrupted and the second segment is continuant.

2.51. As a consequence of Rule MS 11d (sec. 2.3) it is also unnecessary to specify the feature continuant vs. interrupted in sequences consisting of nonnasal acute noncompact (dental) consonants.

2.6 *Voiced vs. Voiceless.* Within a morpheme the feature of voicing is distinctive before a sonorant – i.e., before a vowel, a liquid, a nasal consonant, and the glide – before one or more consecutive {**v*} followed by a sonorant, and at the end of the morpheme. In all other contexts the feature of voicing is governed by rules which hold also across morpheme junctions, and which, therefore, must be included among the P rules. Since nothing is to be gained by setting up the same rules twice, once as part of the MS rules and again as part of the P rules, the feature of voicing in these contexts will be left unspecified, until the P rules are applied. It is important to note that this decision does not affect the possibility of transforming the distinctive feature table into a branching diagram.[8]

In representing morphemes the feature of voicing need not be specified in position before an obstruent, except before one or more consecutive {**v*} followed by a sonorant. In these positions, the correct assignment of the voicing feature is given by the P rules 1b, 2 and 3a, *q.v.*

2.7 *Sharped vs. Plain* (*Palatalized vs. Unpalatalized*). The sharping feature is distinctive for the liquids, the noncompact consonants except {*c*}, and moreover marginally for {**k*}. As in the case of voicing, the constraints on sharping within morphemes can best be handled as part of the P rules. In the following sections these constraints will be surveyed and indications will be given of how these constraints are to be incorporated in the representation of individual morphemes.

2.71 The continuant liquid {**l*} has the smallest number of constraints. Sharping is predictable in this segment before {**e*} by the same rule which governs its distribution across morpheme junctions, and which is, therefore, included in the P rules. In representing morphemes in which {**l*} is followed by {**e*}, sharping is left unspecified; e.g., {**l***es*} "forest"

2.72 In the liquid {**r*}, and in noncompact consonants, except {*c*}, the feature of sharping is distinctive – and must therefore be specified – only before vowels (including #) other than {**e*} and at the end of lexical morphemes and nonfinal suffixes. In all other contexts sharping is distributed in accordance with the rules which hold also across morpheme junctions, and which, therefore, are given among the P rules.

2.721 At the end of prefixes and unaccented prepositions and the end of inflectional suffixes noncompact consonants are always plain. In these morphemes it is, therefore, not necessary to specify the feature of sharping.

[8] Cf., Chapter I, sec. 1.53.

2.722 Historical digression: The above accounts for the loss of sharping in certain word-final consonants which formerly preceded a front jer. Sharping was lost, thus, in the instr. sg. endings of nonfeminine nouns – e.g., {rab+'om} "slave" from Old Russian *rabъmъ* – and in the 1. sg. present tense ending of verbs like {d'am} 'I give" (perf.), because in word-final suffixes ending with a consonant sharping is not admitted. The retention of sharping in {j'e*st,} "there is" and at the end of accented prepositions is due to their being single lexical morphemes that are further unanalyzable and are hence not subject to the restrictions of sec. 2.721.[9] Sharping is also retained in the infinitive suffix since this suffix ends not with the consonant {t,}, but rather with {t,#}, as is shown by such forms as {v,oz-t,'i} "to convey".

2.73 Sharping in compact consonants is distinctive only for {**k*} before {**a*}, {**o*} and {**u*}. Before the other vowels, {**k*} as well as other compact grave (velar) consonants is sharped, whereas in position not before vowels and at the end of the word, all compact grave (velar) consonants are grave. In representing morphemes containing {**k*}, it is, therefore, necessary to specify the feature of sharping only in position before {**a*}, {**o*} and {**u*}; e.g., {*tk,'om} "we weave". In all other contexts, sharping is distributed in accordance with rules which hold also across morpheme junction and which are part of the P rules.

2.8 *Accented vs. Unaccented.* This feature, which is distinctive only for the vowels, is governed in large measure by the morphological rules of the language. As already noted morphemes having a "movable" accent are represented with vowels in which the feature accented-unaccented is left unspecified. This feature is then specified by the morphological rules of the language. Since Russian polysyllabic words must contain exactly one accented vowel, the assignment of the accent to a particular vowel determines also that all other vowels in the word will be unaccented.

2.9 *Gemination.* Within a morpheme, clusters consisting of two or more identical segments are not admitted, with the sole exception of the cluster {**n***n*}. Since voicing and sharping in clusters is governed by various phonological rules and hence is not distinctive, consecutive segments must differ in at least one of the other features.

3. The Phonological (P) Rules

The function of the phonological rules is to complete the process of specifying the sentence. The P rules, which are applied after all morphological rules, operate on representations which consist entirely of morphonemes (both fully as well as incompletely specified) and of boundaries.

The P rules have the following results:

a) They complete the specifications of all incompletely specified morphonemes; i.e. like the MS rules, they replace zeros with pluses or minuses.

[9] See R. Jakobson, "Die Verteilung der stimmhaften und stimmlosen Geräuschlaute im Russischen," *Festschrift für Max Vasmer* (Berlin, 1956), pp. 199–202.

b) They specify all features which play no distinctive role in the language but are not randomly distributed. This, too, can be viewed as assigning pluses or minuses to zeros.

c) They change the values assigned to certain features in particular contexts.

As already noted, the P rules are of the form "rewrite x as y", without reference to the "derivational history" of the symbols. Thus, if in consequence of the application of some P rule a segment acquires a particular feature, the segment thereafter is subject to exactly the same rules as all other segments possessing that feature. This simple form of the rules can be maintained only so long as there is a partial ordering among the rules. The system of numbering used for the MS rules will also be used for the P rules: each rule is assigned a number and a letter, of which only the number reflects its order of application. E.g., rules 6a, 6b and 6c can be applied in any order, so long as they are applied after rules numbered 1 through 5, and before rules numbered 7 and higher.

There are several rules for which no unique place in the hierarchy can be defined: they have to be applied before a certain rule, but not *after* any particular rule. The convention has been adopted to apply such rules immediately before the rule that requires them. Thus, e.g., if a particular rule has to be applied before P rule 75, it will be assigned the number 74.

Rule P 1a.[10] Within a phonological word – i.e., not across phonemic phrases or word boundaries – liquids and noncompact consonants, except {c}, are sharped before (accented or unaccented) {*e}.

E.g. {*les+'a} turns into {l,es+'a}.

The following three rules govern the distribution of the feature of voicing.[11] The domain in which the voicing rules hold is delimited by phonemic phrase boundaries and/or the – boundaries in abbreviations.[12] In stating the rules, other boundaries are, therefore, disregarded unless otherwise indicated. Thus, for instance, in the statement of the voicing rules the sequence {AB%C=DE} will not be normally distinguished from the sequence {ABCDE}.

In order to state the voicing rules in a simple fashion, it is necessary to set up the following classes:

Sonorants; i.e., vowels, liquids, the glide, and the nasal consonants.

Obstruents; i.e., all other morphonemes except {*v}.

{*v} functions as a sonorant if followed by a sonorant, and as an obstruent if followed by an obstruent.

Rule P 1b. Unless followed by an obstruent, {c}, {č} and {x} are voiceless.

The effect of this rule is to turn the three obstruents, which up to this point have remained unspecified with respect to voicing, into voiceless obstruents. Note, how-

Rule P 1a does not apply in certain words of foreign origin. These cases are discussed in sec. 3.1.
Cf., Jakobson, "Die Verteilung der stimmhaften und stimmlosen Geräuschlaute im Russischen."
For the treatment of voicing in abbreviations, see also p. 65 below.

ever, that the rule applies only in the stated contexts; elsewhere the three obstruents remain unspecified.

A single obstruent or several consecutive obstruents occurring in sequence regardless of intervening preposition or word boundaries will be termed an *obstruent cluster*. This definition makes it possible to state simply the distribution of the voicing feature in all remaining contexts where voicing is nondistinctive.

Rule P 2. If an obstruent cluster is followed by a word boundary or by a phonemic phrase boundary, all segments in the cluster are voiceless.

Rule P 3a. If an obstruent cluster is followed by a – (dash) boundary or by a sonorant, then with regard to voicing the cluster conforms to the last segment; if it is voiced, so are all other segments in the cluster; if it is voiceless, so is the entire cluster.

Examples: Within a word: before voiced obstruents: {*pr'os,b+a} [pr'oz,bə] "request"; {ot=*br'os,it,} [adbr'os,it,] "to cast off" (cf. {ot=vl,'eč} [atvl,'eč] "to divert" with a voiceless [t] before {v} followed by a sonorant); {ot=*vdov'i} [advdav'ɨ] "from the widow", {k=*vdov'e} [gvdav,'ɛ] "to the widow" (cf., [ka^{t}c'u][13] "to the father"). Before voiceless obstruents [f,'e^{t}kə] "hypochoristic form of the name Theodor" (cf., another hypochoristic of same name {f,'ed,a}); {*bez=t'om+a} [b,ist'omə] "without the volume" (cf., [b,izr'omə] "without the rum"). Across word boundaries: {gr,'ob d'olgo} [gr,'o^{b}d'olgə] "he rowed for a long time" (but [gr,'opm,'ed,l,inə] "he rowed slowly"). It is to be noted that the rule holds across word boundaries preceding enclitics – e.g., [gr,'obžɨ] "but he did row", but [gr,'opl,i] "did he row" – and following autonomous prepositions e.g., {*bl,'iz, dor'ogi} [bl,'iz,dar'og,i] "near the road", but [bl, 'is,akn'a] "near the window" and [bl,'is, c'erkv,i] "near the church".[14] Note that {v} and {v,} play no independent role. Everything transpires as if {v} or {v,} had been absent; e.g., {m'og vo=jt,'i} [m'ok vajt,'i] "he could enter", [m'okv,irn'ut,] "he could return", but [m'ogvzdaxn'ut,] "he could sigh".[15]

At the end of a phonemic phrase, where a pause is admissible in the utterance, only voiceless obstruents can occur. For this reason in a slow, solemn style of speech, where there is a tendency to treat every accented word as a separate phonemic phrase and hence to pause between words, there is also a tendency to unvoice all obstruents before the word boundary. Before enclitics and after proclitics, and in sequences of words in close contact, like {kn,'az, bor,'is} "Prince Boris", where a phonemic phrase boundary is not admitted, this phenomenon does not occur.

A raised letter indicates that the stop is imploded; see rule P 10a.

Prepositions like {č'er,iz,} "through" which also possess unaccented free variants – i.e. {čir,iz=} – provide an interesting test case for the rules as set up. One can say either [č'er,is, akn'o] "through the window" or [čir,iz – akn'o]. Cf. 2.721 above.

[15] In a review of W. Steinitz, *Russische Lautlehre*, A. V. Isačenko asserts that the "pronunciation of the sequence *rež' bulku* with [žb] is ... unusual. Before [b] there appears a sound whose beginning is voiceless, but which towards its end may become voiced as a result of the excursions (Exkursionstätigkeit) of the vocal cords that are prepared for the phonation of [b]." *Zeitschrift für Phonetik*, *8*, 415 (1955). My own observations include cases like the one described by Isačenko, but these appear to be no more common than cases with entirely voiced clusters. See also Trubetzkoy, *op. cit.*, p. 21.

The feature of voicing follows rules of its own in the nonfinal components of abbreviations like {pr'of–*de*leg'at} "trade union representative". These *ad hoc* formations enjoyed great popularity in Russian in the period between the two World Wars. At present their popularity is on the wane, but the device is still productive.

Many formations of this type have been absorbed into the language as common words. In the process they have also taken on the form of ordinary Russian words preserving the accent only on a single vowel, and subjecting other phonemes to the same phonological rules as normal words; e.g., {s*ovx'oz} [safx'os] "state farm" In a synchronic description there is no reason to differentiate these from other words.

In many other abbreviations the components are treated rather differently. Such abbreviations can contain more than one accented vowel, and hence are similar to such composita as {*s'e*vera z'apad} "north west". The feature of voicing, however, is treated differently than before a word boundary. For this reason the – boundary was introduced. E.g., {p'art–b,i*l'et} "party card"; {m'os–g'as} "Moscow Gas Works"; {s'ov–p'art–šk'ol+a} "Soviet Party School" – all of which contain sequences of voiced and voiceless obstruents, which are not admitted anywhere else.

Unless specifically noted, the following rules hold only within the phonological word; i.e., not across word boundaries.

Rule P 3b. Before acute compact (palatal) consonants, strident acute noncompact (dental) consonants become compact (palatal).

E.g., {*b*ez=č'e*st,ij+o} [b,iščʼ,'es,t,ijə] "dishonor", {*b*ez=ž'alo*st,n+o} [b,ižž'aləsnə] "pitiless", {s=š'um+om} [šš'uməm] "with noise".[16]

The neutralization does not take place across word boundaries, including those preceding enclitics. In normally careful speech the consonant sequence in {v,'oz ži} "but he drove" is distinguished from that in {r'ož ži} "but (it is) rye"; in rapid speech, however, these clusters become homophonous. In a very solemn, rhetorical style, the assimilation of compactness does not always take place across = boundaries between prepositions and the following; e.g., {*bez=ž'alo*st,i} "without pity" is pronounced with [zž], although [žž] is the more common pronunciation.[17]

Rule P 4. If in the sequence {šč} and {žž} no $\dot{=}$ boundary intervenes, both segments are sharped (palatalized) and the {č} becomes a continuant.[18]

E.g., {b'o*ršč} [b'or,š,š,] "beet soup", {d*ožž+'a} [daž,ž,'a] "rain" (gen. sg.), {b*o*rščk+'u} [bar,š,š,k'u] "some beet soup", {*r'ezče} [r,'eš,š,i] "harsher". (In the last example rules 1a, 1b, 3a and 3b were applied before rule 4.)

Rule P 5a. All strident consonants except {c} are continuant. All mellow nonnasal consonants are interrupted (stops); {*k} and {g} are mellow.

[16] Cf., R. Avanesov and S. Ožegov, *Russkoe literaturnoe udarenie i proiznošenie – Opyt slovarja-spravočnika* (Moscow, 1955), pp. 562–563.

[17] V. Vinogradov *et al.*, eds., *Grammatika russkogo jazyka* (Moscow, 1952), p. 80.

[18] Rule P 4 is one of the characteristic traits of the old Moscow standard. It is not mandatory in the present norm. Cf., Vinogradov *et al.*, eds., *op. cit.*, secs. 146 and 148.

Rule P 5b. Except for the cases mentioned in rule P 4, the feature of sharping is assigned to the following segments, which up to this point have remained unspecified with respect to this feature: {č} and {j} are sharped; {c}, {š} and {ž} are plain.[19]

The following rules describe the nondistinctive functioning of sharping (palatalization). Russian admits considerable latitude in nondistinctive sharping, particularly across morpheme boundaries in complex words. Sharping, moreover, presents serious difficulties to the observer – especially difficult are stops – with the result that on occasion even the most reliable phoneticians disagree as to the evidence.[20] The rules are, therefore, tentative, subject to, perhaps radical, revision in the light of future research.

In stating the rules for nondistinctive sharping, it is convenient to have a special term to refer to the liquids and the noncompact consonants except {c}, all of which have distinctively sharped and plain cognates. Hereinafter these morphonemes shall be called *paired morphonemes.*

Rule P 6a. Before the glide {j}, paired morphonemes are normally sharped.

Examples: In simple words: [b,j'u] "I beat", [v,j'užnij] "stormy", [kr'ov,ju] "blood" (instr. sg.), [pap'ad,ji] "the pop's wife" (gen. sg.). In complex words: [v,=j'užnij] "to a Southern", [s,=j'est] "congress", [at,=j'est] "departure", [pad,=j'olkəj] "under a Christmas tree", [ab,=j'om] "volume"

Rule P 6b. Before acute compact (i.e. palatal) consonants which are (non-distinctively) sharped, acute consonants are sharped.

It is to be recalled that by operation of rule P 3b, strident noncompact acute (dental) consonants become compact (palatal) in this context. For these consonants we, therefore, get the following: {s=č'a*st,ju} [š,č,'as,t,ju] "with a part", {*b*ez=šč'a*st,ij+a} [b,iš,š,'as,t,ijə] "without luck". (Sequences of more than two identical segments are pronounced the same as two.)

In this position all mellow acute consonants are sharped; e.g. {n}: {k'ončit,} [k'on,č,it,] "to end", {bu*b'enčik} [bub,'en,č,ik] "little bell", {ž'enščin+a} [ž'en,š,š,inə] "woman"; mellow nonnasal consonants: {sa*v'etčik} [səv,e^{t,}čik] "one who gives unsolicited advice", {ot=č'a*st,i} [a^{t,}č,'as,t,i] "partially", {ot=ščip,'it,} [a^{t,}š,š,ip,'it,] "to chip off".

Before plain {š} and {ž}, to which rule P 6b does not apply, the feature of sharping is distinctive for {*n}; e.g., {*den,ž'onki} "money" (diminutive), {t'on,še} "thinner" as opposed to {xanž'oj} "hypocrite" (instr. sg.) and {bar'onše} "baronness" (dat. sg.).

[19] It is to be noted that compact grave (velar) consonants still remain unspecified with respect to sharping. See below, rule P 6d.

[20] Cf., e.g., the following remark by a leading modern phonetician: "The degree of softening of consonants before soft consonants can vary greatly: in addition to soft and hard consonants, consonants can be pronounced with varying degrees of softness; they are not entirely hard nor entirely soft ("semi-hard" or "semi-soft") consonants, with regard to which it is difficult to decide whether they belong to the hard or to the soft consonants." R. I. Avanesov, *Russkoe literaturnoe proiznošenie* (Moscow, 1954), pp. 79–80.

In this position all other acute consonants are plain; e.g., {n,'izšij} [n,'iššij] "lowest", {pod=š'a*šk+oj} [paˈš'aškəj] "under the sword".

Rule P 6c. Before plain acute noncompact (dental) consonants and before plain liquids, {*r} and noncompact (labial and dental) consonants are plain.

This rule accounts for various "morphophonemic" alternations between sharped and plain consonants and avoids the necessity of giving multiple representations for the morphemes in question. Thus, the representation of the infinitive of the verb {m'it,} "wash" can also figure in the reflexive form {m'it,sa}, although in the former instance the final consonant is sharped, while in the latter instance, it is plain.

Examples. Before plain acute noncompact (dental) consonants: {*d*v'er,c+a} [d,v,'ɛrcə] "little door" – cf. {*d*v'er,} "door"; {p,s+a} [ps'a] "dog" (affective) (gen. sg.) – cf. nom. sg. {p,'os}; {g*um,n+'o} [gumn'o] "threshing floor" – cf., gen. pl. {g*um,'on}; {m'it,sa} [m'iˈcə] "wash" (refl.)[21] – cf., nonreflexive {m'it,}; (cf., also the homonyms {boj'atsa} "fear" (3. pl. pres.) and {boj'at,sa} "fear" (inf.)); {čud,n'a} [čudn'a] "strange" (fem.) – cf. {čud,'on} "strange" (masc.).

Before {l}: {or,l+'a} [arl'a] "eagle" (gen. sg.) – cf., nom. sg. {or,'ol}; {koz,l+'a} [kazl'a] "goat" (gen. sg.) – cf., nom. sg. {koz,'ol}; {kab,l+'a} [kabl'a] "stump, stake" (dial.) (gen. sg.) – cf., nom. sg. {kab,'ol}.

Before {r}: {kov,r+'a} [kavr'a] "rug" (gen. sg.) – cf., nom. sg. {kov,'or}; {s,o*st,r+'a} [s,istr'a] "sister" – cf., gen. sg. {s,o*st,'or}.

Rule P 6c may seem to be contradicted by such examples as the imperative sg. [zab'ut,sə] "forget (yourself)!". It has, however, been pointed out by R. Jakobson that the imperative sg. form is obligatorily followed by a word boundary.[22] The above example is, therefore, represented here as {za=b'ud,%sa} with an intervening word boundary, across which rule P 6c does not hold.

Rule P 6d. Before {*i} and {*e}, compact grave (velar) consonants, which up to this point have remained unspecified with respect to sharping, become sharped. E.g., {r,'ige} [r,'ig,i] "threshing barn" (dat. sg.), {n*og'e} [nag,'ɛ] "foot" (dat. sg.).

Since the following rule differs in the present norm from that in the Moscow standard pronunciation, it is given in two versions, of which the first reflects the Moscow standard, and the second, the present literary pronunciation.

Rule P 7a. Before sharped noncompact (labial and dental) consonants and before {l,}, paired acute noncompact (dental) consonants are sharped.

Rule P 7a′. Paired acute noncompact (dental) consonants in position before sharped noncompact (dental and labial) consonants and before {l,}, are subject to the following treatment:[23]

Within a simple word: acute noncompact obstruents tend to be sharped;[24] {n} is

[21] {s} becomes [c] by rule P 9b, below.

[22] R. Jakobson, "Russian Conjugation," *Word*, *4*, 159 (1948).

[23] In this as in most of the discussion of the distribution of the sharping feature, I follow primarily the description of Avanesov, *Russkoe literaturnoe proiznošenie*, pp. 79–97.

[24] Because of this, it is important that rule P 6c be applied before rule P 7a. Consider the example

sharped before sharped acute noncompact (dental) consonants, and plain before sharped grave noncompact (labial) consonants.

In complex words across a = boundary: Acute noncompact (dental) continuants are sharped before sharped acute noncompact (dental) consonants and before {l,}; but are plain before sharped grave noncompact (labial) consonants. Acute noncompact (dental) stops in this position differ from plain stops in other positions by a tendency towards delabialization. Complete sharping, consisting of both delabialization and widening of the pharynx, is relatively infrequent for stops in this position.[25] {n} does not occur in this position, since no preposition or prefix ends with {n}.

Examples (for the present literary pronunciation only): Nonnasal consonants within a simple word before sharped noncompact consonants and before {l,}: [t,v,'ordɨj] "hard", [m,id,v,'et,] "bear", [s,p,'et,] "to mature", [s,m,it'anə] "sour cream", [z,m,ij'a] "snake", [iz,b,'onkə] "little hut", [s,t,in'a] "wall", [s,l,'ɛt] "trace", [s,n,'ɛk] "snow", [z,d,'es,] "here", [z,l,'it,] "to anger", [m,'ed,l,it,] "to tarry", [maz,n,'a] "daub", [p,'at,n,icə] "Friday", and [d,n,'a] "day" (gen. sg.).

{n} in a simple word before sharped noncompact (dental and labial) consonants: [z'on,t,ik] "umbrella", [p,'en,s,ijə] "pension", [pr,it,'en,z,ijə] "grievance", but [kanf,'ɛta] "candy", and [kanv,'ɛrt] "envelope", [sanl,'ivɨj] "sleepy".

In a complex word across a = boundary after a preposition or prefix, the tendency towards palatalization is weaker, but even when palatalization is absent – e.g. in the stops – the consonants are delabialized.

Continuants before sharped acute noncompact (dental) consonants: [s,=t,'or] "wiped off" (masc.), [s,=t,'ot,ij] "with the aunt", [s,=n,'os] "took down" (masc.), [z,=d,'ɛdəm] "with grandfather", [raz,=d,'ɛl] "division", [b,iz,=n,'ix] "without them".

Continuants before sharped grave noncompact (labial) consonants: [s=p,'ɛl] "sang" (masc. perf.), [s=m,'er,il] "measured" (masc. perf.), [iz=b,'it,] "to beat up", [biz=v,'ɛrɨ] "without faith", [č,r,iz=m,'ɛrṅə] "excessive".

Continuants before {l,}: [s,l,'ɛp] "blind" (masc.), [z,l,'it,] "two anger", [s,=l,'it,] "to pour together", [č,ir,iz,=l,'ɛs] "through the forest".

Stops before acute noncompact (dental) consonants: [pat=s,'ɛl] "sat down next to" (masc. sg.), [at=n,iv'o] "from him", [pad=n,'os] "brought up" (masc. sg.).

Stops before grave noncompact (labial) consonants: [at=m,'er,it,] "to measure off", [pad=b,'it,] "to interline".

{s,o*st,r+'a} "sister". If rule P 7a were to be applied first, the {*s} before the {t,} would have to be sharped, and then, by applying rule P 6c twice, turned back into a plain consonant. If, instead, rule P 6c is applied first, the {t,} becomes plain, and there is no further possibility of applying rule P 7a. This is another example of how economy of statement can be obtained by ordering the rules. – The word "tend" appears in the above rule as well as in some others, because at present there are no fixed standards. Cf., e.g., "in cases where in different forms of the same word, the sequence [st] of the suffix *stv* appears not only before [v,] but also before [v], there *may* be no softening of the sequence." Avanesov and Ožegov, *op. cit.*, p. 559 [my italics – M.H.].

[25] Concerning labiovelarization of plain consonants and liquids, see O. Broch, *Slavische Phonetik* (Heidelberg, 1911), pp. 224–230.

Stops before {l,}: [dl,i=v'as] "for you", [tl,'et,] "to decay", [at=l,'ič,nə] "excellent", [nad=l,iž'at,] "to be obligated".

Rule P 7b. Grave (labial and velar) consonants and {r} tend not to be sharped nondistinctively before sharped consonants, with the exception of the following special cases.

Before sharped grave compact (velar) consonants nondistinctive sharping of grave noncompact (labial) consonants is now considered archaic or substandard; e.g., [d'ɛfk,i] "girls" is preferred over [d,'ef,k,i]. {x}, the only compact grave (velar) consonant admissible in this position, is, however, nondistinctively sharped; e.g., [l,'öx,k,ix] "lungs" (gen. pl.).[26]

Before sharped noncompact grave (labial) consonants, nondistinctive sharping of noncompact grave (labial) consonants is optional, although the formerly predominant tendency towards nondistinctive sharping is on the decrease. [bom,b,'it,] "to bomb" seems to be in free variation with [bomb,'it,], and [l'am,p,i] "lamp" (dat. sg.) with [l'amp,i].

Before {l,} and before sharped acute (dental and palatal) consonants, grave consonants are always plain (unpalatalized). E.g., [l'okt,i] "elbows", [l,'ɛxč,i] "easier", [l'apt,i] "bast shoes", [kr,'ɛpč,i] "stronger", [z,'ɛml,i] "lands"

Nondistinctive sharping of {r} in position before sharped consonants was the rule in the old Moscow standard. At present this is considered substandard in most instances. There are, however, a number of words in which sharping appears to be obligatory. E.g., [p,'er,m,] "Perm", [p,'er,v,in,ic] "first born", [b'or,š,š,] "beet soup", etc.[27] It has been suggested by Boyanus that nondistinctive sharping of {r} takes place in "a stressed syllable which has a soft consonant, a **j** or a **š**, **ž** preceding its vowel".[28] This rule is, however, contradicted by examples like [b'or,š,š,] "beet soup", which is given by Avanesov as the prevailing norm.[29]

Rule P 7c.[30] If in a simple word a plain consonant or liquid precedes its sharped cognate (gemination), the former is also sharped; e.g., [v'an,n,i] "bath" (dat. sg.). In a complex word, across a = boundary, this regressive assimilation of sharping is not mandatory; it is, however, preferred; e.g., {v=v,o*st,'i} "to introduce" is pronounced [v,v,is,t,'i] as well as [vv,is,t,'i].

Rule P 7d. Before {*n} the acute noncompact (dental) stop is dropped in position after an acute noncompact (dental) continuant.

Examples: {*l'e*st,nij} [l,'ɛsnij] "flattering" - cf., {*l'e*st,} "flattery". (Note that

[26] Cf., Avanesov, *Russkoe literaturnoe proiznošenie*, p. 95.

[27] *Ibid.*, pp. 90–92. Examples quoted below are from this work unless specially noted.

[28] S. Boyanus, *A Manual of Russian Pronunciation* (London, 1946), pp. 132–134.

[29] It seems that the facts are not clear even to the investigators most immediately concerned with this problem. Thus, for instance, Avanesov and Ožegov's dictionary, published only a year after the monograph of Avanesov, from which most of the above examples were taken, differs from the latter in a number of crucial examples; see, e.g., their entries for *boršč* and *pervenec*, where there is no indication of sharping for {*r}, although these two words are specifically cited in the monograph as instances of nondistinctive sharping.

[30] Cf., sec. 2.9 above.

rule PR 6c turns the sharped {t,} into a plain consonant; it must, therefore, precede rule PR 7d.) {na=j'e*zd,n,ik} [naj'ez,n,ik] "rider" – {j'e*zd,it,} "to ride". {*sv,'i*stnut,} [s,v,'isnut,] "to whistle" – cf., {*sv,'ist} "whistle".

Rule P 7e. {*a} is of high tonality (i.e., unrounded).

Rule P 7f. With the exception of certain verbal forms[31] grave compact (velar) consonants not followed by {*i} or {*e} are plain.

The following rules govern the phonetic implementation of the unaccented vowels (*ikan'e* and *akan'e*). The present literary norm differs from the Moscow standard in its treatment of unaccented vowels after compact acute (palatal) consonants. Whereas the Moscow standard requires that {a} become [i] after all compact acute (palatal) consonants, the present literary norm requires this change only after sharped compact acute (palatal) consonants. Hence {žar'+a} "heat" and {šag'at,} "to step" are pronounced [žir'a] and [šig'at,] in the Moscow standard, and [žar'a] and [šag'at,] in the present literary norm.

The differences between the dialects are reflected in different P rules. In order to bring out this difference, corresponding rules are given next to each other. Rules P 7g′ and P 9a′ apply to the old Moscow standard; Rules P 7g and P 9a, to the contemporary norm; while Rule P 8 is common to both as well as to most Southern Russian dialects.

Rule P 7g′. In position after {c} unaccented nondiffuse noncompact vowels – i.e., {o} and {e} – become diffuse and of high tonality; i.e., {i}.

Rule P 7g. In position after {c} {š} and {ž} unaccented noncompact nondiffuse vowels become diffuse and of high tonality.

Rule P 8. Unaccented nondiffuse noncompact vowels become compact and nonflat; i.e., {a}.

Rule P 9a′. After compact acute (palatal) consonants and after all sharped segments – but not after the + boundary[32] – unaccented compact vowels become diffuse and of high tonality.

Rule P 9a. After all sharped segments – but not after the + boundary[32] – unaccented compact vowels become diffuse and of high tonality.

Examples: {cen'a} [cɨn'a] "price"; {žon'a} [žɨn'a] "wife"; {v'i=šol} [v'išɨl] "went out" (masc.). Although these examples are valid for both dialects, they are consequences of different rules. In the contemporary norm, all three result from the application of Rule P 7g. In the Moscow standard the first results from the application of Rule P 7g′, and the last two, from the application of Rule P 8 followed in turn by the application of Rule 9a′; i.e., {žon'a} → {žan'a} → {žin'a} and {v'i=šol} → {v'i=šal} → {v'i=šil}; which ultimately are converted into [žɨn'a] and [v'išɨl], respectively. Rule P 8 accounts for cases like {vol'i} [val'i] "oxen", as well as for {v=m'or,+o} [vm'or,ə] "into the sea"

[31] Primarily forms of the verb {*tka} "weave".

[32] Cf., the discussion in Chapter I, sec. 2.2 above concerning the contexts in which the + boundary is postulated.

and {p'old,n,+om} [p'oldn,əm] "noun" (instr. sg.), which are not subject to Rule P 9a or P 9a′ because of the intervening + boundary. In this, the latter two differ from cases like {v=m'or,e} [vm'or,i] "in the sea" or {v=s,'in,om} [fs,'in,im] "in the blue", where no + boundary precedes the last vowel and Rule P 9a or P 9a′ applies. The latter rules account also for examples like {kl,al'a} [kl,il'a] "cursed" (fem.), which are pronounced alike in both dialects. {šal'un} "imp", on the other hand, is subject only to Rule P 9a′ and not to Rule P 9a, and is, therefore, pronounced differently in the two dialects. In most instances the different P rules of the dialects account for the different pronunciations. In a few instances, however, the differences between the dialects reflect differences in phonological composition. Thus, e.g., the old Moscow [cal'uju] "I kiss" and the now all but universal [cɨl'uju] *idem.* are reflexes of {cal'uju} and {cel'uju}, respectively.

Rule P 9b. Within a simple word – and in rapid speech also across a = boundary – acute (dental and palatal) continuants become interrupted (affricates) in position after interrupted acute noncompact (dental) consonants (stops).

Examples: {m'it,sa} "to wash" (refl.) is pronounced with a [c] preceded by a long closure. Similarly {kol'od,ca} "well" (gen. sg.) as well as {kol'ot,sa} "to sting" are pronounced [kal'o^{t}ca]. On the other hand, {ot=cepn'oj} "uncoupled" and {ot=sipn'oj} "doled out" are distinguished in careful speech, the first being pronounced with an imploded stop followed by a continuant (gradual onset of fricative noise), and the second, with an imploded stop followed by an affricate (abrupt onset of fricative noise), i.e., [a^{t}cɨpn'oj] and [a^{t}sɨpn'oj]. This distinction is often lost in conversational speech.

Rule P 9c. Before {*i}, {j} is dropped in position after a vowel.

Examples: {'aji*st} ['aist] "stork"; {taj'it} [ta'it] "hides"; {*d'elajot} [d,'ɛlait] "does" (note that {jo} becomes {ji} by rules P 8 and P 9a or P 9a′). This rule does not hold across the = boundary; e.g., {za=ježž'at,} [zəjiž,ž,'at,] "to drive in".

Rule P 10a. If an interrupted consonant is preceded by a mellow interrupted (stop) consonant, the cluster is pronounced with a single explosion. If the cluster does not follow a pause, the period of occlusion is prolonged (*dolgij zatvor*). If the cluster is preceded by a vowel, the first (stop) is frequently imploded.[33]

Examples: {ot,c+'u%$_{o}$pa=l,ixč'ajot} "father is beginning to feel better", and {ot=s'up+a%$_{o}$l,ixč'ajot} "one feels better because of the soup". In normal discourse the first consonant cluster of both utterances consists of an imploded [^{t}] followed by [c]. In the first example, however, the silent interval is longer than in the second since the cluster contains two stops. In rapid speech this distinction is often obscured or altogether overlooked. Note also that as a consequence of rule P 6c, {t} in the first example becomes plain before {c}.

{pad=č'ašk+oj} [pa^{t,}č,'aškəj] "under the cup" and {pod=š'ašk+oj} [pa^{t}š'aškəj]

[33] See *Preliminaries*, sec. 2.3111 and Chapter V, sec. 7.1 below. In the present work, implosion is denoted by a raised letter. Note that rule P 10a does not apply to sequences with {c} and {č}, since these are not mellow.

"under the sword" differ from each other as a consequence of rule P 6b, which requires that the final stop of the preposition be sharped (palatalized) before (the sharped) [č,].

For the position after pause there are examples only before {šč}: {*t*ščet+'a} [č,š,it'a] "uselessness" and {*ščit+'a} [š,š,it'a] "shield" (gen. sg.).[34]

Rule P 10b. Before plain segments and before word boundaries, {'e} becomes more compact (more open); i.e. ['ɛ].

This rule accounts for the distinction in the vowels in words like {*v'es} [v,'ɛs} "weight" and {*v'es,} [v,'es,] "entire", as well as for the distinction between utterances like {po=g*or'e l,i} [pəgar,'ɛl,i] "by way of the mountain?" and {po=g*or'el,i} [pəgar,'el,i] "have they burned down?"[35]

Up to this point the only tonality feature specified for vowels has been flat vs. natural (rounded vs. unrounded). The following four rules specify the second tonality feature, grave vs. acute (back vs. front), which functions nondistinctively in the vowels.

Rule P 10c. High tonality vowels are acute (front).

Rule P 10d. Low tonality as well as compact vowels are grave (back).

Rule P 11. In position between sharped nonvowels, grave vowels tend to become acute.[36]

Rule P 12. If no pause intervenes, accented or unaccented {*i} becomes grave (back) – i.e. [ɨ] – in position after plain segments, regardless of the presence or absence of other boundaries.[37]

Rule P 12 obviates the necessity for postulating [ɨ] as a separate phonological entity distinct from {*i}. The rule predicts the appearance of [ɨ] in cases like {ž'oni} "wives", {v*ol'i} "oxen" It also accounts for the distinction between utterances like {k='ire} [k'ir,i] "to Ira" and {k'ire} [k,'ir,i] "Kira" (dat. sg.).[38] In the latter word, the sharping of {k} is a consequence of the operation of rule P 6d.

[34] As an illustration of the operation of the phonological rules, the derivation of the correct phonetic value of the cluster {*t*šč} is given here:

{*t*šč} becomes {tšč} by rule P 3a
{tšč} becomes {tš,š,} by rule P 4
{tš,š,} becomes {t,š,š,} by rule P 6b
{t,š,š,} becomes {t,č,š,} by rule P 9b
{t,č,š,} not after pause becomes [č,š,] by rule P 10a.

[35] Example from R. Jakobson and M. Halle, *Fundamentals of Language*, p. 18.

[36] Boyanus, *op. cit.*, p. 30: "[a] may occur in every position except between two palatalized consonants." Cf., also Avanesov, *Fonetika*, p. 96.

[37] In formulating this rule I follow Avanesov and those scholars who take issue with Ščerba for considering pronunciations like [br'at ɨv'an] "brother Ivan" as instances of an "incomplete style" Cf., Avanesov, *Fonetika*, p. 48.

[38] The example is taken from A. A. Reformatskij, "Fonologičeskie zametki," *Voprosy jazykoznanija*, *6*, 2, 101–102 (1957), where a solution quite similar to the one above is advanced. See also *Fundamentals*, p. 18.

Rule P 13a. When next to plain segments, acute (front) vowels have [u]-like transitions.[39]

Rule P 13b. When next to sharped segments, grave (back) vowels have [i]-like transitions.[39]

Rule P 14a. In pretonic and in absolute word initial position unaccented [a] becomes more lax. In all other positions unaccented [a] is, moreover, heavily reduced, i.e. [ə].[40]

Rule P 14b. Unaccented diffuse vowels are reduced and lax; i.e. [u], [ɪ], [ı] and [ə].[40]

3.1 *The Phonological Rules of Foreign Words.* In the pronunciation of words of foreign origin, great vacillations can be observed. On the one extreme, there are speakers who attempt to pronounce such words as they are pronounced in the language of origin, while on the other extreme there are those who "russify" the words completely. Certain conventions are, however, fairly generally observed by educated speakers.

The rules stated in sec. 3 do not describe the distribution of the feature of sharping in foreign words.

In proper names of West European origin {š} and {ž} are sharped before {*i}; e.g., [ž,'id] "Gide"

In French and German loanwords of all types compact (palatal and velar) consonants are sharped before {*u}; e.g., [ž,'ül,] "Jules", [ž,ür,'i] "jury", [g,üg'o] "Hugo", [k,'üx,ilbɛk,ɛr] "Küchelbecker", [k,üv,'etkə] "shallow basin", [braš,'üra] "brochure"

Rules P 1a, 8 and 9a do not apply in words of foreign origin. Plain (unpalatalized) consonants and liquids appear before {*e}; e.g., [kašn'ɛ] "scarf", [vol,t'ɛr] "Voltaire", [ab'ɛ] "abbé", [šos'ɛ] "highway, Fr. chaussée", [ant'ɛnna] "antenna". Unaccented {e} and {o} are not transformed into [i] and [a] respectively. Examples in addition to those already quoted: [bol'ɛro] "bolero", [ot'el,] "hotel".

The hierarchical order of the two rules is reflected in their application to foreign words. Thus, it is impossible to apply rule P 1a to foreign words without applying also rules P 8 and 9a, but it is possible to apply the latter rules without applying rule P 1a. E.g., Voltaire's name may be pronounced either [vol,t'ɛr] or [val,t,'ɛr], but never *[vol,t,'ɛr]; or the cognate of the French *hotel* may have the variants [ot'el,], [at'el,] and [at,'el,], but not *[ot,el,].[41] The sharping of noncompact grave (labial) consonants in this position – i.e., pronunciations like [ab,'ɛ] "abbé" or [p,'ɛns] "pence" – is considered substandard.

Rules P 2 and 3a do not govern the distribution of the voicing feature. The exact

[39] For acoustic information about formant transitions in vowels next to plain and to sharped segments, see Chapter V, sec. 9.1 below.

[40] In rapid speech the differences among reduced vowels are often obscured. [ə] and [ɪ] in words like {po*lev'oj} "pertaining to a field" and {pi*lev'oj} "dusty" are particularly liable to confusion. See data reported below in Chapter V, sec. 4.1 and Table V-3. Cf., also Avanesov, *Fonetika*, pp. 118–119.

[41] Avanesov and Ožegov, *op. cit.*, give the pronunciation [ot'el,] on pp. 282 and 542, and the pronunciation [at,'el,] on p. 540.

facts could not be established. It seems, however, that {f} is never voiced in position before voiced consonants. Thus, for example, the name of Hitler is pronounced ['adol,fg,itl,'ɛr] and not ['adol,vg,itl,'ɛr]. Note also, ['ɛfd'ur] "f major" and [gr'af b'ob,i] "Graf Bobby"

3.2 *On Prominence and Accent.* From the point of view of the phonological system of Russian, vowels are either accented or unaccented. In an actual utterance, however, vowels are pronounced with varying degrees of prominence (dynamic force).[42] Some of these variations are due to idiosyncracies of the speakers and are outside the scope of a linguistic description. Other variations in prominence can be shown to be lawfully related to the phonological context; specifically, to the feature accented vs. unaccented and to the position of particular vowels with respect to other vowels and to certain boundaries. Prominence is, therefore, an "allophonic" feature which need not be indicated in the representation, for it can be inferred from the phonological representation by means of the rules given in the following pages.[43]

We distinguish several degrees of prominence. These will be indicated by numbers: 1 being assigned to the vowel of highest prominence; 2, to the vowel next highest in prominence; 3, to the one still lower, etc. Only five degrees of prominence are distinguished in the present exposition. This is probably more than are usually met with in actual speech. The cases where fewer degrees of prominence are distinguished can always be described as coalescing into one, the lowest two, three, etc., of the degrees of prominence that are distinguished in the maximally explicit style which serves as the basis for the present exposition.

The following rules apply only within the phonemic phrase; i.e., not across the boundary.[44]

Rule P 15. In a phonemic phrase containing two accented vowels, the prominence on the first accented vowel is lower by one degree than that on the second.

This rule applies equally to phrases like {st'$\overset{2}{\text{a}}$rij % barab'$\overset{1}{\text{a}}$nščik} "old drummer" as well as to composita like {*tr,'$\overset{2}{\text{o}}$x %gol'$\overset{1}{\text{o}}$vij} "three headed" and {p'$\overset{2}{\text{a}}$rt–b,i*l'$\overset{1}{\text{e}}$t} "party card".

Rule P 16a. Within a phonemic phrase, unaccented vowels are lower in prominence than all accented vowels.

Rule P 16b. The word boundary *before* unaccented words is eliminated unless it coincides with a phrase boundary.

This rule makes enclitics and other unaccented words part of the preceding accented

[42] No objective definition of "prominence" has as yet been devised, nor was it possible to devise one in the course of these studies. The term is, therefore, used without reference to a specific physical measure. If such a measure is defined, however, it will have to conform to the rules given in this section.

[43] Emphatic stress is not considered here. It can best be handled as a transformation of the normal (unemphatic) prominence relations.

[44] Cf., Chapter I, sec. 2.2 on the contexts in which the phonemic phrase boundary is postulated.

word, if any, within the same phonemic phrase. This rule is necessary in order to state rule P 17 in the most general fashion.

Rule P 17. The domain of this rule is delimited by two successive word boundaries. Within this domain unaccented vowels have different degrees of prominence as indicated in the following list:

1 degree lower than the accented vowel: vowels in pretonic position or in absolute initial; i.e., immediately following the word boundary.

2 degrees lower than the accented vowel: vowels in absolute final position; i.e. immediately preceding the word boundary.

3 degrees lower than the accented: vowels in syllables following and preceding the word boundaries, but not in absolute terminal or in pretonic position.

4 degrees lower than the accented vowel: vowels in all other positions.

Examples: $\{\mathrm{d\overset{4}{o}{=}{*}st{*}\overset{5}{\scriptstyle A}{=}pr,\overset{5}{i}{=}{*}m\overset{2}{e}\check{c}'\overset{1}{a}t,{*}\overset{5}{A}l,n\overset{5}{o}{*}st,{+}\overset{5}{a}m,\overset{3}{i}}\}$[45] "tourist sights" (instr. pl.); $\{\mathrm{\overset{2}{i}v\overset{2}{a}n'\overset{1}{o}v}\}$ "Ivanov"; $\{\mathrm{d\overset{2}{o}{=}n'\overset{1}{o}s}\}$ "denunciation", but, n.b., $\{\mathrm{d\overset{3}{a}\%n'\overset{1}{o}s}\}$ "and the nose, too", since in the proclitic the vowel is in absolute final but not in pretonic position; $\{\mathrm{d'\overset{2}{o}br\overset{4}{i}j\ m'\overset{1}{a}l\overset{3}{i}j}\}$ "good fellow"; $\{\mathrm{p\overset{4}{o}d{=}p\overset{2}{o}{=}r'\overset{1}{u}\check{c}\overset{5}{i}k\overset{5}{i}\%\check{z}\overset{3}{i}}\}$ "but the lieutenants".

Rule P 18. If in a phonemic phrase one or more of the positions listed in rule P 17 are missing, the vowels in positions having lower degrees of prominence are moved up one degree.

Examples: $\{\mathrm{d\overset{3}{o}r\overset{2}{o}g'\overset{1}{o}j}\}$ "dear" (nom. masc.) but $\{\mathrm{d\overset{4}{o}r\overset{2}{o}g'\overset{1}{o}m\overset{3}{u}}\}$ "dear" (dat. masc.).

Rule P 19. If by operation of rules P 17 and 18 degrees of prominence lower than 4 are required, they are all coalesced into level 5.

Examples: $\{\mathrm{d\overset{5}{o}r\overset{3}{o}g'\overset{2}{o}m\overset{4}{u}\ \overset{2}{i}v'\overset{1}{a}n\overset{3}{u}}\}$ "dear Ivan", but cf. $\{\mathrm{d\overset{4}{o}r\overset{2}{o}g'\overset{1}{o}m\overset{3}{u}}\}$.

[45] The symbol *A represents a further unspecified nondiffuse vowel.

PART II

ACOUSTICS

CHAPTER III

A NON-TECHNICAL ACCOUNT OF SOME ELEMENTARY CONCEPTS AND TECHNIQUES OF ACOUSTICS

1. Introduction

Relatively recent developments in the techniques of electrical measurements have made possible significant advances in our understanding of the acoustical properties of speech. A number of new research tools has been put at the disposal of the investigator, and a fairly voluminous literature on the acoustics of speech has come into being. Since a large proportion of the works on speech acoustics is written in the language of physics, few linguists have been able to follow the new developments except on a fairly superficial level. As a result, acoustic phonetics is still something of a novelty for students of language, many of whom treat it either with great suspicion or with uncomprehending admiration, but few of whom feel competent to come to grips with the problems which this new expansion raises.

This state of affairs is an unhealthy one, for not only has the linguist a great many valuable facts to learn from the acoustician, but he also has many insights to contribute which would materially accelerate the acoustical investigations of language. A change is, however, not likely to occur until linguists have gained some understanding of the methods of acoustics and of the relevance which their knowledge has to the problems that are coming to the fore as a result of the new evidence collected. The following pages were written with the aim of doing something towards remedying this unsatisfactory situation.

2. Elementary Exposition of Some Concepts and Techniques of Acoustics

To the physicist "sound" signifies a mechanical disturbance that is propagated in a medium. To get a better insight into disturbances of the kind that are of interest here, we consider the movements of a very special particle, a small weight attached to a spring (Fig. III-1). In the beginning the particle is in its *rest position.* We now pull the particle downward and release it. The particle then moves upward through the rest position to some point above the latter, where it stops and reverses its direction of travel, again passing through the rest position to some point below it. The particle

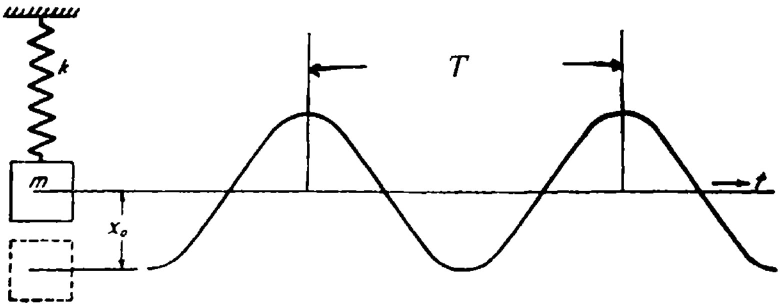


Fig. III-1. The particle m on the spring k executes sinusoidal motions as represented by the curve; x_0 is the amplitude and T is the period of the sine wave.

will execute this movement several times, but each time it will move through a shorter distance, until finally it will stop at the rest position.

If we attach some kind of marking device, say a piece of lead, to our particle, and move a piece of paper at right angles to the direction of travel of our particle, we shall obtain a curve like that in Fig. III-1. It is clear that for any given pull the motion may continue for a longer or shorter time, depending on the quality of the spring, the friction of the air, etc. In the ideal, frictionless case the particle will move forever through the same distance above and below the rest position.

Curves like that in Fig. III-1 are known as *sine* or *cosine* waves. The maximum displacement from the rest position is the *amplitude* of the wave. The time it takes for the particle to return to the initial position is the *period* or *cycle*. The number of periods or cycles per second (cps) is the *frequency*. When, as is the usual case, the amplitude of the vibration decreases with time, we speak of a *damped sine wave*. The definition of the frequency of the damped wave is the same as that of the undamped one.

A tine of a vibrating tuning fork behaves almost exactly like our particle, except that it executes many more vibrations per second, or to use our new language, vibrates with a higher frequency. The vibrations of the tuning fork cause the same motion in the air, whence it is transmitted to our eardrums. Sounds like those produced by the tuning fork, which ideally are nothing but single sine waves of any amplitude and frequency, are called *pure tones*.

When the vibrations of a pure tone strike our ears and are perceived by us, we do not describe our sensations in terms of frequency and amplitude of the stimulus, but rather in terms of the perceptual attributes of *pitch* and *loudness*, since what we perceive has little in common with our perception of the usual kind of vibrations; e.g., the bobbing up and down of spring.

Variations in frequency of a pure tone are perceived primarily as variations in pitch, while variations in amplitude affect mainly our perception of loudness. However, variations in pitch affect the loudness of a pure tone and variations in amplitude affect its pitch. It is a well-known fact that, by and large, constant ratios between frequencies of two pure tones correspond to constant pitch intervals. Thus, a ratio of 1:2 in the frequencies of two pure tones corresponds to an interval of an octave; from which it follows that a tone one octave above middle *a* (440 cps) has a frequency of 880 cps, while the tone two octaves above middle *a* has a frequency of $440 \times 2 \times 2 = 1760$ cps (and not $440+440+440=1320$ cps). Similar constant ratio relationships obtain by and large in our perception of amplitude differences of pure tones. The measure used to specify amplitude differences is the *decibel* (db), and like the octave and other musical intervals, it is a measure not of absolute values, but of their ratios. The decibel is defined by the following expression

$$N_{(db)} = 20 \log_{10} \frac{A_1}{A_0}$$

where A_1 and A_0 are the respective amplitudes of the sounds being compared, and N, their difference in db. Thus if a sound has an amplitude that is ten times larger than that of another sound – i.e., $\frac{A_1}{A_0} = 10$ – their difference is 20 db. It follows from the above definition that it is always necessary to specify a reference amplitude (A_0) with which the measured sound is being compared. This situation may be familiar to the reader from music where an expression such as "2 octaves" normally presupposes some reference tone with respect to which the 2 octaves are specified.

The air particles that strike our ear need not behave as regularly as those just discussed, and, as a matter of fact, they usually do not. Consider the case of two identical sine waves propagating simultaneously. Consider, in particular, the vibration of air particles in a plane which is the perpendicular bisector of the line joining two identical tuning forks when these are struck simultaneously. The nature of the resultant particle movement will depend on the relation between the times at which the two sine waves reach the same point in their cycles. When both sine waves are simultaneously at identical stages of their cycles they will reinforce each other everywhere, i.e., there will be a sine wave of double the amplitude. We say in such a case that they are *in phase*. When they are not in phase they will reinforce each other at times and suppress each other at other times. In one extreme case one of the waves will reach its maximum point exactly at the moment when the second reaches its minimum point. In this case, the two sine waves cancel each other out, and we have then the seemingly paradoxical situation where sound plus sound yields silence.

By adding sine waves of different frequencies, amplitudes and phase relations, it is possible to generate an infinity of different curves. As a matter of fact, it has been shown by the French mathematician, J. B. J. Fourier (1768–1830), that all possible curves can be generated simply by adding a number (in certain cases infinitely many) of sine waves of differing amplitudes, frequencies, and phase relations. Furthermore the frequencies of the different component sine waves are integral multiples of the lowest or *fundamental* frequency. For example, if the fundamental frequency is 100 cps, then the remaining components of the sound are 200, 300, 400, etc., cps, and there are no components in between. The various components constituting a sound are called *harmonics.*

This fact is of crucial importance for the physical description of sound. It means nothing less than that any sound at all can be viewed as the sum of sounds of a single type; i.e., of sine waves that differ only in amplitude, frequency, and phase relations.

Sine wave analysis has been one of the most powerful tools of acoustics. A fairly elementary form of sine wave analysis has dominated the field of psycho-acoustics almost to the exclusion of all other possible methods of analysis. Since in the following we, too, shall discuss sound in terms of elementary Fourier analysis, it is necessary to state at this point that sine waves are not the only class of curves into which all possible curves can be analyzed, and that there are limitations to the applicability of Fourier analysis.[1]

We now turn to a description of some secondary acoustical concepts which we shall explicate in terms of the primitives of acoustics, i.e., in terms of frequency, amplitude, and phase.

When we play the piano in a room it sometimes happens that as we strike a particular key, the window panes emit a sound. We say that the panes *resonate* in response to that tone, or at the frequency of that tone. The phenomenon of resonance is basic for a comprehension of speech production itself as well as for an understanding of the most important apparatus employed in the physical investigation of sound – i.e., filters. We shall, therefore, discuss it in some detail.[2]

The window panes vibrate somewhat in response to any tone; however, only in response to certain tones does their vibration become large enough to be audible. We describe these facts negatively by saying that a resonator tends to suppress all frequencies except those at which it resonates. It follows from this that the response of a resonator to a sound not containing frequency components close to its resonance frequencies will be relatively small. We have seen that the panes suppress essentially all but a certain note, since a note emitted by a piano contains relatively few frequency components of sufficient magnitude. If, however, we were to depress all the piano keys at once – i.e., emit a sound rich in frequency components – we are quite certain to make the panes resonate. (Another way of exciting the resonances of the panes is, as everybody knows, by tapping them: from this it would appear that a tap is a type of sound containing components at many frequencies. We shall see later why this is, indeed, the case.)

Except for the fact that the components at the resonance frequencies must be present with a minimum intensity in the exciting sound, the frequencies at which a body

[1] Cf., J. C. R. Licklider, "Basic Correlates of the Auditory Stimulus," in S. S. Stevens, ed., *Handbook of Experimental Psychology* (New York, 1951), pp. 985–1039.

[2] The phenomenon of resonance was well known in antiquity. Bouasse quotes the Talmud: "It has been said by Ramé, the son of Ezeekiel: 'When a cock shall have stretched his head in the interior of a glass vase and shall have crowed therein in such a manner as to break it, the whole cost shall be payable.'" Cf., A. Wood, *Acoustics* (New York, 1947), p. 96.

resonates are independent of the excitation; they are completely determined by certain geometrical properties of the resonator and the material of which it is made.

It is very illuminating to consider the process of vowel production from this point of view. The vocal tract is a resonator which is excited by vibrations of the glottis or by a whisper. Since both the vocal cord vibrations as well as the whisper are complex, a great number of sine waves of different frequencies are required in order to represent them by means of Fourier's theorem. This, however, is but another way of saying that these excitations contain components at a great many frequencies, and, therefore, the resonances of the vocal tract usually are excited regardless of the nature of the speaker's voice. The resonances, as we have already said, are controlled by geometrical properties of the resonator, and it is by changing the shape and size of the vocal cavity and of its openings to the outer air (primarily the lip opening) that we produce the different vowels. The particular speech sound quality of a vowel, its "tamber", is produced by the configuration of the vocal tract and is entirely due to the resonances of the tract. Differences in the pitch with which a vowel is pronounced are produced by differences in the rate of vibration of the vocal cords, i.e., by the fundamental frequency of the vocal cord excitation.

In the last pages we have considered sounds not as vibrations of air particles – i.e., not as changes in the position of an air particle with respect to time such as one sees in the usual display of sound on the face of a cathode ray oscilloscope – rather have we viewed sounds in terms of certain of their sine wave components, the resonances. We shall now point out the interrelations between these two ways of looking at sound.

We have said before that any wave at all could be represented by a sum of sine waves differing in amplitude, frequency, and phase relations, where the frequencies of the different components (harmonics) are integral multiples of the fundamental. In our discussion of resonance we observed that the components at resonance frequencies have amplitudes above other components. We have thereby made a statement about the relation of the amplitudes among sine waves of differing frequencies. This fact could be represented in a graph like Fig. III-2. Such a graph is commonly known as a *line spectrum*. Since each of the components is a single sine wave of definite frequency, the components are represented here by vertical lines. The solid line connecting the tops of all frequency components is known as the *spectrum envelope*.

A true line spectrum, where each frequency component is represented by a line, which ideally has no width, can be obtained only for sounds of infinite duration. In the case of sounds of finite duration the lines are of finite width, and the shorter the sound, the wider are the lines representing its frequency components.

Some understanding of the reasons for this may perhaps be gained from the following not quite rigorous consideration. When we measure the frequency of a component we are measuring basically the recurrence of an event. It is obvious that given two regularly recurring events, one with a high frequency and the other with a low frequency of occurrence, we would need less time to determine the frequency of the former than that of the latter. Similarly, the more exactness we need in our

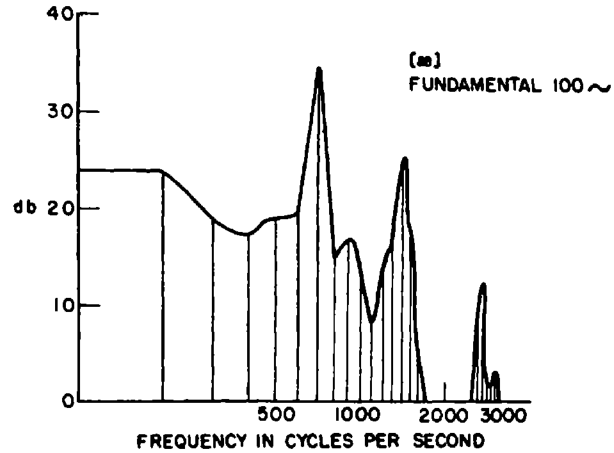


Fig. III-2. Line spectrum and spectral envelope of vowel [æ]. The vertical lines represent different frequency components of the sound. The heavy solid line connecting the tops of the vertical lines is known as the *spectrum envelope*. (Adapted, with permission, from R. K. Potter and J. C. Steinberg, "Toward the Specification of Speech," *JASA*, *22*, 812 (1950).)

determination of the frequency, the longer must we observe the event. In an ideal line spectrum the frequency components are by definition absolutely exact. It follows from the above considerations that a line spectrum where the frequency of each component is given with absolute exactness presupposes an infinitely long sound. If now we think of the thicker lines as being due to a certain amount of indeterminacy in frequency, it becomes clear why sounds of finite duration must have components of finite frequency width. The shorter the duration of the sound, the greater the width of the component. It also follows from the above that a sound lasting an infinitesimally short time – i.e., an idealized click, a pulse – must have components at all frequencies. Such a spectrum is called a *continuous* spectrum. Since the amplitude of the spectrum is the same everywhere, we say that the spectrum envelope is *flat*.

(We may recall at this point the observation made earlier that a tap, which is essentially a pulse, can be used to excite the resonances of window panes regardless of their construction. We are now in a position to appreciate the reason for this.)

Sounds which possess spectra in which all frequency components are present with random phase relations are called *noises*, and if their spectral envelopes are flat – i.e. if all their frequency components are of the same amplitude – we say that the noise is *white*. Pulses and white noises, therefore, have identical spectra. The difference between these two types of sound lies in the phase relations of their components. In the white noise the phase relations among the various frequency components are completely random; the frequency components of the single pulse all reach their maximum amplitudes at the time t=0, and cancel each other at all other times.

In the real world there are, of course, neither infinitely short nor infinitely long events. The closest approximation to a pulse that we meet with in speech is the burst of a stop like [p] or [t]. Fricatives such as [s] or [f] are examples of types of noises.

The ideal cases considered above make it possible for us to extend our picture of speech production. In talking of vowels we have already distinguished between two factors that are involved in the production of any speech sound: an excitation and a resonator. The resonances of the vocal tract are controlled by the shape that the tract is made to assume. To each configuration there corresponds a definite resonance pattern. The resonances alone, however, do not determine the nature of the output completely; in order to specify it completely we need to know the nature of the excitation and the place in the vocal tract where it is applied. The simplest case is that of the vowels and sonorants (nasals and liquids) where the excitation is applied at the very end of the vocal tract, at the glottis. Normally this excitation is due to the vibration of the vocal cords which produces little puffs of air that recur at regular intervals. This regular recurrence determines also the fundamental frequency and hence the spacing of the harmonics in the vowel and sonorant spectrum. Since all these sounds are also of a duration that for practical purposes might be considered infinite, their spectra are line spectra (cf., Fig. III-2). In the case of the whisper, however, the resonator is excited by a noise and the spectrum is continuous.

Stops and fricatives are produced by an excitation that is not located at the glottis end of the resonator. This fact causes certain complications which, however, shall not be discussed here. The excitation in the case of fricatives is noise-like, and in the case of stops it is approximately an impulse. Hence both types of sound have continuous spectra. These spectra, however, are usually not flat; very much like vowel spectra, they show regions of greater and lesser prominence produced by the resonances of the vocal tract. In voiced stops and fricatives the vibration of the vocal cords provides a supplementary excitation. In these cases a discrete, line spectrum is superimposed on a continuous spectrum.

Figure III-2 is a line spectrum of the vowel [æ]. The frequency components of the vocal cord vibrations are represented by the vertical lines, and the resonances are due to the specific configuration of the vocal tract. In the phonetic literature these resonances are known as *formants*. In the case of the vowels the shape of the spectrum envelope can be predicted if we know the frequencies of the formants.[3] This is usually taken to mean that the vowel quality as distinct from the pitch, intonation, etc., with which a given vowel is pronounced, is primarily determined by the position of the formants. Since for theoretical reasons there must be three formants below 3000 cps, and since this region is clearly of maximum importance as far as vowel recognition is concerned, most investigators have chosen to specify vowels in terms of the lowest two or three formants.

It is important to bear in mind that although in specifying vowels by their formants we tend to speak as if vowels were composed of two or three frequency components (pure tones), formants are not pure tones and very little can be said about our perception of vowels on the basis of our knowledge of the perception of pure tones. This

[3] C. G. M. Fant, "On the Predictability of Formant Levels and Spectrum Envelopes from Formant Frequencies," *For Roman Jakobson* ('s-Gravenhage, 1956), pp. 109–121.

fact was illustrated in a recent investigation by J. L. Flanagan, who tried to establish the smallest frequency shifts in a formant of a vowel that could be perceived.[4] He found that shifts of less than 3 to 5% in the formant frequency were not noticed at all, while a shift of 1% in the frequency of a pure tone is readily noticed. He also found that in the case of formants close to each other, decreases in the distance between them are much more readily perceived than are increases. This observation contrasts with the fact that for single pure tones the direction of the shift has little effect.

We have described a resonator as a device which suppresses frequency components other than those at which it resonates. This property makes resonators ideally suited for determining what frequency components are actually present in a sound. When functioning in this capacity, resonators are referred to as *filters.*

In our discussion of resonators above we have concentrated our attention entirely on the fact that only some frequency components are passed by a resonator. We must now consider in some detail the manner in which resonators pass the different frequency components.

When we examine Fig. III-2, we note that frequency components near the resonance frequencies are attenuated less than those farther away. As a matter of fact each resonator attenuates in its own way frequency components not at resonance. The manner of attenuation is determined by certain details in the construction of the resonator. This property is made use of in the design of filters. For some tasks we may be interested in having filters which pass only very few frequency components, while for other purposes we may need filters that pass very many frequency components.

It is customary to specify, in addition to the resonating frequency of a filter, the two points on the frequency curve which have one-half the power of the component at the resonating frequency, i.e., which are attenuated to one-half their original value. These two frequencies are known as the *cut-off frequencies* of the filter, and their difference, as the *band width* of the filter. In Fig. III-3 we have shown the response curve of a filter which passes frequency components between 450 and 2250 cps. Its bandwidth consequently is 1800 cps. Frequency components outside this region are attenuated to at least one-half of their original value.

In addition to band pass filters there are filters which attenuate only frequency components below or above a certain frequency. The former are called *high pass* filters, and the latter are called *low pass* filters. The band pass filter of Fig. III-3 can be thought of as consisting of a high pass and a low pass filter.

The sound spectrograph, or sonagraph,[5] as it is called commercially, is probably the instrument most widely used for speech analysis in the U.S.A. today. A schematic diagram of this instrument, intended to show how it operates, is given in Figure III-4. The sound is recorded onto a revolving magnetic drum (A) which can accomodate up

[4] J. L. Flanagan, "A Difference Limen for Vowel Formant Frequencies," *JASA, 27,* 613–617 (1955).
[5] For a detailed description see *Technical Aspects of Visible Speech, Bell System Technical Monograph B-1415* (1946) = *JASA, 18,* 1–86 (1946).

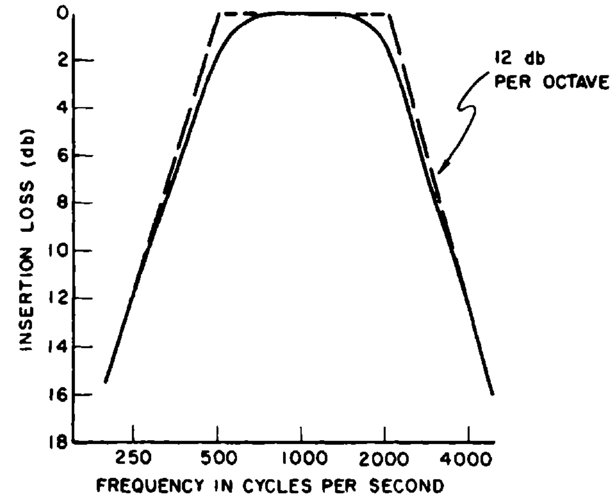


Fig. III-3. Response curve of a band pass filter which passes frequencies between 450 and 2250 cps.

to 2.4 seconds of sound. The sound recorded on the drum is analyzed by a filter (B) of fixed bandwidth but variable center frequency. The output of the filter is connected to a stylus (C). The stylus burns a trace on a piece of specially treated paper wound around drum (D) when the output of the filter exceeds a specified minimum value. The blackness of the trace roughly indicates the intensity of the component at a given frequency and time. Drum D is mounted on the same shaft as the magnetic drum and, therefore, moves in synchrony with it. When the stylus is set at the bottom of the paper, the center frequency of the filter is at 0 cps. The stylus and filter are connected to the driving mechanism in such a way that for each revolution of the drum, the filter moves its center frequency up by a certain amount while the stylus moves a small distance up the paper. Thus the sonagram is a spiral of closely spaced lines.

Another well-known device in present-day speech research is the Pattern Playback developed at Haskins Laboratories.[6] Its operation is essentially the reverse of the sonagraph: it translates visible patterns into sound. Originally photographs of sonagrams were used on the pattern playback, but it was found possible to use handpainted versions of the sonagram, in which the patterns were highly simplified. The use of these handpainted sonagrams has in recent years led to a large number of interesting and important experiments and discoveries regarding the nature of the distinctive parameters of speech perception.

On examining Fig. III-2 again, we find that it contains no information about phase relations. This is an important omission since many different wave shapes may have the same frequency spectrum. In spite of this omission, the spectrum envelope contains sufficient information for the solution of many speech problems.

[6] F. S. Cooper, A. M. Liberman, J. Borst, "The Interconversion of Audible and Visible Patterns as a Basis for Research in the Perception of Speech," *Proceedings of the National Academy of Sciences*, *37*, 318–325 (1951).

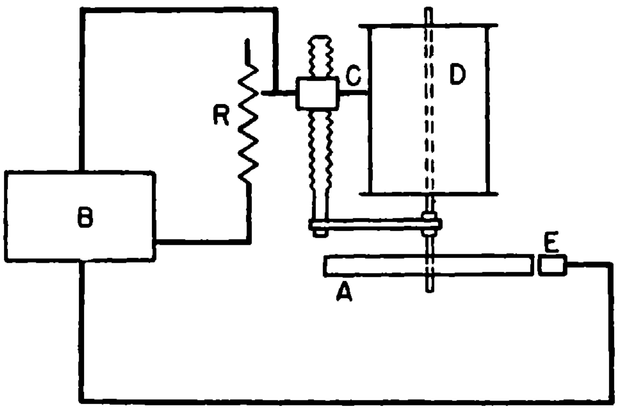


Fig. III-4. Diagram illustrating the principles of operation of the sonagraph. See explanation in the text.

3. Digression on the Relation Between the Descriptive Frameworks of Acoustics, of Psycho-Acoustics and of Phonetics

The elementary terms of acoustics and of phonetics are not simply related: there is no immediately obvious, self-evident way of translating the elementary terms of phonetics – i.e., terms like vowel, nasality, etc., in which phoneticians describe the sounds of speech – into frequency, amplitude, and phase relations of sine waves. This is not surprising if we consider that the two descriptive frameworks were developed quite independently by two groups of researchers, who usually did not even know of each other's existence. The reasons for the choice were, however, the same in the two cases: the particular framework was chosen because it was felt that this framework would yield the simplest adequate description of the data, and that a different framework would lead to more complicated statements.

We have seen that this is rigorously true for the case of acoustics; where waves of any kind of complexity can be decomposed into sine waves differing in frequency, amplitude, and phase. It might be noted here that sine waves are a very special kind of motion and are the consequence of a rather singular configuration of forces acting upon a body. The fact that from the point of view of mechanics rectilinear motion is, in a sense, simpler, has never been used as an argument for employing rectilinear motion as an elementary term in the field of acoustics.

The case of phonetics is somewhat less clear since phoneticians – at least in modern times – have not been in the habit of analyzing the reasons for their most important activities. The problem was, however, faced by the Hindu grammarian Pānini, who apparently was conscious of the grammatical implications of his phonetic classificatory scheme. As Professor Thieme tells us: "The arrangement of Pānini's list of sounds, which at first looks rather disorderly, is explainable as due to the phonetic catalogue of sounds having been adapted to the practical requirements of the grammar in which Paṇini wanted to refer to certain groups of sounds by short expression."[7]

In recent years Roman Jakobson, in attempting to justify the use of a particular set of primitives in phonetics, employed essentially the same line of reasoning. He tried to show that a host of linguistic facts from diverse languages could be simply described in terms of the particular set of distinctive features that he proposed, but would require long lists if another scheme – e.g., another set of features – were adopted.[8]

We summarize: in the case of phonetics and acoustics the choice of the descriptive framework was based on considerations intrinsic to the particular science without reference to the decisions made by other sciences.

There is yet another requirement that terms used in a scientific description are ordinarily made to satisfy: at least some of them must be translatable into the language of physics. This condition played a very secondary role in the development of phonetics, which consequently has never achieved the status of a "hard" science. It exercised an overriding influence on the development of psycho-acoustics. Partly as a consequence of the fact that psycho-acoustics was developed by physicists who were, perhaps, excessively eager to achieve the reduction of a behavioral science to physics, and partly because of the fact that in spite of its non-linear nature, the human auditory system acts in a great many instances much like a linear device and therefore made such a reduction appear a relatively simple matter, classical psycho-acoustics started by taking over in its entirety the framework of acoustics. It attempted to find the psychological correlates of the frequency, amplitude, and phase relations of sine waves, without stopping to answer the question as to the relevance of these terms from its internal point of view. In the initial stages of psycho-acoustics it seemed that a very simple relationship existed between the frameworks of the two sciences. This optimistic belief is still reflected in many textbooks of physics where we are told that frequency is the physical correlate of pitch, and amplitude, that of loudness, and that phase is unimportant, since "man is phase-deaf".

The facts, as they were uncovered by later investigations are, however, much less neat. Frequency and amplitude both affect our judgments of the pitch as well as of the loudness of a pure tone, and under certain conditions man is sensitive to phase. Consequently there is no simple relationship between the two descriptive frameworks. It has furthermore been almost impossible on the basis of the data accumulated for pure tones (single sine waves) to make predictions about the perception of complex acoustic stimuli. Since a principal requirement on elementary terms of a science is that all pertinent phenomena be ultimately reducible to them, this further weakens the position of pitch and loudness in psycho-acoustics. There seems, thus, to be reason to believe that pure tones do not occupy the same crucial position in the perception of

[7] P. Thieme, *Pāṇini and the Veda* (Allahabad, 1935), p. 104.

[8] Cf., the various sub-chapters entitled "Occurrence" in *Preliminaries to Speech Analysis*. See also, M. Halle, "In Defense of the Number Two," *Studies Presented to Joshua Whatmough* ('s-Gravenhage, 1957), pp. 65–72.

sound that sine waves occupy in the physical description of sound. Psycho-acoustics is, therefore, faced at present with the necessity of constructing a more appropriate descriptive framework.

In phonetics the situation is the reverse of that in psycho-acoustics. The traditional methodology of linguistics was fundamentally sound in that it selected its own descriptive framework on the basis of considerations intrinsic to its own aims without regard for the choice made by other related sciences. Until fairly recently it was, however, impossible to give precise physical meaning to the basic terms of phonetics, and many linguists were, therefore, quite willing to follow Twaddell and consider as fictions the phoneme and consequently also all other concepts of linguistics that involve phonemes. This defiant declaration of independence of linguistics from the developments in other fields, particularly in acoustics and psycho-acoustics, was obviously a move in the right direction, since it enabled linguists to continue without waiting for progress in areas over which they had no control. As a long-term solution, however, it is not acceptable, for speech evidently is a physical phenomenon and the terms in which we describe it must, in the final analysis, be translatable into the language of physics, e.g. into frequency, amplitude, phase relations of sine waves. But while we should insist on the necessity of this translation and consider it a central issue in modern phonetics, we want to stress especially that the basic terms of linguistics need not have simple translations into, or be identical with, the basic terms of acoustics. In other words, the elementary terms of linguistics need not be physically simple. The importance of translating the basic terms into the language of physics can hardly be overrated, for should such a translation actually be effected, the status of linguistics within the framework of present-day science would be firmly established. A contribution to this translation is offered in Chapter V below.

CHAPTER IV

A CRITICAL SURVEY OF ACOUSTICAL INVESTIGATIONS OF SPEECH SOUNDS

The earliest descriptions of speech sounds that could lay claim to scientific status were those of the Hindu grammarians.[1] These descriptions were made primarily in terms of the position and action of the organs of speech and would not be relevant to the present discussion except for the fact that they illustrate strikingly certain features of phonetic descriptions in general.

The description begins with a fixed set of parameters of definite structure. Every sound in the language is classified according to a small number of attributes, e.g., degree of closure, point of articulation, etc. All other attributes which a sound may possess are omitted from consideration: they are not significant variables. Furthermore, each of the significant variables can assume only a very restricted number of values: there are five significant points or articulation, two degrees of aspiration and of voicing, etc.; although physically speaking there are, of course, infinitely many.[2]

The Hindus had thus created an important tool for scientific description: a theoretical framework in terms of which articulatory data could be recorded. For acoustical phenomena the beginnings of such a framework existed in the musical notation, which was developed in the Middle Ages. It was, however, not until much later that musical notation was used to describe acoustical properties of speech. In his book *Mathesis Mosaica* (Cologne, 1679),[3] the German scholar Samuel Reyherr published data about the "characteristic pitches" of French and German vowels. These were given in musical notation and correspond fairly well with data for the second formant of vowels.

In 1681 the famous British physicist Robert Hooke demonstrated before the Royal Society "a way of making *Musical and other Sounds*, by the striking of Teeth of several Brass Wheels, proportionately cut as to their number, and turned very fast round, in

[1] See W. D. Whitney, *Sanskrit Grammar* (Cambridge, 1941), pp. 1–34; and also W. S. Allen› *Phonetics in Ancient India* (London, 1953).

[2] Compare the important distinction between "scientific" and "selective" information made by D. MacKay, "The Nomenclature of Information Theory," in H. von Foerster, ed., *Proceedings of the Eighth Conference on Cybernetics* (New York, 1951), pp. 223–224.

[3] Pp. 432 ff. as quoted in C. Stumpf, *Die Sprachlaute* (Berlin, 1926), p. 148.

which it was observable that the equal or proportional stroaks, that is, 2 to 1, 4 to 3, *&c.* made Musical Notes, but the unequal stroaks of Teeth more answered the sound of the Voice in speaking."[4] This wheel is the earliest objective standard of frequency. Since Hooke's device and procedure could be duplicated with some accuracy, the frequencies of sounds could be established by stating the number of revolutions per second that the wheel executed when a particular note was heard.[5] This device was, however, not destined to play a role in speech studies; it was too cumbersome and heavy. The frequency standard used widely in speech studies was the tuning fork, which was invented in 1711 by a British musician.

Synthesis of speech, which had first been attempted by Hooke, attracted much interest in the eighteenth century when fascination with automata of all kinds was at a peak. Between 1750 and 1780 the French abbé Mical constructed several speaking machines, about which unfortunately there seems to be no detailed information.[6] In 1779 the Imperial Academy of Sciences of St. Petersburg announced a contest for the solution of the following two questions: What is the nature and character of the vowel sounds a, e, i, o, u? and, Is it possible to construct an instrument like the *vox humana* of the organ, which can exactly express the sounds of these vowels? The following year the prize was won by Dr. Christian Gottlieb Kratzenstein, who had constructed an organ containing pipes of various shapes and configurations which could produce the desired vowel sounds.[7]

About a decade later, Wolfgang von Kempelen[8] produced a more elaborate speaking machine. He was the first to give an extensive treatment of the production of consonants. His work anticipates in many ways very modern approaches to the entire problem of speech synthesis and analysis. According to Kempelen, all vowels possess the following three properties:

1. The glottis emits sound while the nose is closed.
2. The voice upon leaving the throat is conducted by the tongue, as by a channel, directly to the lips.
3. The greater or lesser opening of the mouth completes the formation of the sound and gives it its characteristic quality (*netteté*).[9]

Kempelen defined consonants negatively. After listing the characteristics of vowels, he says: "...toute lettre qui n'a pas ces qualités n'est qu'une consonne."[10] He

[4] R. Waller, *The Life of Robert Hooke* (1705), reprinted in R. T. Gunther, *Early Science in Oxford, 6*, (Oxford, 1930), p. 57.

[5] Though invented by Hooke, the device is commonly known as *Savart's Wheel*, after a French physicist of the nineteenth century.

[6] G. O. Russell, *The Vowel* (Columbus, 1928), pp. 6–8.

[7] Cf., C. G. Kratzenstein, *Tentamen coronatum de voce* (St. Petersburg, 1780). A French summary of this work appears under the title, "Essai sur la naissance et la formation des voyelles," *Journal de Physique, 21* supplement, 358–379 (1782).

[8] Wolfgang de Kempelen, *Le mécanisme de la parole suivi de la description d'une machine parlante* (Vienna, 1791). The German edition, inaccessible to me, bears the title *Mechanismus der menschlichen Stimme nebst der Beschreibung einer sprechenden Maschine* (Vienna, 1791).

[9] *Ibid.*, pp. 194–195.

[10] *Ibid.*, p. 229.

classified the vowels in terms of two properties – degree of mouth opening, and "tongue channel opening". Increase of mouth opening gives the series u-o-i-e-a (diffuse-compact, in our terminology), and increase of "tongue channel opening" gives the series i-e-a-o-u (acute-grave).[11] In producing the sounds of speech on his machine, Kempelen controlled the following properties: (1) vocal cord vibrations, (2) role of nose, (3) position of the tongue, (4) role of the teeth, (5) position of the lips, (6) manner in which the sound was produced, especially for stops and r, and (7) volume of resonance cavity.

The speaking machine itself was described by Sir Charles Wheatstone as follows:

De Kempelen's machine consisted externally of a square box, and a bellows, of the ordinary construction, placed on a board. The bellows were pressed down by the right arm, and it expanded itself, by means of a weight or spring, when the pressure was removed. There were two openings in the box, for the purpose of introducing the hands to act on the keys, etc. within, and numerous small round apertures concealed by silk, to prevent stifling the sound. When the cover was removed, a small box or wind-chest was seen communicating with the pipe of the bellows at one end, and with an India-rubber bell or funnel, from which the sound ultimately issued, on the other; on this wind-chest were various keys to be touched by the fingers of the right hand for the sounds S, SH, R, etc. The sound was produced by an ivory or brass reed, covered on the side which vibrated against the edges of the aperture, with very thin leather; this reed, representing the larynx of the vocal organs, was placed within, between the wind-chest and the narrow part of the India-rubber funnel or mouth. The sound P was produced by suddenly removing the left hand from the front of the mouth, which it had previously completely stopped; the sound B by the same action but, instead of closing the mouth completely, a very minute aperture was left, so that the sound of the reed might not be entirely stifled; M was heard on opening two small tubes, representing the nostrils, placed between the wind-chest and the mouth, while the front of the mouth was stopped, as for P. A few vowel modifications and the sounds F and V, were produced by modifying the form and size of the aperture of the mouth by the left hand; the continuous consonants S and SH were obtained by causing the wind to pass through small tubes of particular forms, the passage of the wind being governed by keys; and R was imitated by occasioning a vibration or trembling of the reed when an appropriate lever was depressed.[12]

Very much like modern engineers,[13] Kempelen viewed the production of speech as a process of imposing modulations upon a carrier. This carrier could be (1) purely periodic (reed source), or (2) noise (passage of wind from bellows through appropriately constructed tubes). Various ways were employed by Kempelen to give the carrier its resonance properties: modification of the form and aperture of the mouth, modification of the size of the resonating cavity, or finally, passage through specially constructed resonators (nasals or fricatives). Special arrangements existed for sounds depending for their perception upon rapid fluctuations in the envelope; e.g., the stops, the trilled [r], and the nasals.

On the other hand, Kempelen was quite naive regarding the relation between

[11] In speaking of vowels Kempelen used the terms "grave" and "acute". Cf. *ibid.*, p. 201.

[12] *The Scientific Papers of Sir Charles Wheatstone* (London, 1879), pp. 363–364; cf., also H. Dudley and T. H. Tarnoczy, "The Speaking Machine of Wolfgang von Kempelen," *JASA*, *22*, 151–166 (1950).

[13] Cf., H. Dudley, "The Carrier Nature of Speech," *Bell System Technical Journal*, *19*, 495–515 (1940).

spelling and pronunciation, as a consequence of which in Kempelen's system nasal vowels are impossible by definition.[14] He also had no conception of the role played by the pharynx in the production of speech sounds.[15]

In 1835 a modified version of Kempelen's machine was demonstrated by Wheatstone before the British Association for the Advancement of Science. This machine, which was publicly displayed as recently as 1923 by Sir Richard Paget, produced, according to the latter, the vowels [æ] (hat), [ʌ] (up), [ɔ] (more), [u] (who), and the consonants [p], [m], [s], and [ʃ], and was, therefore, less versatile than Kempelen's.[16]

Early in the nineteenth century, Willis became interested in Kempelen's work, and successfully repeated his vowel synthesis, using a funnel-shaped cavity and a free reed. Afterwards he went on to experiment with the effect of various lengths of cylindrical tube, mounted like an organ pipe. From these experiments he concluded that vowel quality depended on the length of the tube and was independent of the pitch of the exciting reed. He later tested this hypothesis by utilizing, like Hooke before him, a toothed wheel (equivalent in effect to the reed in the other experiment) and varying lengths of watch spring held against the wheel (the watch spring being equivalent to the resonator tube in the other experiments). He found that by varying the length of the watch spring, he obtained a variety of vowel sounds. On the basis of his experiments Willis suggested that the vowel quality was due to the damped vibrations produced by the vocal tract.[17]

The next contribution of importance to vowel theory was made by Wheatstone, who discovered that a resonator produces not a single, but a multiplicity of resonances. He also noted the fact – later to be emphasized by other investigators, particularly Helmholtz – that when the resonance frequency of the cavity is an integral multiple of the exciting frequency, the "energy of the resonance is so greatly augmented as to produce the effect of a superadded musical sound".[18]

In his famous treatise of 1862, *Die Lehre von den Tonempfindungen*,[19] Helmholtz made the next advance in our understanding of the nature of the vowel. He pointed out that vowels were produced by exciting a resonator, the vocal cavity, and proposed that the vowel quality is determined solely by the resonance frequencies of the vocal cavity.

[14] Kempelen's first criterion for vowels is: "The glottis emits sound while the nose is closed." *Op. cit.*, p. 194; cf. above, p. 92.

[15] The first to realize this was the Czech physiologist, J. E. Purkyně, in his *Badania w przedmiocie fizyologii mowy ludzkiej* (Cracow, 1836). This work was unavailable to me, and my statement is based on P. Grützner, *Physiologie der Stimme und Sprache* (Leipzig, 1879), p. 158.

[16] R. Paget, *Human Speech* (New York–London, 1930), p. 19.

[17] According to Willis, therefore, a vowel is produced by an air pulsation travelling from the glottis to the lips, where part of the pulsation force is radiated as sound. The remainder is reflected back to the glottis, where it is again reflected and returned to the lips – the amplitude gradually diminishing until the next pulsation is sent out from the glottis by an opening of the vocal cords. Cf., T. Chiba and M. Kajiyama, *The Vowel, Its Nature and Structure* (Tokyo, 1941), p. 51.

[18] Russell, *op. cit.*, pp. 41–42.

[19] H. v. Helmholtz, *Die Lehre von den Tonempfindungen*, 6th ed. (Braunschweig, 1913); English trans. by Alexander J. Ellis, *On the Sensations of Tone*, 4th ed. (London, 1912).

When first advanced these views were felt to be very different from those of Willis, so that two opposing schools developed and battled one another with much heat. It was, however, noted by Rayleigh as early as 1877 that the differences between the two theories were much less radical. It will be recalled that according to Willis the quality of a vowel is produced by the damped vibrations of the vocal cavity. If it is possible to assume that the damped vibrations continue indefinitely – in the case of vowel sounds this assumption is a sufficiently close approximation – then, given the damped vibrations, it is possible to calculate mathematically the frequency spectrum and hence also the resonance frequencies which, according to Helmholtz, are the primary determinants of vowel quality. Helmholtz and Willis, therefore, are in agreement that the resonance frequencies of the vocal cavity are necessary for specifying vowel quality. They differ in that for Helmholtz the resonance frequencies are necessary as well as sufficient, whereas according to Willis more information is required, viz., the amplitudes and the phase angles of all spectral components.

Helmholtz's neglect of all spectral components, except those at the resonance frequency, can be justified by the fact, established theoretically by Fant, that in the case of vowels it is possible to predict the spectral envelope, given the resonance frequencies.[20] To justify his neglect of phase relations, Helmholtz cited "Ohm's Law", according to which the ear is "phase-deaf". Although this "law" has lost much of its force as a result of recent investigations,[21] in the case of vowels there seems little reason to introduce phase relations among the spectral components as a significant variable.

Some investigators, e.g., Trautmann, insisted on a "harmonic relationship" between the frequency of excitation (vocal cord vibration) and the resonance frequencies. They argued that only a harmonic relationship could assure sufficient intensity to the resonance components. This can be seen to be rather unimportant since the resonances of the vocal tract are not infinitely sharp (totally undamped) and are, therefore, excited also by components that do not bear them an exact harmonic relationship. Moreover, for any resonance frequency which is of interest in speech there are several excitation frequencies that bear a harmonic relationship to the former. Consequently, given any configuration of the vocal tract, which automatically determines also the resonance frequencies of the tract, a strong component of the resonance frequency is to be expected regardless of the frequency of the glottis excitation.

Helmholtz separated the vowels into two classes: *u*, *o*, *a*, which he specified by a single resonance, and *ü*, *ö*, *ä*, *i*, *e*, which according to him, had two resonances.[22] He determined the resonant frequencies as follows:

[20] C. G. M. Fant, "On the Predictability of Formant Levels and Spectrum Envelopes from Formant Frequencies," *For Roman Jakobson* ('s-Gravenhage, 1956), pp. 109–121.
[21] Cf., J. C. R. Licklider, "Basic Correlates of the Auditory Stimulus," in S. S. Stevens, *Handbook of Experimental Psychology* (New York, 1951), pp. 1024 ff.
[22] *On the Sensations of Tone*, p. 109, note. The frequencies are those of the corresponding musical

u 175	ü	175	1468	i	175	2349
o 466	ö	349	1109	e	349	1976
a 932	ä	587	1568			

Helmholtz was apparently unable to separate formants when they were too close to each other, or even to perceive them when they were too greatly attenuated. He, therefore, specified only a single formant for the back vowels, and two formants for the other vowels, instead of three or more. The frequency positions specified are remarkable for their accuracy.

Helmholtz offers almost no information on consonants except some vague remarks about the resonant structure of nasal sounds and the rather obvious observation that onset characteristics play an important role in the perception of stops.

The earliest acoustical analyses of consonants were attempted in the 1870's by Grassmann, Michaelis, Trautmann, and others. These investigators carefully separated two types of cue: a) the nature of the "noise", and b) the "characteristic pitch" (*Eigenton*) of the consonant. Not much could be done about describing the former since the necessary experimental equipment was lacking. Data on the "characteristic pitch", however, were published by a number of investigators, beginning with Grassmann.[23] It is fairly obvious from this material that the "characteristic pitch" refers to the pitch of the second formant of the adjacent vowel, probably at the terminal (beginning or end) point of its transition. This interpretation explains the remark of Lloyd that the "characteristic pitch" of a consonant can vary considerably without in the least affecting the identifiability of the consonant.[24] It also explains the observation of the Russian phonetician, Thomson, who insisted that the difference between Russian "hard" and "soft" consonants lay in the fact that the latter had a "higher characteristic pitch" than the former.[25] The "hub" and the "locus"[26] – recently advanced as primary cues for the perception of consonants – were thus anticipated by workers active at the end of the nineteenth century.[27]

---

notes by which Helmholtz specified the formants; no importance should be attached to the fact that they are given to four places.

[23] H. Grassman, "Über die physikalische Natur der Sprachlaute," *Annalen der Physik und Chemie*, series 3, *1*, 623–639 (1877).

[24] "...When freed from connection with any vowel, the resonance of *f* can be carried a long way both up and down in pitch, without at all spoiling the *f* itself. It becomes clear that the essential quality of *f* is but vaguely linked with the actual pitch of its resonance." R. J. Lloyd, "On Consonant Sounds," *Proceedings of the Royal Society of Edinburgh*, *22*, 224 (1898).

[25] A. I. Tomson, *Obščee jazykovedenie* (Odessa, 1910), p. 198., Cf. also A. Thomson, "Über die weichen Konsonanten," *Zeitschrift für slavische Philologie*, *8*, 92–101 (1931), and *idem.*, "Die Erhärtung und Erweichung der Labiale im Ukrainischen," *Zapiski ist.-fil. vidd. A.N. Ukr. S.S.R.*, *13–14*, 253–263 (1927).

[26] For "hub" see p. 97 below. "Locus" is a more elaborate definition of the second formant transition used by workers at the Haskins Laboratories; their work is discussed on pp. 106–7 below.

[27] Of interest is also the suggestion of Rosapelly that in the production of a consonant there is always a vowel-like transition, for which he coined the term *vocaloid*. Cf., S. Petrovskij, *Nabljudenija nad prodolžitel'nost'ju i vysotoj proiznošenija zvukov v slove* (Kazan', 1903).

At the turn of the century the abbé J. P. J. Rousselot published his influential *Principes de phonétique expérimentale* (Paris, 1897–1906), in which he described a number of devices used in phonetic analysis, such as the kymograph, palatograms, various types of mechanical wave analyzers, and reported on an impressive array of detailed experiments performed with the aid of these apparatus. Rousselot inspired many other investigators who came to work with him in Paris early in the century. His laboratory became a model for phonetic institutes that were established at different universities at this time, of which we might mention here the laboratories of Panconcelli Calzia in Hamburg, of Scripture in Vienna, and of Ščerba in St. Petersburg.

Rousselot influenced Russian phoneticians, particularly Usov and Ščerba, who were his students, as well as Thomson and Bogorodickij.

Usov's work contains the earliest kymograph traces of Russian vowels.[28] It is open to criticism on a number of points of methodology and experimental techniques and is only of historical interest.

Ščerba's work[29] is a much more substantial contribution. This is due to his skillful use of the kymograph traces for the measurement of duration and relative intensities of vowels, as well as to his utilization of tuning forks and some special devices of his own design for obtaining spectral information on vowels. His treatment of the vowels remains to this day the most extensive that has been published in Russia.

Thomson possessed a truly remarkable sense of pitch and performed his analyses entirely by ear.[30] His determinations of the formant frequencies of Russian vowels are most reliable and were not superseded until the development of the sonagraph made large-scale spectral analysis of vowels a relatively simple matter.

Bogorodickij's publications cover over half a century. He attempted to determine the acoustical properties of Russian speech sounds.[31] Unfortunately Bogorodickij was not too reliable an observer, and acoustical equipment except of a very rudimentary kind was not available to him.

We might mention also the work of the physicist A. Samojlov, who subjected to Fourier analysis the wave shapes of five vowels. His recording equipment was very poor and consequently his results are not very good. He also failed to state whether or not the analyzed vowels were Russian.[32]

In spite of the relatively great activity, the advances in understanding the acoustical properties of speech were not very marked. Great hopes were pinned on the Fourier analysis of traces of phonograph recordings, but the results obtained[33] were rather

[28] N. Oussof, "Etudes expérimentales sur une prononciation russe," *La Parole*, *1*, 676–686 and 705–717 (1899).

[29] L. V. Ščerba, *Russkie glasnye v kačestvennom i količestvennom otnošenii* (= *Zapiski istoriko-filologičeskogo fakul'teta imperatorskogo St. Petersburgskogo universiteta*), *107* (1912).

[30] A Tomson, "Fonetičeskie ètjudy," *Russkij filologičeskij vestnik*, *1905*, no. 2, 199–244. See also footnote 25 above.

[31] V. A. Bogorodickij, *Fonetika russkogo jazyka* (Kazan', 1930), and literature cited there.

[32] A. Samojlov, "Zur Vokalfrage," *Pflügers Archiv*, *78*, 1–37 (1899).

[33] Cf. Lloyd, *op. cit.*, and L. Hermann, "Phonophotographische Untersuchungen," *Pflügers Archiv*, 45–150, *passim* (1899–1913).

disappointing. It may almost be said that from a linguistic point of view the only interesting results of acoustical investigations were such by-products as Pipping's acoustical vowel triangle,[34] or Rosapelly's and Thomson's anticipation of the "hub" principle. It is, therefore, not surprising that the question should have been raised among linguists whether acoustical analysis could contribute anything of value to linguistics. The negative side, persuasively argued by Jespersen,[35] carried the day, with the result that for a long time thereafter linguists showed no interest whatever in acoustical studies of speech.

Fortunately interest in the acoustic analysis of speech sounds on the part of non-linguists continued in the twentieth century. Late in the twenties, two significant works were published – *Die Sprachlaute* (1926) by the German psychologist Stumpf, and *Speech and Hearing* (1929) by the American engineer Harvey Fletcher, of Bell Telephone Laboratories. Although only three years separate the publication dates of these two books, the former is in most respects obsolete, while the latter is still one of the fundamental books on the acoustics of speech. The reasons for this are not hard to find.

In 1913 when Stumpf began his investigations his equipment, consisting of elaborate systems of pipes of adjustable length which functioned as a set of acoustic filters, was as modern as any that could be found. By the time he published his results, electronics had progressed so far as to make Stumpf's equipment as obsolete as a horse and buggy. But even more than the technological changes, changes in the intellectual climate have outmoded the many subjective speculations that are such an outstanding feature of Stumpf's book. How far removed Stumpf's approach is from our own can perhaps best be illustrated by examining the following statement in which Stumpf sought to describe the difference between vowels and consonants: "Vokale sind sprachlich herstellbare Klänge oder Geräusche mit ausgeprägter Färbung; Konsonanten sind sprachlich herstellbare Geräusche ohne ausgeprägte Färbung."[36] Disregarding the obvious mistake of equating consonants with noises (which would class sonorants like [m], [l], etc., among the vowels), Stumpf's statement lacks physical meaning. There is no attempt at explaining what "sprachlich herstellbar" might mean, i.e. how we could know whether or not a given sound is "sprachlich herstellbar" In trying to explain what is meant by "ausgeprägte Färbung", Stumpf ranges far and wide over the entire field of perception, comparing sounds of speech to those of musical instruments, drawing analogies with the perception of colors, but again fails to state by what measurement procedure it would be possible to know whether or not a certain sound possesses "ausgeprägte Färbung".

[34] He correctly pointed out that the dimension "close-open" of the vowel triangle was associated with the frequency of the lowest formant. Like Helmholtz he saw that front vowels had a clearly defined second formant, which back vowels lacked. Cf., H. Pipping, "Über die Theorie der Vocale," *Acta societatis scientiarum fennicae, 20:11*, 29–31 (1894)

[35] O. Jespersen, *Phonetische Grundfragen* (Leipzig, 1904), Chs. IV and V.

[36] C. Stumpf, *op. cit.*, p. 100.

It is not, however, my intention to deny that Stumpf's work had considerable merit. He called attention to a number of problems in the analysis of speech sounds which continue to occupy investigators. For instance, it is obvious that not all maxima of a vowel spectrum should be called "formants" (a term originated by Hermann and used constantly since then). Furthermore, not all *bona fide* formants play distinctive roles in characterizing the differences between vowels. It would hence be convenient to be able to categorize formants as major or minor. Stumpf was the first who faced this problem squarely. Many of Stumpf's ideas showed great insight, and it is to be regretted that no one has tried to reformulate Stumpf's statements in terms that would meet modern standards of scientific objectivity and rigor. His attempts at discovering similarities in various fields of perception might well have led to interesting experiments and to a broadening of our understanding of this complicated process.

In the field of consonant investigation Stumpf was apparently the first to study the effects of filtering on the perception. By means of his acoustic filters he high-passed and low-passed various sounds and noted at what frequencies incorrect identification became frequent. The results of his studies show clearly that continuants are dependent for correct identification on high frequency components. Not much was discovered about properties of other consonants, for in Stumpf's experiments they were either all correctly identified or became all unidentifiable. This is doubtless due to the fact that his filtering was not sharp enough.

In properties other than those of the acoustic spectrum Stumpf showed little interest. He wrote:

> Finally it is not to be overlooked that the differences between consonants generally do not depend only on their acoustic components in the narrow sense of that expression [i.e., on their spectral properties – M.H.]. Differences in their onset, temporal course, duration and intensity are well-known and have always been used primarily for purposes of classification.[37]

Thirty years after the publication of the above, it is difficult to share Stumpf's optimism about the degree to which "differences in onset, temporal course, duration and intensity" are well-known and understood. It is unfortunate that Stumpf should have passed over so carelessly some of the gravest gaps in our knowledge. Stumpf's view was limited by the capabilities of his equipment, and he seems to have been somewhat too hasty in considering as trivial, problems for which his equipment failed to provide an answer.

A comparison between Fletcher's famous *Speech and Hearing* (New York, 1929) and Stumpf's book demonstrates strikingly how completely the introduction of the vacuum tube changed the entire field of acoustic research. In speech studies vacuum tube techniques were first used by I. B. Crandall,[38] one of Fletcher's colleagues at the Bell Telephone Laboratories. These studies mark the beginning of the modern phase in the study of the acoustics of speech. In this period the Bell Laboratories were the

[37] *Ibid.*, pp. 128–129.
[38] I. B. Crandall, "The Sounds of Speech," *Bell System Technical Journal*, *4*, 586–626 (1925), and "A Dynamic Study of the Vowel Sounds," *ibid.*, *6*, 100–116 (1927).

undisputed leaders in the field, and Fletcher's book, which is a summary of the work done there up to 1929, owes much of its value to the many-sided research activities that were supported there in the early period.

Crandall performed Fourier analysis of the wave forms of vowels and other periodic sounds.[39] These were recorded by means of an oscillograph, and the Fourier spectrum computed by standard mathematical methods. The results so obtained showed that the determinations made by such earlier investigators as Helmholtz were remarkably accurate.

Tests somewhat similar to those of Stumpf were also conducted by Fletcher and his colleagues: the speech sounds were subjected to high- and low-pass filtering in order to discover the effects of these types of distortion upon the perception.[40] The results were basically similar to those of Stumpf. They also show interesting parallels in behavior between sounds having similar "points of articulation" These experiments differed from those of Stumpf in two respects: (1) instead of using acoustical filters, electrical filters were employed throughout, and (2) qualitative data about the errors made as a result of the distortions were suppressed in favor of quantitative data which reflected the effects of the particular distortion on the "articulation score" – i.e., on the percentage of correctly identified words in a list which was supposed to constitute a representative sample of the language. The data obtained from amplitude distortions were treated in the same way. It is obvious that this is a much less illuminating way of presenting the results, for it tells us less about the parameters destroyed as a result of a given type of distortion than does the so-called "confusion matrix" in which the nature of the error is recorded as well.

Furthermore, Fletcher published a discussion of the wave shapes of various consonants and reported results of Sacia and Beck[41] on the intensity of the various speech sounds. These measurements are to this day almost the only data we have on the intensities of the consonants. It is therefore unfortunate that the methods used in obtaining these data were not unexceptionable.

The greatest weakness in Fletcher's work lies in his relative lack of sophistication with regard to experiments in which human subjects are involved. The achievements of modern psychology, the discussions as to proper experimental procedures, the developments in the field of linguistics, had but a slight influence on Fletcher and his collaborators.

Two other books of this period should be mentioned briefly. One is *The Vowel* by George Oscar Russell, Director of the Phonetics Laboratories at Ohio State University, which contains a summary of the history of vowel analysis to the twentieth century, covering the main developments in acoustic analysis and vowel synthesis and a study

[39] Crandall was not the first to perform such analyses. He was preceded by several investigators of whom D. C. Miller, *The Science of Musical Sounds* (New York, 1916), is to be noted especially.
[40] Fletcher, *op. cit.*, pp. 400 ff.
[41] C. F. Sacia, "Speech Power and Energy," *Bell System Technical Journal*, *4*, 627–641 (1925); and C. F. Sacia and C. J. Beck, "The Power of Fundamental Speech Sounds," *ibid.*, *5*, 393–403.

of the articulation of speech sounds with the help of x-rays. The other book is Sir Richard Paget's *Human Speech.* Some of Paget's ideas on language and linguistic analysis are bizarre,[42] but his enthusiasm for the analysis and synthesis of speech sounds adequately compensates for this fault. Paget, who had perfect pitch discrimination, determined the formant frequencies of his own vowels and those of French and Russian speakers. For most vowels he lists at least two formants, and his results for English agree fairly well with those of Crandall, whose data he includes in an appendix. He also attempted to analyze some consonants by ear, but his results were not conclusive. In addition, he built a number of plasticene models of the vocal tract for vowel synthesis, thus carrying on the type of experiment begun by Kratzenstein and Kempelen.

At the end of the 1920's devices were constructed for the automatic frequency analysis of sounds. These devices were apparently developed simultaneously in Germany and in the United States.[43] Some of these devices consisted of filters of fixed bandwidth whose pass bands covered the entire spectrum. When these filters were put into operation simultaneously, the records for the whole filter bank gave an instantaneous frequency analysis of the signal. In the 1940's the principle of heterodyning was applied to this problem – i.e. a filter of fixed bandwidth was used in conjunction with a variable oscillator and modulator system by which any portion of the sound spectrum was brought within the frequency range of the filter.[44] In spite of the availability of adequate recording techniques in the 1930's, it was not until the end of the decade that automatic analyses were generally performed from recordings instead of from a human subject who was required to intone the same vowel for a long time.[45]

Thanks to these technical advances it became possible to study connected speech. The signal could be split into short consecutive segments, and the Fourier spectrum of these segments could be calculated. The resulting series of spectra represented the speech signal as a function of both frequency and time. One of the earliest portrayals of the speech signal as a function of both frequency and time was produced by J. C. Steinberg of Bell Telephone Laboratories.[46] From an oscillogram of the speech wave, he calculated the spectrum of each pitch period in the vowels, sonorants, and voiced consonants. He presented his results in a two-dimensional plot quite similar in appearance to the Visible Speech sonagram.

E.g., his views on the origin of language, which, however, were taken seriously by as important a worker as Fletcher,

[43] M. Gruetzmacher, *Elektr. Nachr. Techn.*, *4*, 553 (1927); C. R. Moore and A. S. Curtis, "An Analyzer for the Voice Frequency Range," *Bell System Technical Journal*, *6*, 217–229 (1927); E. Gerlach, *Z. techn. Phys.*, *8*, 815 (1927).

[44] W. Koenig, H. K. Dunn, and L. Y. Lacy, "The Sound Spectrograph," *JASA*, *18*, 21 (1946).

[45] Cf., e.g., the paper of E. Thienhaus and L. Barczinski, "Klangspektren and Lautstärken deutscher Sprachlaute," *Archives néerlandaises de la phonétique expérimentale*, *11*, 47–69 (1935).

[46] J. C. Steinberg, "Application of Sound Measuring Instruments to the Study of Phonetic Problems," *JASA*, *6*, 16–24 (1934).

Fourier analysis of wave shapes of speech, primarily of vowels, was also performed by Thienhaus,[47] Gemelli and Pastori,[48] and Hála.[49] These investigations brought little that was new in either methodology or technique, and their main importance lay in the fact that they confirmed and extended the results obtained by Fletcher and his collaborators.

Shortly after the war the *sound spectrograph* or *sonagraph* was made commercially available. This device portrays an acoustic signal as a function of time, frequency, and intensity.[50] Since its operation is a relatively simple matter, it has become standard equipment in phonetics laboratories. It has acquainted many linguists with the elementary facts of the acoustics of speech, and has reawakened interest in speech acoustics among linguists.

In the initial paper on this device[51] Kopp and Green showed how certain classes of speech sounds could be recognized by their characteristic patterns on the sound spectrogram. For example, vowels and sonorants are marked by the presence of a number of heavy horizontal lines, the formants; stops are recognized by a vertical white area (silence) followed by a vertical line (burst); fricatives, by irregular vertical striations. The authors also called attention to the transitions between vowels and consonants, and showed certain regularities in the behavior of the second formant (consonant "hub"). Finally, they tried to draw conclusions about the energy distribution in the consonants proper. It is strange that no attempt at automatic recognition of at least the clearest of these distinctions was made.

The first linguist to employ the sound spectrograph extensively in a detailed analysis of speech was Martin Joos.[52] His book is valuable in that it contains a very simple and understandable discussion of the acoustical principles underlying speech. It does good service as an introduction for linguists who are not acquainted with such standard terms as *harmonic, fundamental, formant,* etc. Unfortunately, Joos did not utilize the instrument in the manner in which it would have provided him with the most interesting set of answers. His exclusive use of the narrow band (45 cps) filter, which he justifies by unproven neurophysiological speculations, obscured important properties of vowels and made the study of consonants well nigh impossible.

Interesting information about the acoustical properties of speech was obtained by investigations into the perception of distorted speech. Studies of the perception of

[47] Thienhaus and Barczinski, *op. cit.*

[48] A. Gemelli and G. Pastori, "Quelques recherches sur la structure des consonnes," *Archives italiennes de biologie, 92,* 97–126 (1934). See also A. Gemelli and G. Pastori, "L'analisi elettroacustica del linguaggio," *Università Cattolica del S. Cuore Pubblicazioni,* series VI (biological sciences), Vol. II (Milan, 1934).

[49] B. Hála, *Akustická podstata samohlásek* (= *Rozpravy české akademie věd a umění,* Class III, no. 78) (1941).

[50] A description of this device is given in *Technical Aspects of Visible Speech* (= *Bell System Technical Monograph* B-1415) (1946), and *JASA, 18,* 1–89 (1946). Cf., also Chapter III, sec. 2 above.

[51] G. A. Kopp and H. C. Green, "Basic Phonetic Principles of Visible Speech," *JASA, 18,* 74–89 (1946).

[52] M. Joos, *Acoustic Phonetics* (= *Language Monograph* no. 23) (Baltimore, 1948).

filtered speech had been made by various investigators, beginning with Helmholtz. During the last war the need for reliable military communications stimulated a number of inquiries into the perception of speech which had been subjected to different types of distortion common in poor telephone and wireless links. The most important fact to emerge from these studies was that speech remained surprisingly intelligible under very severe distortions. Only a few types of distortions, such as high pass filtering and "center clipping" – i.e., distortions which affect most radically the low frequency region – impair intelligibility completely.[53]

In these studies intelligibility was conceived as a gross property of speech and was measured by means of "articulation scores", a measure due to Fletcher and his collaborators, which, however, is uninformative from the point of view of phonetics, since it tells us nothing about the effects of a given type of distortion on individual acoustic features of speech. In recent years G. A. Miller has repeated some of these experiments with the purpose of investigating the qualitative effects produced by different kinds of distortion.[54] Accordingly he presented his data in "confusion matrices"; i.e., in tables showing the mistakes made by the listener for each condition of distortion. He has shown that different types of distortion affect individual distinctive features differently, and has suggested that these facts be taken into consideration when devising special "higly intelligible" vocabularies. Miller has also attempted to express intelligibility as a function of the correct reception of individual distinctive features.

Very important for the study of the acoustical properties of speech have been studies of the relationship between articulation and acoustical output. In principle, it is possible to compute mathematically the acoustical output of any cavity system, given its geometric configuration and the location and nature of the sources which excite the system. Calculations of this sort were carried out by Helmholtz, Crandall, Benton,[55] and several other investigators. The agreement between calculated and observed values was often poor because of the many approximations and short cuts which had to be resorted to, owing to the labor involved in the computations. In addition the investigators were frequently not too well acquainted with the geometry of the vocal tract, since good x-ray pictures of speech articulations did not become common until the 1930's.

The first to tackle the problem in all its complexity were T. Chiba and M. Kajiyama in their monograph on the vowel.[56] From radiographs and palatograms these authors

[53] For a very readable summary of the results of these experiments, see J. C. R. Licklider, "The Manner in Which and the Extent to Which Speech Can Be Distorted and Remain Intelligible," in H. von Foerster, ed., *Cybernetics*: *Transactions of the Seventh Conference* (New York, 1950), pp. 58–122.

[54] G. A. Miller and P. E. Nicely, "An Analysis of Perceptual Confusions Among Some English Consonants", *JASA*, *27*, 338–352 (1955), and G. A. Miller, "The Perception of Speech," in *For Roman Jakobson* ('s-Gravenhage, 1956), pp. 353–360.

[55] W. E. Benton, "A Note on the Double Resonator Theory of Vowel Sounds," in Paget, *op. cit.*, pp. 275–298.

[56] T. Chiba and M. Kajiyama, *The Vowel, Its Nature and Structure* (Tokyo, 1941).

obtained exact measurements of the vocal tract of a speaker articulating the six Japanese vowels. They then calculated mathematically the acoustical outputs of the six cavity configurations and compared the calculated values with those actually measured. The agreement between the two sets of data was fairly good, which shows that the methods and assumptions involved in the calculations were basically sound.

Chiba and Kajiyama's work also illustrates some of the difficulties inherent in their method. The greatest of these is the large amount of work required for the computation. Though the equations have all been solved in principle, the numerical solutions of actual cases are so time-consuming as to make imperative a number of not altogether desirable simplifications. Another disadvantage is that the results of the computations – i.e., the actual sounds – cannot be easily subjected to auditory test, which are of obvious importance in this kind of work.

These disadvantages are overcome in the various analog devices for the production of speech, all of which can be considered as computers for the generation of sounds from data obtained by measuring various dimensions of the vocal cavity. And it is in this sense mainly, and not as speaking gadgets, that analog devices are of interest to students of speech. Devices like those of Kempelen and other investigators are, from this point of view, functional and anatomical analogs of the vocal tract. In the early devices the different resonating cavities were simulated by actual vessels made of clay, rubber, or some other material, and the excitation was provided by some mechanical sound source, most commonly a reed of some kind. The analog devices developed in the last thirty years differ from the former in that the resonators and the excitation are simulated by electrical circuits. The output of such a circuit is equivalent to the output of the corresponding acoustical system. The great advantage of all analog devices is that the output is accessible to immediate checking by ear. In the case of the electrical analog there is also the added advantage that the spectrum of the output can be easily measured by standard techniques and need not be laboriously computed.

The various analog devices constructed in recent years differ from one another in the extent to which details of the speech-producing apparatus are simulated. The fewest anatomical details were analoged in the Bell Laboratories' Voder.[57] In the Voder there was a choice between two excitations (voicing or noise) which were modulated by a limited number of pre-set spectral envelopes, that simulated the resonance properties of the vocal tract in articulating the individual speech sounds. In H. K. Dunn's vocal tract[58] the three independent variables were those in terms of which vowels are normally described: the size of the mouth opening and the position and the extent of the tongue constriction.

Dunn's analog incorporated several undesirable simplifications, viz., the tongue constriction was treated as a lumped, rather than as a distributed, element, and the

[57] H. Dudley, R. R. Riesz, S. S. A. Watkins, "A Synthetic Speaker" (Voder), *Journal of the Franklin Institute*, *227*, 739–764 (1937).

[58] H. K. Dunn, "The Calculation of Vowel Resonances and an Electrical Vocal Tract," *JASA*, *22*, 740–753 (1950).

unconstricted part of the vocal tract was assumed to have a constant cross section. The electrical vocal tract analogs (EVTA) due to Fant and to Stevens *et al.* simulate anatomical features in greater detail, representing the cavity by a network equivalent consisting of more than thirty cylindrical sections of variable diameter and height.[59] This gives the EVTA enough flexibility to eliminate the need for any undesirable simplifications.

For practical research purposes the increase in independently variable quantities proved, however, only a partial advantage, since it made the utilization of the tract a very complex and unwieldy affair. In their research with the EVTA Stevens and House[60] found it necessary to reduce the number of independent variables to the same three that were incorporated in Dunn's analog. The flexibility of their equipment made it possible for them to improve on Dunn's analog by treating the tongue constriction as a distributed element and to approximate more closely the unconstricted portions of the vocal tract. Stevens and House studied the changes in the position of the three lowest formants which were produced by varying the three mentioned variables. They also extended their study to include data of formant positions associated with total closure of the tract, from which inferences can be made about the formant transitions in vowels adjacent to stops and fricatives. Finally they investigated the nature of nasalization by simulating a nasal cavity, which was switched into the EVTA at a point corresponding to the nasal pharynx. An important feature of Stevens and House's work was the perceptual testing of the synthetic stimuli. These experiments established that the variables of the electrical analog produced the same perceptual effects as the articulatory variables (rounding, fronting, nasalization, etc.) that were being simulated.

In a series of studies extending over a decade Fant has investigated the acoustical correlates of most articulatory features known to be distinctive in language.[61] As a result of his work, this problem appears to be well on the way towards a definitive solution. Among the many new results obtained by Fant, the following are the most important:

Fant has shown that vowel spectra can be completely specified if the frequencies of the four lowest formants are given.

He has materially advanced our understanding of the production mechanism of fricative consonants. He has shown that for sounds like [s] it is appropriate to assume

[59] K. N. Stevens, S. Kasowski, C. G. M. Fant, "Electrical Analog of the Vocal Tract," *JASA*, *25*, 734–742 (1953), and C. G. M. Fant's device in operation at the Royal Institute of Technology in Stockholm.

[60] K. N. Stevens and A. S. House, "Development of Quantitative Description of Vowel Articulation," *JASA*, *27*, 484–493 (1955); *idem*, "Auditory Testing of a Simplified Description of Vowel Articulation," *JASA*, *27*, 882–887 (1955); *idem*, "Analog Studies of the Nasalization of Vowels," *Journal of Speech and Hearing Disorders*, *21*, 218–232 (1956); and A. S. House, "Analog Studies of Nasal Consonants," *Journal of Speech and Hearing Disorders*, *22*, 190–204 (1957).

[61] All statements regarding Fant's work are made on the basis of a manuscript of his *Acoustic Theory of Speech Production* ('s-Gravenhage, in press), which, thanks to his kindness, I have been able to examine in detail.

a source located at the teeth and that the cavities behind the tongue constriction have only minor effects on the acoustic output.

He has demonstrated that the terminal point of a vowel formant transition (i.e. the "hub" or "locus") is more radically affected by palatalization in the adjacent consonant than by differences in the point of articulation of the consonant.[62]

Another approach to the study of the invariants of speech is by systematic variation of selected acoustical cues, which then are presented to listeners who are instructed to judge them as sounds of speech. This method of attack goes back to the very beginnings of speech investigations; e.g., the experiments of Hooke and of Willis were essentially of this kind. In recent times a series of interesting experiments has grown out of an attempt to discover which cues present in a Visible Speech display are essential to comprehension and which cues can be omitted without serious impairment of intelligibility. The group at the Haskins Laboratories, who carried out these experiments, first developed a device capable of converting Visible Speech patterns back into sound.[63] After preliminary experimentation it was found possible to produce satisfactory synthetic speech from greatly simplified patterns prepared in the laboratory. A systematic study was then undertaken to test the role of individual variables in the perception of speech.

The procedure was generally as follows: Artificial patterns were prepared and played back on the play-back device. The output of the latter was recorded and a test tape of such stimuli in random order was prepared. The stimuli were presented to a group of listeners who were instructed to identify these as words or syllables of a natural language. In most of the tests the subjects were given a response vocabulary from which to choose their answer, and no other answers were allowed.

The results obtained by the Haskins group have deeply influenced the thinking of almost everyone working on the acoustics and psycho-acoustics of speech. Above all, their work has served to establish the formants as the most important class of cues in the perception of speech. Thanks mainly to the experiments of these workers we now know that formants are not only sufficient cues for vowel identification, but also that the changes in formant position (transitions) play an exceedingly important role in the perception of many types of consonant.

The most important experiments of the Haskins groups have to do with the perception of stops.[64] The results of these experiments caused some anguish among linguists, since they seemed to indicate that identical responses can be elicited by different acoustical cues, and that apparently identical acoustical cues can consistently elicit different responses. These conclusions are based on the following facts:

[62] See Table V-5 below.

[63] F. S. Cooper, J. M. Borst, A. M. Liberman, "The Interconversion of Audible and Visible Patterns as a Basis for Research in the Perception of Speech," *Proceedings of the National Academy of Sciences*, *87*, 318–325 (1951).

[64] F. S. Cooper *et al.*, "Some Experiments in the Perception of Synthetic Speech Sounds," *JASA*, *24*, 597–606 (1952); and by the same authors, "The Role of Selected Stimulus Variables in the Perception of Unvoiced Stop Consonants," *American Journal of Psychology*, *65*, 497–516 (1952).

1) Stop responses were elicited not only by short bursts of sound, but also by rapid variations (transitions) of the formant frequencies in the terminal (beginning or end) portions of vowels.

2) In experiments in which the stop cues were short bursts of sound centered at different frequencies, it was found, e.g., that a burst centered at 1440 cps elicited a majority of /k/ responses when adjacent to the vowel /ɔ/, but a majority of /p/ responses when adjacent to the vowel /i/.

3) The transitional cues eliciting particular stop responses – described by the so-called "locus" hypothesis[65] – have from an acoustical point of view very little in common with the bursts producing the identical response.

These facts, however, do not exhaust the information that is relevant here. It should not be overlooked that in spite of clearly apparent differences, rapid formant transitions and stop bursts are quite akin to each other. An instantaneous transition followed by a stationary vowel formant is identical with a burst followed by a stationary vowel formant. Moreover, when reproduced on the Haskins Pattern Playback, even moderately rapid transitions cannot be differentiated from bursts followed by stationary formants.[66]

It is not surprising that one and the same stimulus elicits different responses in different context. The only conclusion to be drawn from the cited evidence is that the burst is not an independent variable, and the correct statement of the desired invariance must include a reference to the context. As I have proposed elsewhere,[67] the correct formulation of the observed behavior of listeners should read as follows: "/k/ judgements are correlated with a concentration of energy (i.e., placing the center frequency of the burst) in the region somewhat above the second formant of the adjacent vowel; /p/ and /t/ judgements are correlated with an absence of energy concentration in this region. In the latter case, /t/ judgements predominate when energy is concentrated in the high frequencies; /p/ judgements predominate when energy is not concentrated in the high frequencies." This formula covers also the observations embodied in the "locus" hypothesis, which, however, is to be expected in view of the essential similarity of bursts and transitions.

In the last decade acoustical measurements of speech sounds, particularly of vowels, have become very common. The majority of these measurements, however, have been made on single utterances of one or two speakers, and the results obtained have served mainly to corroborate previously known facts. Of more general interest are the results of studies in which a relatively large number of vowels was measured.[68] As might have been expected, these studies showed great variability in the data. Different

[65] P. C. Delattre, A. M. Liberman, and F. S. Cooper, "Acoustic Loci and Transitional Cues for Consonants," *JASA*, *27*, 769–773 (1955).

[66] Private communication from Fant, July, 1956.

[67] M. Halle, Review of C. F. Hockett, *Manual of Phonology*, *JASA*, *28*, 509–511 (1956).

[68] G. E. Peterson and H. L. Barney, "Control Methods Used in a Study of the Vowels," *JASA*, *24*, 175–184 (1952). C. G. M. Fant, *Analys av de svenska vokalljuden* (= *L. M. Ericsson protokoll H/P – 1035*) (1948).

utterances of the same vowel (even by one and the same speaker) had widely divergent formant frequencies. On the other hand, in some instances of perceptually different vowels, the formant frequencies were almost identical. As yet no generally accepted explanation has been offered for these observations. The following suggestions may merit further investigation.

In identifying vowels in connected discourse, the listener is provided with more information than the frequencies of the lowest two or three formants. It has, therefore, been suggested that the listener uses the additional information in order to readjust his identification criteria. Fant has proposed correction factors which take into account the age and sex of the speaker.[69] Ladefoged and Broadbent have attempted to show by experiment that the formant frequencies of the vowels in an introductory phrase provide the listener with a sort of reference against which he judges all vowels that come later in the utterance.[70] Their idea is based on a suggestion made by Joos that the function of phatic phrases like "how do you do" is to enable the listener to establish a sort of reference vowel triangle in terms of which he can identify his interlocutor's vowels.[71]

The above explanations all take for granted that formant frequencies are the significant variables in vowel perception. It is, however, conceivable that the difficulties which have been brought to light in the experiments reviewed above are primarily due to the fact that a poor set of variables was selected and that they would largely disappear if a better set were to be chosen. Fant has shown that formant frequencies determine completely the spectral envelope of vowel sounds. From this it does not, however, follow that vowels are perceived in terms of their formant frequencies. It may well be that the process of perception involves some nonlinear transformation of the spectral envelope which cannot be simply expressed in terms of formant frequencies. The difficulty with this suggestion is that before it can be experimentally tested, a specific set of variables has to be postulated. Unfortunately there are many candidates, and at present we possess no information that would enable us to choose one set over the rest. A basic inquiry is therefore called for.

The present decade has also seen the first publication of reliable measurements of acoustic properties of consonants. In the late 1940's Fant investigated the Swedish consonants with specially designed filtering equipment and produced the first spectra of stop consonants.[72] An investigation into the acoustical properties of Danish consonants is the work of Miss E. Fischer-Jørgensen.[73] This study contains many valuable observations about the difference between tense and lax stops, and about the role of formant transitions in vowels adjacent to stop sounds. The data on spectral

[69] C. G. M. Fant, *Transmission Properties of the Vocal Tract with Application to the Acoustic Specification of Phonemes* (= *M.I.T. Acoustics Laboratory Technical Report* No. 12) (1952).
[70] P. Ladefoged and D. E. Broadbent, "Information Conveyed by Vowels," *JASA*, *29*, 98–104 (1957).
[71] Joos, *op. cit.*, p. 61.
[72] C. G. M. Fant, *Analys av de svenska konsonantljuden* (= *L. M. Ericsson protokoll H/P – 1064*) (1949).

properties of stop bursts are, unfortunately, not reliable, since they were obtained by estimates from broad band sonagrams. A large body of data on the acoustical properties of English fricative and stop consonants was published by me in joint papers with G. W. Hughes and J.-P. A. Radley.[74] The methods used in the collection of these data were essentially similar to those described in Chapter V below.

[73] E. Fischer-Jørgensen, "Acoustic Analysis of Stop Consonants," *Miscellanea Phonetica*, *2*, 42–59 (1954).

[74] G. W. Hughes and M. Halle, "Spectral Properties of Fricative Consonants," *JASA*, *28*, 303–310 (1956); and M. Halle, G. W. Hughes and J.-P. A. Radley, "Acoustic Properties of Stop Consonants," *JASA*, *29*, 107–116 (1957).

CHAPTER V

THE ACOUSTICAL CORRELATES OF THE DISTINCTIVE FEATURES OF RUSSIAN

In Part I we reviewed various aspects of the phonic side of Russian and attempted to show that the distinctive features form a reasonable and adequate framework for the description of these facts. The purpose of the experimental work reported in this chapter is to establish the acoustical correlates of the distinctive features of Russian. This is a necessary step in a description of a language that pretends to completeness; for it establishes the link between language viewed as an abstract system and language as a physical phenomenon. In the last analysis language is sound, and a description of language that omits a discussion of its acoustical properties is incomplete.

In *Preliminaries* tentative statements were made about the acoustical correlates of the distinctive features. These statements had to be general and often vague, for at that time our investigations had just begun. The work reported here is considerably more specific. Measurement procedures are outlined by means of which individual features can be identified. These identification procedures were tested on a fair-sized corpus, and the results of the tests are given in the text, thereby enabling the reader to judge the reliability of the proposed criteria. The identification procedures are often excessively cumbersome and perhaps even arbitrary. It is my conviction, however, that procedures essentially similar to those proposed below will ultimately be shown to be appropriate for the acoustical description of the elementary information bearing signals of language.[1]

1. Subjects and Material

The subjects were the following four Russian speakers, all natives of the city of Moscow:

For further work in this direction, see G. W. Hughes and M. Halle, "Spectral Properties of Fricative Consonants," *JASA*, *28*, 303–310 (1956), and M. Halle, G. W. Hughes, and J.-P. A. Radley, "Acoustic Properties of Stop Consonants," *JASA*, *29*, 107–117 (1957).

K, female, born about 1895; of an intellectual family; left Moscow in the early 1920's; habitual language, Russian; occupation, housewife.

D, male, born about 1910; left Moscow in the early 1940's; habitual language, Russian; occupation, actor, formerly with the Moscow Art Theater.

I, male, born about 1895; left Moscow in the early 1920's; habitual language, Russian; occupation, painter and poet.

J, male, born about 1895, left Moscow in the early 1920's; one of the habitual languages, Russian; occupation, university professor.

The material consisted of the following:

a) A list of Russian syllables composed of a vowel preceded or followed by a consonant. All syllables of this structure admitted in Russian were recorded and investigated.

b) A list of polysyllabic words in which all variations in the unaccented vowels with respect to the positions of the word accent were exemplified.

c) Connected discourse, used primarily for the study of certain junctural phenomena and for purposes of comparison.

Objections have been raised against the use in phonetic investigations of materials like the above on the grounds that in normal discourse both distinctive and redundant features and even entire phonemes are often obscured or altogether omitted. Material like the above, in which an attempt is made to pronounce carefully all relevant cues, is, therefore, said to be highly artificial, "spoonfed"

These objections are not quite relevant in the case of the present investigation, the purpose of which is to state the acoustical correlates of the distinctive features of the language as they are implemented in the different phonemes. As was pointed out in Part I, the relation between phonemic symbol and acoustical event is an indirect one, mediated through a series of fairly complex rules. As a consequence of these rules, in many utterances individual distinctive features are omitted or replaced by others, and an acoustical analysis of such an utterance could do no more than show this omission or substitution. It follows, therefore, that if we are interested in a particular feature, we should investigate samples in which this feature is known to be present. Isolated CV and VC syllables in which contextual influences are minimized constitute, from this point of view, the most suitable material. Only after all features have been described fully, is it possible to discuss the various modifications due to the operation of the phonetic rules.

The oft-cited evidence that speakers and listeners are not aware of the operation of these modifications, that somehow they are able to reconstitute the correct utterance even in the absence of a great many relevant cues, is really beside the point here. This ability is due to the knowledge that the speaker and listener have of the grammar of the language, of its statistical structure, and of the external circumstances in which the given speech event is uttered. It is not due to information conveyed by the acoustical signal, and is, therefore, of secondary interest in a study devoted to the acoustical specification of the distinctive features.

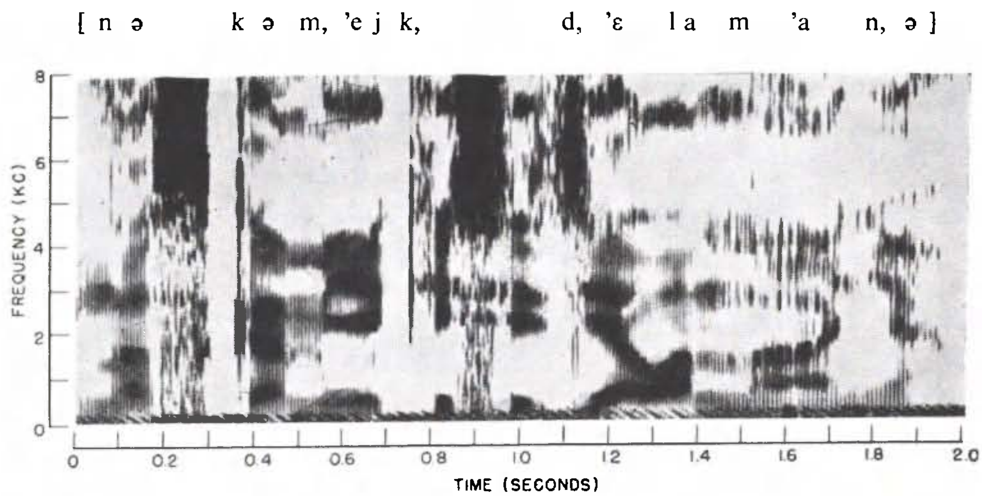


Fig. V-1. Sonagram of connected discourse. }na =*ska*m'ejk+e s,i*d'el+a m'an,+a{ Speaker X. The vowels are marked by the dark, sharply defined, horizontal bars (formants); cf. the formants beginning at 0.1; 0.4; 0.55; 0.82; 0.97; 1.15; 1.33; 1.52; 1.8 secs. The nasals and liquids, on the other hand, have much more weakly defined formant regions; cf. the nasals at 0; 0.45; 1.4; 1.72 secs, and the /l/ at 1.25 secs. The /l/ shows a reduction in intensity of the second and, more strikingly, of the third formant as compared with that of the adjacent vowels. Unlike the vowels, the nasals have almost stationary formants. In the vowels the formant transitions serve as cues for the adjacent consonants. Note in particular the characteristic coming together of formants 2 and 3 adjacent to /k/ at 0.4 and 0.68 secs. The fricatives are represented in the sonagram by irregularly striated areas primarily in the high frequencies; cf. 0.175 and .85 secs. The stops are marked by the absence of energy in the frequencies above the voicing components followed by a sharply defined vertical line — the explosion; cf., the "silences" at 0.3 and 0.7 secs. The regular vertical striations are due to the vibrations of the vocal cords. At the end of the sentence there is a very marked drop in pitch, which is shown in the sonagram by the wider spacing of the striations at 1.75 secs than at 0.1 sec. Note also the total absence of any pauses between words.

2. *Measurement Procedure*

The data on which the following discussion is based were obtained by means of electrical measurements. These measurements were conducted mainly in the Research Laboratory of Electronics of the Massachusetts Institute of Technology. The most important sources of the data were Visible Speech sonagrams, power spectra obtained with a special technique, and comparison measurements of power in various broad frequency regions.

2.11 Since Visible Speech sonagrams have been described above,[2] nothing further shall be said about them here. It is, however, necessary to discuss briefly the procedure used in reading off the vowel formant frequencies from sonagrams, since in most typical samples of vowels, particularly of vowels in consonantal contexts, the formants are not stationary for their entire duration. (Cf., Fig. V-1.) If in such cases there was a portion in which the formants were stationary, this segment was chosen as representative of the vowel and its formant frequencies were recorded. If there was no steady state portion – as, e.g., in the [ɨ] variant of {i} – a point in the middle of the vowel interval was chosen and the formant frequencies were read off there.

Difficulties of another sort arise from the fact that it is sometimes hard to decide which marks on the sonagram represent a formant. When two formants are close together they may appear on the sonagram as a single black horizontal area. Because of the limited amplitude range of the sonagraph, a weak formant may not appear on the sonagram altogether. Since only two fixed bandwidth filters are at our disposal, spurious formants may appear for certain speakers. From theoretical considerations it is known that the human vocal tract in articulating nonnasal vowels will produce three formants in the region below 3000 cps, of which the third will be above 1500 cps. By taking account of this fact, it is possible to locate the "true" vowel formants in most instances.[3]

Little use was made of the sonagraph sectioner, which gives a frequency vs. intensity display of a selected time interval of the signal. It was felt that the heavy investment in time required for making sections was hardly justified by the relatively small amount of relevant information about vowels that is obtained from sections in addition to that contained in a wide band sonagram.

2.12 For the measurement of consonant spectra the sonagraph proved inadequate because of its limited dynamic range and the great uncertainty in locating the exact time interval over which the sectioner samples the signal. The procedure in measuring consonant spectra was the following:

(1) The sample studied, which was usually a syllable containing a vowel preceded or followed by a consonant, was recorded onto an endless loop of magnetic tape. Special care was taken to minimize distortions in the recording and play back stages.

[2] Cf., Chapter III, pp. 87-88 above and literature cited there.

[3] C. G. M. Fant, "On the Predictability of Formant Levels and Spectrum Envelopes from Formant Frequencies," *For Roman Jakobson* ('s-Gravenhage, 1956), pp. 109–119.

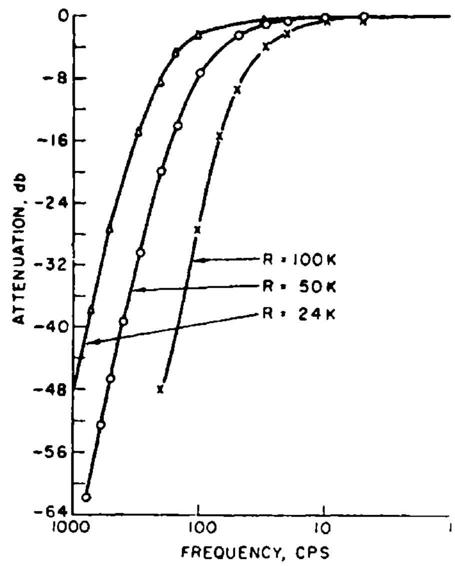


Fig. V-2. Response curves of the Hewlett-Packard 300 A Wave Analyzer.

Ampex tape recorders with substantially flat (+2db) responses up to 10 kc were used throughout the study.

(2) Since it was evidently desirable to study separately the spectrum of the consonant and that of the vowel, rather than average the two together, a gating circuit was constructed. This device passes only the part of the sound that lies between the opening and closing of the "gate". Since the opening and closing can be controlled with great precision it is possible to select a portion of any desired length anywhere in the syllable. The decision as to where to open and close the gate was made after inspecting sonagrams and an oscilloscope display of each syllable, as well as listening to the gated portion.

(3) The gated sample was next passed through a filter. For spectrum measurements we used a Hewlett-Packard Wave Analyzer, model 300A, whose filter had a pass band that could be adjusted to any desired value between 50 cps and 300 cps (cf., Fig. V-2). The center frequency of the filter could be varied between 0 and 16,000 cps, thus making it possible to cover whatever frequency range was of interest.

For our comparison measurements we used a Spencer-Kennedy model 302 electronic filter. The Spencer-Kennedy filter was always used as either a high-pass or a low-pass filter. This made it possible to cascade the two sections of the filter and to obtain an attenuation of 36 db per octave.

(4) The filtered output was then passed through a calibrated attenuator, amplified and fed into a square law device.[4]

[4] The square law device used by us was developed by J. S. Rochefort; see his *Design and Construction of a Germanium Square-Law Device*, M. S. Thesis in the Department of Electrical Engineering, Massachusetts Institute of Technology, 1951.

In most previous investigations the filtered signal was rectified rather than actually squared. It was felt originally that it was important to have a true power spectrum and, therefore, all signals were squared. It has since been shown that for the signals of interest here there is no great difference in the spectra obtained by full wave rectification and those obtained by squaring.[5]

(5) The squared signal was passed through an RC integrator whose output was kept constant for all measurements by adjusting the calibrated attenuator (step 4 above). The readings of this attenuator are the recorded data. This way of measuring has the advantage that only the attenuator needs to be a high precision instrument.

Fig. V-3 presents the measurement procedure outlined in the preceding pages, in the form of a block diagram.

Fig. V-3. Block diagram of the measurement procedure.

The length of the sample measured and the bandwidth of the filter require some discussion. While on the one hand it is desirable to use as short a sample as possible in order to note all temporal changes in the spectral distributions of energy, we also want to make as exact frequency measurements as possible. These two requirements are contradictory in the sense that exact frequency measurements can only be made on long samples. It is, therefore, necessary to make a compromise between these two requirements.

In the case of all consonants except the stops, we chose a sample 50 msecs long. This segment was usually selected somewhere in the middle of the sound. A sample 50 msecs long can safely be subjected to filtering with a 50 cps filter. The stops, however, presented special problems because stop bursts often last 10 or 15 msecs. A broader filter was, therefore, used for stops (120 cps), and the gate set so as to include all of the burst or the first 20 msecs of it, whichever was shorter.

3. Vocalic vs. Nonvocalic, and Consonantal vs. Nonconsonantal

The fundamental oppositions of the Russian phonemic system are the oppositions vocalic vs. nonvocalic and consonantal vs. nonconsonantal. These two features divide the Russian phonemes into the following four classes: (1) The vowels, which are vocalic and nonconsonantal. (2) The consonants, which are consonantal and nonvocalic. (3) The liquids ({l} {l,} {r} {r,}) which are vocalic and consonantal. (4) The glide {j}, which is non-vocalic and nonconsonantal.

The following definitions of these features were arrived at after a thorough exami-

[5] Hughes and Halle, *op. cit.*

nation of the available evidence. Since, however, it was not possible to test them in all respects, they cannot be viewed as definitive formulations.

Vocalic vs. Nonvocalic: Vocalic phonemes are characterized by a spectrum having formant structure and a major concentration of energy in a specific frequency region (somewhere between 300 and 800 cps).

Consonantal vs. Nonconsonantal: Consonantal phonemes are characterized either by an absence of a formant structure or by a discontinuity in the transitions of the formants to and/or from the adjacent phoneme. If a consonantal phoneme is adjacent to a vowel, the consonantal phoneme has the lower overall level.

In sounds possessing "formant structure" the energy is concentrated in a small number of narrow frequency bands.

It will be noticed that the lack of formant structure is a sufficient condition for classifying a phoneme as a consonant; i.e., as consonantal and nonvocalic. Since the absence or presence of formant structure can be easily determined by an examination of the sonagram, nothing further shall be said at this point about consonants whose distinctive property is lack of formant structure, i.e. the stops, the fricatives, and the affricates.[6] (See Fig. V-1.)

We must now investigate more closely the remaining phonemes with particular reference to the definitions given above. All these phonemes are, according to one phonetic terminology, "sonorant"; i.e., have formant structure. It is convenient to consider individually the following classes: vowels; {l} and {l,}, {r} and {r,}; the glide {j}; and the nasal consonants.

The vowels as a class are characterized by the following properties:

(1) All vowels have formant structure.

(2) The higher formants in vowels are less attenuated than in other phonemes possessing a formant structure.

In all of these phonemes the first formant contains a very large fraction of the total energy. For all formants except the first, the higher the formant frequency, the greater is its attenuation.[6] In vowels, however, the attenuation of the upper formants is less than in other phonemes. This property of vowels has been used in several devices for the automatic recognition of vowels.[7]

To test this property further the following measurements were made: In 50 msec long, gated portions of vowels and of the period of oral closure in nasal consonants, the energy below and above 300 cps was measured. This value was chosen because in vowels the first formant rarely is below 300 cps. It was found that in the nasal

[6] It should be noted at this point that while it is rather easy to recognize in a sonagram the presence or absence of formant structure, it is a matter of considerable difficulty to construct a device that would be capable of performing this judgment automatically; cf., J. L. Flanagan, *A Speech Analyzer for a Formant-Coding Compression System*, ScD thesis (M.I.T., May 1955).

[7] C. P. Smith, "Selective Compression of Speech Sounds," *JASA*, *29*, 832 (1953). Similar results were obtained by J.-P. A. Radley, see M.I.T. Research Laboratory of Electronics, *Quarterly Progress Report*, April, 1954, and by J. L. Flanagan – oral communication.

consonants, the low band was much stronger than in the vowels. Except for /i/,[8] the differences were fairly striking, as can be seen in the following table in which are given for vowels and nasal consonants the differences in db between the 300 cps high pass output and the 300 cps low pass output:

All vowels	from	−8 db to +15 db	(113 cases 3 speakers)
/i/	from	−8 db to +4 db	
other vowels	from	−2 db to +15 db	
Nasal consonants	from	−17.5 db to −3 db	(74 cases 3 speakers)

The overlap in the table above is due to the crude nature of the measurement which assigns an inordinately great importance to small downward variations of the first formant below 300 cps. It seems probable that the overlap could be overcome by a more refined measurement procedure in which the higher vowel formants would be given special amplification and where the overall level of the sound would also be taken into consideration.

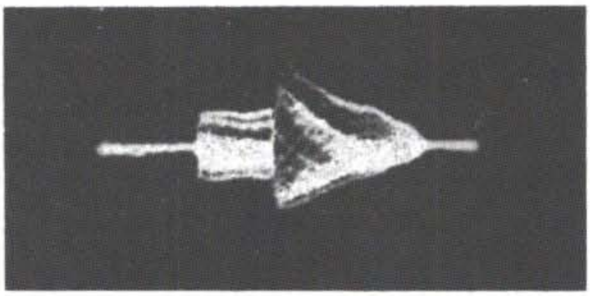


Fig. V-4. Oscillogram of the syllable /lu/. Subject K. The initial [l]-portion is markedly lower in intensity than the vowel.

(3) The vowels often have greater intensity than other phonemes. (See Figs. V-1 and V-4.) This fact is of importance in distinguishing, in the spoken chain, vowels from other phonemes having a formant structure. When in the syllable /lu/ the vowel was eliminated (by gating it out) and the remaining portion was presented to naive listeners, they judged the sound as /u/. When the listeners were presented with a stimulus containing both the low level (/l/) and the high level (vowel) portion of the syllable, they identified the syllable correctly as /lu/. This shows that an isolated segment cannot always be identified correctly, and that comparisons between adjacent segments must sometimes be made.

The phonetic entities represented by symbols enclosed in diagonals satisfy Condition (3a), sec. 1.3.

The higher level of the vowels is a consequence of the fact that in articulating a vowel the vocal tract is a more efficient transducer of kinetic energy into acoustical energy than in articulating any other class of phonemes. The most efficient configuration is the one most closely approximating a horn (cf. the shapes of megaphones used by cheerleaders); i.e., the articulatory position for [a] with its maximally wide mouth opening and its compressed pharynx. Due to the large mouth opening, the first formant will be at a high frequency, while the constriction in the pharynx will tend to lower the second formant.[9] These facts – the higher level and the concentration of the energy in a central region of the frequency spectrum – are the primary cues for the feature of compactness. They explain why compact vowels are *ceteris paribus* louder than other vowels.

Another consequence of the fact that vowels are more efficiently produced than other sounds is the easily observed phenomenon that in pronouncing vowels it is possible to vary the intensity over a wide range, while in pronouncing phonemes of other classes this range is much smaller. This is very clearly illustrated by the extreme cases of [a] and of the voiceless stops: tremendous variations can easily be produced in the former, while only a very slight increase in level can be obtained in the voiceless stops.

/l/ and /l,/ have spectra that closely resemble those of the vowels; i.e., possess formant structure. The spectra of /l/ and /l,/ differ from those of the vowels in that they have a higher attenuation in the upper formants (see Fig. V-5). As in the nasals, the first formant of /l/ is generally lower in frequency than that of the vowels [i] and [u]. The attenuation of the upper formants, however, is not quite as great as in the nasals and, therefore, when the band below 300 cps is compared with that above 300 cps, the differences are not as great as in the nasals.[10] /l/ and /l,/ have formants which are not continuations of the vowel formants. A discontinuity in the formants is thus an indication that a vowel or glide immediately adjoins a phoneme other than a vowel or a glide possessing formant structure. The decision as to which is the vowel and which is the other phoneme is then made on the basis of their relative intensities, the vowels always having the greater intensity.

The Russian /r/ and /r,/ are usually pronounced with a trill. These trills are about 40 to 70 msecs long, and not more than one is needed to give proper identification (see Fig. V-7). The sound is thus made up of two separate parts, one of which is primarily vowel-like and the other either a silence or a noise. In order to perceive a good trilled [r] it is necessary that both parts be present. There are, however, examples where no trill can be observed, and the phoneme contains much noise, (see Fig. V-6).

In the vocalic part of [r], the formants are less clearly defined than in the vowels, and the intensity of the frequency band between 600 cps and 1500 cps is much greater

[9] T. Chiba and M. Kajiyama, *The Vowel, Its Nature and Structure* (Tokyo, 1940), pp. 149–154, and K. N. Stevens and A. S. House, "Development of a Quantitative Description of Vowel Articulation," *JASA*, *27*, 484–493 (1955) and Fant, *op. cit.*

[10] Cf., data about liquids, below, sec. 7.2.

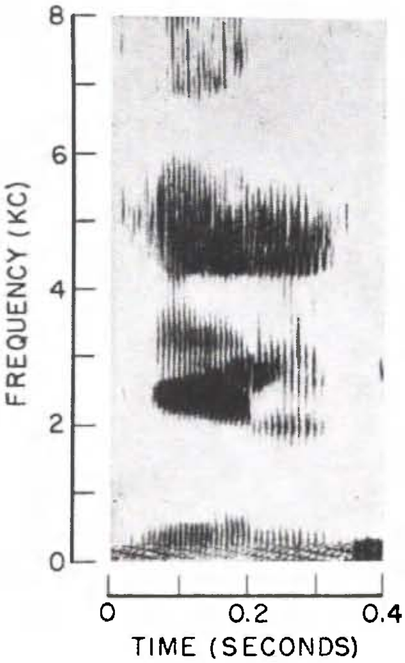


Fig. V-5. Sonagram of the syllable /il,/. Subject D. The upper formants are more attenuated in the liquid (at 0.2 secs) than the vowel.

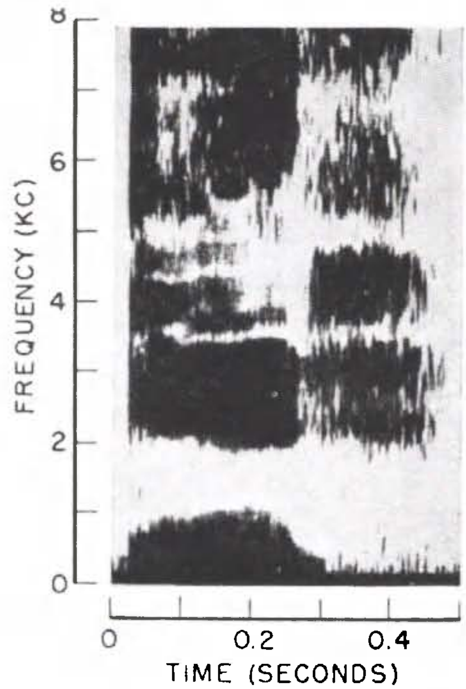


Fig. V-6. Sonagrams of the syllable /ir,/. Subject K. The /r,/ is not trilled here. It has a very strong noise component as shown by the irregularly striated high frequency region.

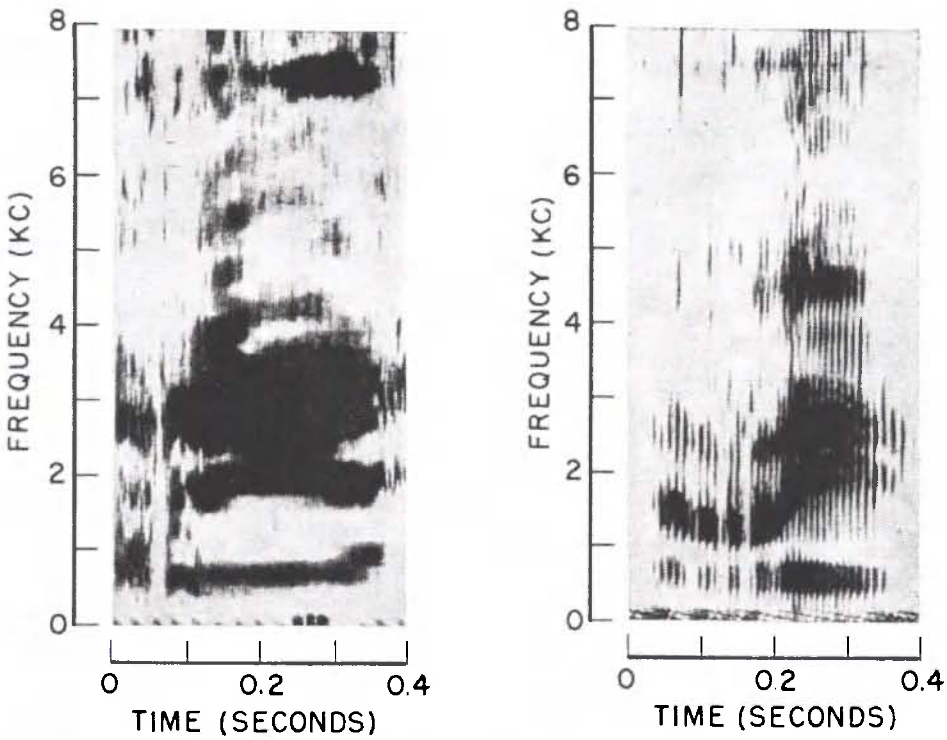


Fig. V-7. Sonagrams of the syllable /re/. Subject K (left) and D (right). Subject K has a single trill; Subject D has three distinct trills. Note also that /r/ as pronounced by both subjects has evenly spaced formants characteristic of the neutral vowel [ʌ].

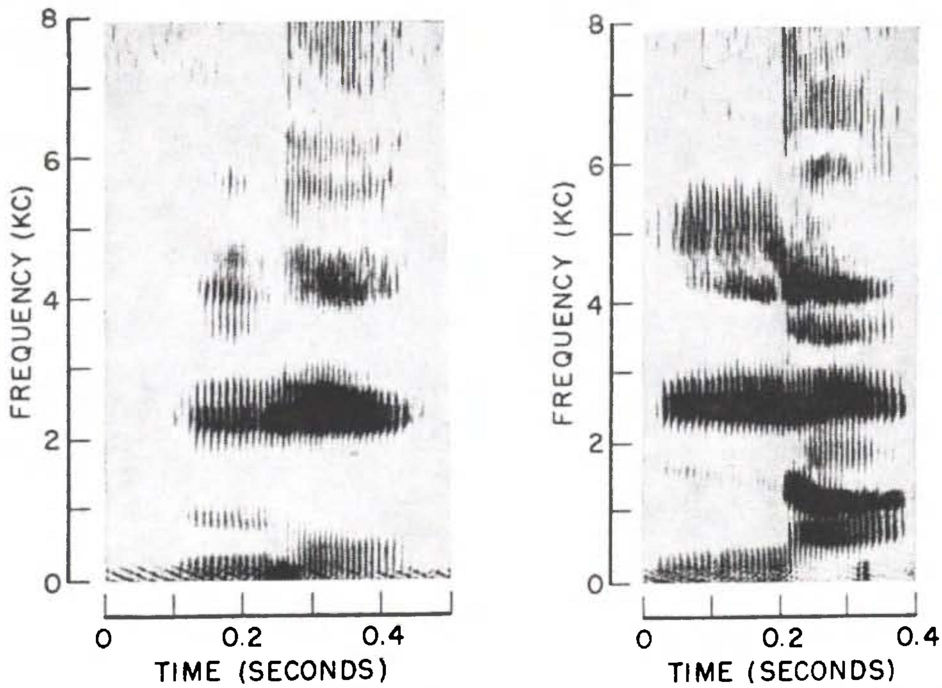


Fig. V-8. Sonagrams of syllables /mo/ (left) and /no/ (right). Subject D. The period of oral closure shows very strong formants. This, however, is an individual characteristic of Subject D. (Cf., Fig. V-9).

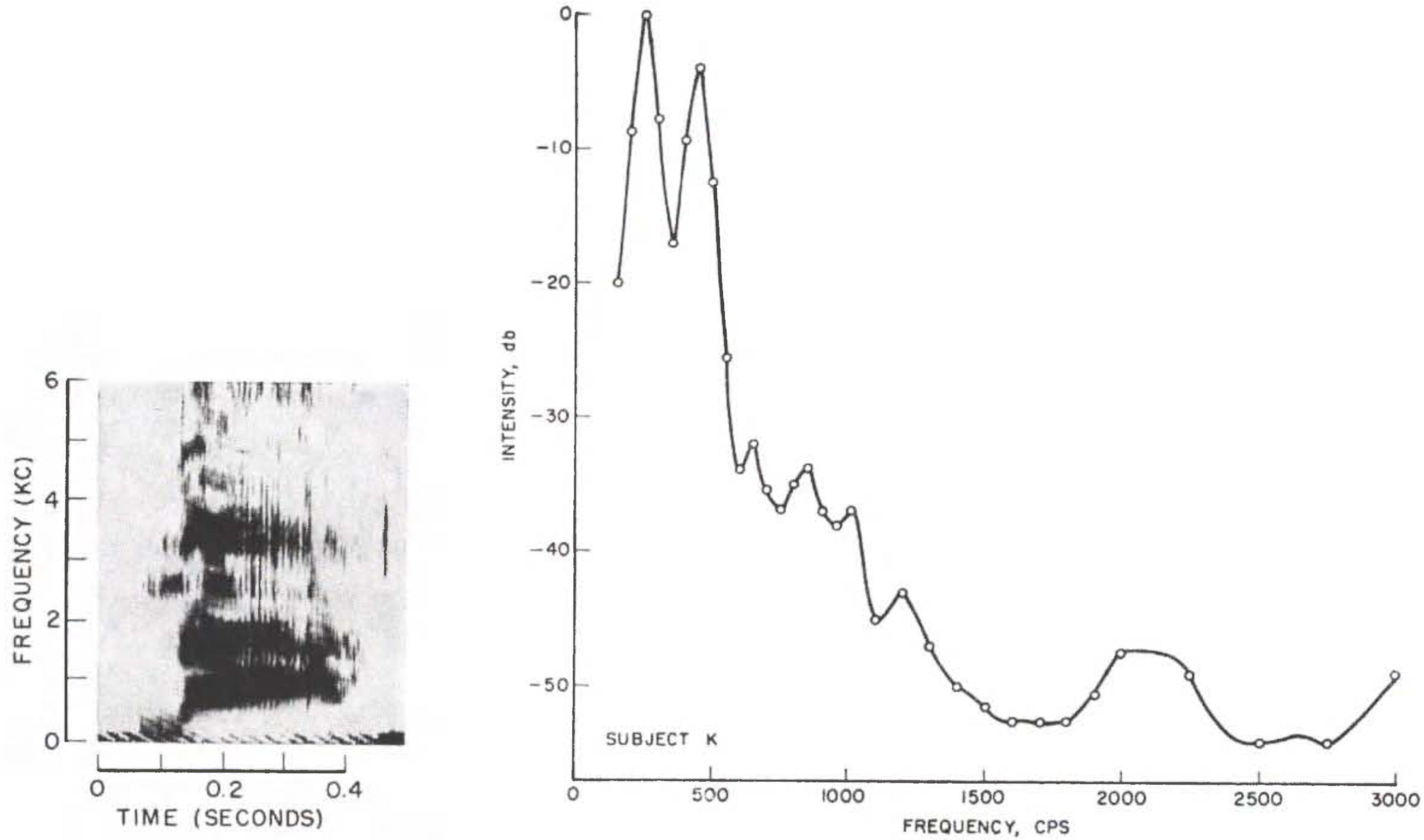
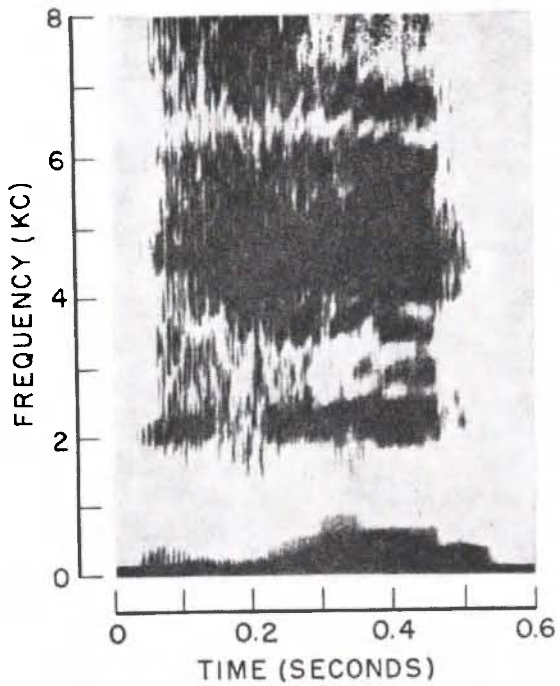


Fig. V-9. Sonagram of the syllable /na/ and energy density spectrum of the nasal portion. Subject K. The spectrum was taken at 0.2 secs from the beginning of the sound. Note the great attenuation of the upper formants in the nasal portion.



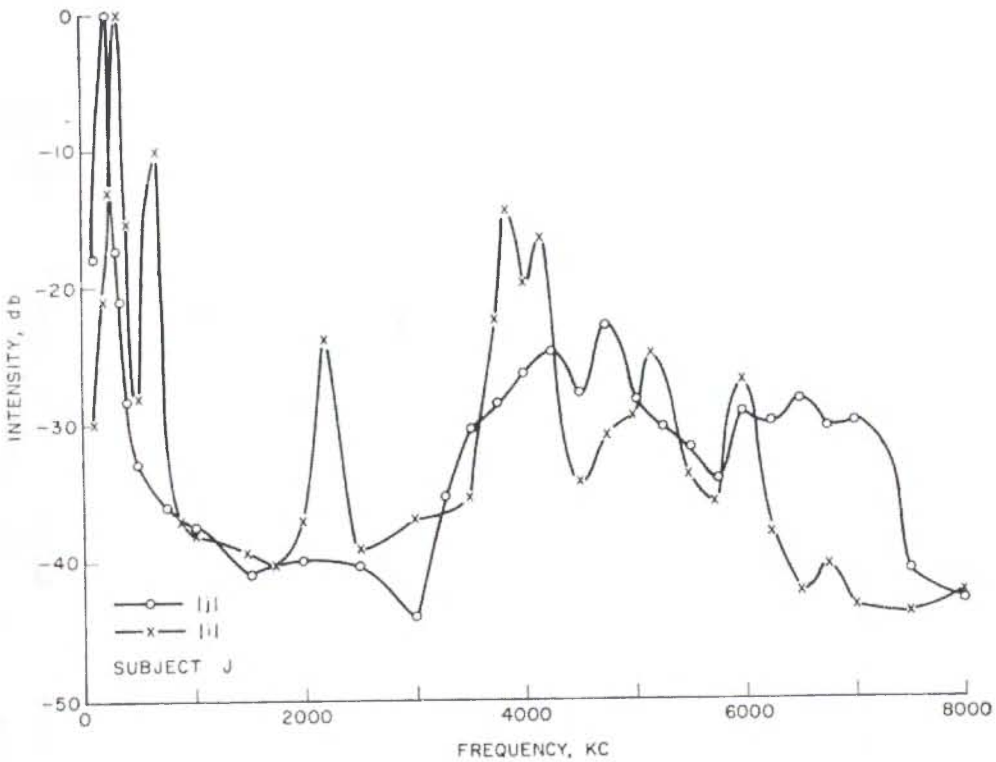


Fig. V-10. Sonagram and energy density spectra of the syllable /ji/. Subject J. The spectra were measured at 0.12 and 0 38 secs, respecttively, from the beginning of phonation. The glide /j/ has a lower first formant than the vowel /i/, and its upper formants are not clearly defined. Note also the sharply defined formants at 2000 cps and 4000 cps in the vowel.

relative to the first formant than in the nasals and in most vowels. In a few cases we observed a peak at 1500 cps; there were, however, other cases in which no energy maxima in this region could be found. The best way to describe the vocalic parts of the [r] is to say that they resemble most the neutral vowel [ʌ], strongly influenced by the adjacent phonemes (cf., Fig. V-7).

In view of the evolution of Common Slavic palatal [r] into strident /ř/ as in Czech and subsequently into /ž/ as in Polish, Slavicists will be interested to learn that we found several cases in which the segment corresponding to the closure (noise) when played in isolation gave a distinct [ž] impression. All these were instances of the sharped /r,/.

Physically the nasal consonants – and the stops as well – consist of two distinct events: one of these corresponds to the oral closure, and the second, to the transition to and/or from the closure. The latter is ordinarily identical with the transition of the cognate stop, the former, ordinarily referred to as the "nasal" portion, is characterized by a concentration of energy in the low frequencies, below 300 cps. This concentration is very marked and is the most important single property that all nasal consonants possess in common; cf., p. 119 above, and secs. 4.4 and 6 below. In the "nasal" portion the upper formants are present with varying degrees of intensity. In the cases investigated strong upper formants were found in a number of instances, particularly for speaker D; cf., Fig. V-8. For the other speakers the upper formants were usually attenuated by at least 30 db with respect to the first formant and consequently failed to show up in the sonagrams; cf., Fig. V-9.[11]

In sonagrams of connected discourse it can be observed that the nasals have formants that unlike those of the vowels are essentially stationary (i.e. during the period of oral closure there are no variations in the spectral composition of the sound), and are discontinuous with the formants of the adjacent vowels (Fig. V-1).

The glide /j/ resembles the vowel /i/. In a great many instances it was impossible to be sure that the subject had actually uttered /j/ and not the vowel. In the cases, however, in which there was no question that the subject had uttered a glide, the following characteristic properties were noted:

The first formant of the glide is somewhat lower in frequency than the first formant in the vowel /i/. This difference was very consistent: the first formant in the vowel is at approximately 275 to 300 cps; in the glide it is at approximately 200 cps. The second formant of the glide is considerably lower in intensity than that of the vowel; at times it is altogether absent. The second and higher formants are less clearly defined in the glide than in the vowel (see Fig. V-10). Comparisons of the energy below 300 cps with that above 300 cps indicate that the glide has a lower center of gravity than the vowel, although in the glide it is not as low as in the nasals.

[11] These observations agree with the experimental results obtained with artificially produced Visible Speech patterns at the Haskins Laboratories, as reported by P. Delattre at the 1953 meeting of the Modern Languages Association. See also A. Malécot, "Acoustic Cues for Nasal Consonants." *Language*, *32*, 274–284 (1956).

In connected discourse the formants of the glide form a regular continuation of the adjacent vowels. They differ in this from the liquids and nasal consonants, where an abrupt change in the formant position can usually be observed (see Fig. V-1).

4. *Compact vs. Noncompact, Diffuse vs. Nondiffuse, and Low vs. High Tonality*

The features of compactness, diffuseness, and tonality are based on properties that are most clearly observable in the energy density spectrum. For this reason these three features are treated together. However, because of special problems connected with each of the following classes of phonemes, the discussion will be divided into four parts: 4.1 the vowels; 4.2 the normaly strident or continuant consonants; 4.3 the normally mellow stops; and 4.4 the nasal consonants.

4.1 *The Vowels.*[12] The vowels of Russian are distinguished by means of the prosodic feature of accented vs. unaccented and the three inherent features of compact vs. noncompact, diffuse vs. nondiffuse, and flat vs. natural. The following discussion deals with the acoustical correlates of the inherent features.[13]

It has been shown by Fant that in vowels the frequencies of the formants completely determine the spectral envelope.[14] The relative intensity and the bandwidth of each of the formants can be derived mathematically, once the formant frequencies are known. In talking about the inherent features of vowels we shall, therefore, concentrate primarily on their formant frequencies and refer to other spectral properties only occasionally.

As a consequence of the phonological rules of Russian in words of Russian origin, unaccented {e} is not differentiated from unaccented {i}, and unaccented {o} is not differentiated from unaccented {a}.[15] In other words, for unaccented vowels the feature compact vs. noncompact has no acoustical correlate. It follows, therefore, that the identification procedure for the *unaccented* vowels need not include the feature compact vs. noncompact.

The feature of diffuseness is distinctive only for the vowels. It is acoustically signalled by a maximally low first formant. This can be produced by constricting the resonator more in front than further back; i.e., by giving the cavity a shape similar to that of a Helmholtz resonator.[16] In Russian this is produced either by lip rounding or by widening the pharynx. Lip rounding is used in the flat vowels; widening of the pharynx, in the natural vowels.

[12] For a discussion of the linguistic terminology employed here, which differs in certain details from that of *Preliminaries* and *Fundamentals*, see Chapter I, secs. 3.32 and 3.33. The data in this section (4.1) were gathered by L. G. Jones. For more details as well as a theoretical discussion by Jones, see the Excursus, pp. 157–167 below.

[13] The acoustical correlates of the Russian accent were not investigated in these studies. For a discussion of some linguistic aspects of the feature of accent, see Chapter II, sec. 3.2.

[14] Fant, *op. cit.*

[15] See Chapter II, sec. 2.

[16] R. Jakobson and M. Halle, *Fundamentals of Language* ('s-Gravenhage, 1956), p. 29.

Context	CV	VC	C,V	VC,	CVj	CVC	CVC,	C,VC	C,VC,	Total
					Syllables (273 per speaker)					
Speaker										
D	4	5	0	2						11
J	2	2	1	3						8
K	16	0	3	0						19
					Accented vowels in words (48 per speaker)					
D	2				1	0	1	0	1	5
J	1				0	0	0	2	0	3
K	1				3	1	1	0	2	8
					Unaccented vowels in strong position in words (27 per speaker)					
D								1	0	1
J								0	2	2
K								3	1	4
					Unaccented vowels in weak position in words (54 per speaker)					
D	0		0		0	2	2	3	1	8
J	2		0		1	4	2	1	0	10
K	1		1		1	2	3	4	3	15

	Totals	
Speaker	Errors	%
D	25	6.2
J	23	5.7
K	46	11.4
Total	94	7.8

Table V-1. Failures of the proposed criterion for the feature diffuse-nondiffuse. The table shows how many times and in what contexts there were violations of the rule that diffuse vowels have a first formant below 350 cps (450 cps for female speakers). See detailed data in Appendix I below.

Among the vowels the feature of compactness is actualized only in those that are accented and nondiffuse. Compactness is signalled by a maximally high first formant. It is produced by giving the resonator a horn-like (forward flanged) shape; i.e., by having the vocal tract assume a position in which the cross sectional area increases (or at least does not decrease) as we go from the glottis to the lips.

The tonality feature of flat vs. natural is produced by lip rounding, which in many contexts is combined with a constriction at the rear of the vocal tract. In position between two sharped ("soft") consonants, however, the tongue constriction is relatively far forward and the low tonality is produced by lip rounding alone, cf., e.g., /t,'ot,a/ = [t,öt,ə] "aunt" The acoustic consequence of lip rounding is a lowering of all formants. Of principal interest in the present instance are the effects on the second formant,

Context	CV	VC	C,V	VC,	CVC	CVC,	C,VC	C,VC,	Total
				Syllables (163 per speaker)					
Speaker									
D	9	8	3	2					22
J	3	4	3	2					12
K	7	2	1						10
				Accented vowels in words (36 per speaker)					
D	3	0			0	0	0	0	3
J	0	0			3	2	1	0	6
K	0	1			1	0	1	1	4

	Totals	
Speaker	Errors	%
D	25	12.5
J	18	9
K	14	7
Total	57	9.5

Table V-2. Failures of the proposed criterion for the feature compact-noncompact in vowels. The table shows how many times and in what contexts there were violations of the rule that compact vowels have a first formant above 650 cps (750 cps for female speakers). See detailed data in Appendix I below.

since the lowering of the first formant, which, as we have seen, can be achieved also by means other than lip rounding, serves as a cue for diffuseness and compactness of vowels. The influence of lip rounding on the third and higher formants is relatively minor.

In sum, compactness and diffuseness in Russian vowels are functions of the first formant, while tonality is a function of the second formant. The next problem is to give this statement numerical form.

It has been pointed out by Fant that "when formant positions are used as criteria for phoneme detection, it is necessary to take into account the significantly higher formants of the speech of women and children as compared with that of men."[17] The critical values for the female speaker K were 100 cps above the threshold values postulated for male speakers. On the basis of the data gathered by L. G. Jones, which are given in complete detail in Appendix I, the following critical values were established:

Diffuse vowels have a first formant below 350 cps (450 cps for the female speaker).

Compact vowels, which are all accented, have a first formant at 650 cps (750 cps for (female) speaker K) or above.

[17] C. G. M. Fant, *Transmission Properties of the Vocal Tract with Application to the Acoustic Specification of Phonemes* (= *M.I.T. Acoustics Laboratory Technical Report No. 12*) (Jan. 1952), p. 13.

Speaker D		Speaker J		Speaker K	
		Compactness in accented vowels			
	Fl		Fl		Fl
[t,išin'a]	600	[pəpad'at,]	600	['ɛtət]	750
[t,ir,im'a]	600	[t'opət]	700	[s'oncə]	750
[t,ut,un'a]	600	[s'oncə]	750	[t,'ot,im,i]	800
		[v'orət]	750	[pul,im,'ot]	900
		[pul,im,'ot]	750		
		[fsk'or,i]	750		
		Diffuseness in accented vowels			
[zv,'er,ə]]	250	[kət,ir,'inə]	350	[v,'er,im]	400
[pəlav'oj]	300	[təpar'ɨ]	350	[zv,'er,əm]	400
[m'il,u]	350	[pupav,'inə]	350	[pɨl,iv'oj]	400
[təpar'ɨ]	350			[pəl,iv'oj]	400
[pətam'u]	350			[pəlav'oj]	400
				[v'ɨrit]	500
				[təpar'ɨ]	600
				[v'ɨnut,]	600

Diffuseness in unaccented vowels

(If the word contained more than one unaccented vowel, the vowel in which the criterion failed is underlined. In cases where there was doubt about the pronunciation or about the correct interpretation of the sonagram, a question mark precedes the word.)

Speaker D		Speaker J		Speaker K	
?[t'opət]	250	?[pətam'u]	950	[t,išin'a]	450
?[t,'ot,im,i]	300	?[apas,'en,jə]	200	[stət,is,t,'ičiskəj]	500
?[təpar'ɨ]	350	?[kət,ir,'inə]	200	[v'ɨrit]	500
[v'ɨrit]	350	[tupav'at]	350	[v,'er,im]	500
[v'ɨrit,]	400	[təpar'i]	350	[t,išin'a]	500
[stət,is,t,'ičiskəj]	400	[kət,ir,'inə]	400	[t,ir,im'a]	500
[k'ol,im]	450	[v,'er,im]	400	[t,ir,im'a]	500
[v,'er,im]	450	[stət,is,t,ičiskəj]	400	?[pət,ir'at,]	500
[pəp,ir'at,]	500	[v'irit,]	400	[pəp,ir'at,]	500
		[v'iriti]	400	[v'irit]	600
		[v'ɨriti]	400	?[v'ɨnut,]	600
		[v'ɨrit]	400	?[m'il,u]	600
				?[v'ɨnut]	600
				?[m,'et,it,]	600
				?[v'ɨrit,]	750
				?[stət,is,t,'ičiskəj]	850

Table V-3. Summary of the failures of the criteria for the features compact-noncompact and diffuse-nondiffuse in words. For detailed data see Appendix I below.

These criteria were then tested on all the vowels in the sample, those in syllables as well as those in words. The results of these tests are summarized in Tables V-1 and V-2. The words in which "errors" occurred are given in Table V-3. As shown in Table V-1, the criterion for diffuseness holds in approximately 92% of the 1206 vowels tested. The criterion for compactness (see Table V-2) holds in better than 90% of the 597 vowels tested. The difference in the size of the sample for the two criteria is due to the fact that the feature compact vs. noncompact is neutralized in all vowels except /'a/, /'o/ and /'e/.

It is to be noted that the efficacy of the proposed criterion is particularly low in the unaccented vowels in weak position, i.e., in unaccented vowels neither in word initial nor in pretonic position. Phonetically at most three vowels are distinguished in this position; there is, however, vacillation with regard to the distinction between [i] and [ə].[18]

It is significant that unaccented [a] had in the majority of instances (65%) first formants that were below the critical value 650 cps (750 cps) established for the compact accented /'a/. This fact lends support to the identification scheme proposed here, where the feature compact vs. noncompact is neutralized in (and hence not relevant for the identification of) the unaccented words.

Unlike the features diffuse vs. nondiffuse and compact vs. noncompact, the third inherent vowel feature flat vs. natural poses no problem. Its acoustical correlate is the position of the second formant. On the basis of the data the following rule was formulated:

Flat vowels, both accented and unaccented, have a second formant at 1300 cps (1400 cps for female speaker K) or below.

In the 873 vowels of the sample, there were only 6 exceptions to this rule.

4.2 *Strident or Continuant Consonants.* These consonants are produced by blowing air past an obstacle capable of generating turbulence. They differ from each other in their initial phase: all continuants have a gradual onset; /c/ and, in most contexts, /č/, have an abrupt onset. /x/, /š/, /ž/ and /č/ are compact; of these /x/ is grave, and the rest are acute. /f/, /f,/, /v/, /v,/, /s/, /s,/, /z/, /z,/, /c/ are noncompact; of these, the first four are grave, the last five acute. The spectra of the interrupted strident consonants do not differ from those of their continuant cognates; cf., Appendix II. The following discussion, therefore, applies to the continuants as well as to /c/ and /č/.

In the syllables which served as our samples, strident consonants lasted for well over 50 msecs; it was, therefore, possible to use a 50 msec gate everywhere. It is

[18] The neutralization of all vowel distinctions except for flat vs. plain has often been noted in the literature. See, e.g., the following remark of V. A. Bogorodickij: "In the words *nadlomit'*, *sapogi* the unaccented *a* changes into a sound that is shorter and vacillates between an unclear *a* and an unclear *i*." V. A. Bogorodickij, *Očerki po jazykovedeniju i russkomu jazyku* (Moscow, 1939), p. 115. The neutralization is a standard feature of much Russian verse after Puškin Cf., e.g., B. Tomaševskij, "K istorii russkoj rifmy," *Trudy otdela novoj russkoj literatury, 1*, 264ff. (1948).

known that high frequencies are involved in the perception of these sounds. Hence frequencies between 500 and 10,000 cps were studied. The pass band of the Hewlett Packard Wave Analyzer was adjusted to be substantially flat (—3 db) over 240 cps (see Fig. V-2, R=24K). Between 500 and 10,000 cps points at 500 cps intervals were measured. A measurement was also taken at any maximum and minimum that did not coincide with the standard measurement points. Thus, about 20 points were obtained for each sample investigated. Individual spectra are given in Appendix II.

The production of these consonants has recently been investigated from an acoustical point of view by Fant and by K. N. Stevens.[19] The following discussion is based primarily on results obtained by these investigators.

It is convenient to consider the continuants as the product of several factors: the exciting source, the transfer characteristics of the vocal tract, and the location of the source. In the continuants the exciting source is a turbulent noise generated by air flow past some obstacle. The spectrum of the turbulent noise so generated has a single spectral maximum. It appears that in continuants two sources of this type are present: one, having a high frequency maximum, is produced by the air blown past the teeth; and the other, having a low frequency maximum, is produced by air flowing past some blunter obstacle, such as the lips or tongue. The contributions of the two types of source vary, however, in importance. It has been shown by Fant that from the point of view of the one-dimensional model used generally in all discussions concerning the acoustical properties of speech, the turbulent source must be considered as being of low impedance. Accordingly cavities behind the source cannot, in general, be disregarded, as they can be in the case of the high impedance glottis source.

The geometry of the vocal tract in articulating the particular phoneme is of obvious importance, since it determines the resonance properties of the cavity. All continuants are produced with a narrow constriction in the tract. This constriction may be very short, as, e.g., in /f/; or it may be quite long, as, e.g., in /s,/. Behind this constriction the vocal tract broadens out considerably. It is probable that the effect of these large back cavities is minor and can be disregarded, particularly in a qualitative exposition like the present one. If the constriction is very narrow, the cavities behind it may be considered uncoupled from those in front of it, and sometimes be neglected.

Finally the location of the source must be considered. Since the source is assumed to have low impedance, the spectrum of the output signal will contain zeros (anti-resonances) when the source is located in the middle of the tract rather than at its end. The reason for this can be readily appreciated if one considers the vocal cavity with a source in the middle as the equivalent of a configuration in which a source is located between two resonators. It is evident that at the frequencies at which one of the resonators resonates,* the other resonator will receive little or no

[19] Personal communications from Fant and Stevens.

* This applies only to the resonance at which the impedance of the first resonator is a maximum.

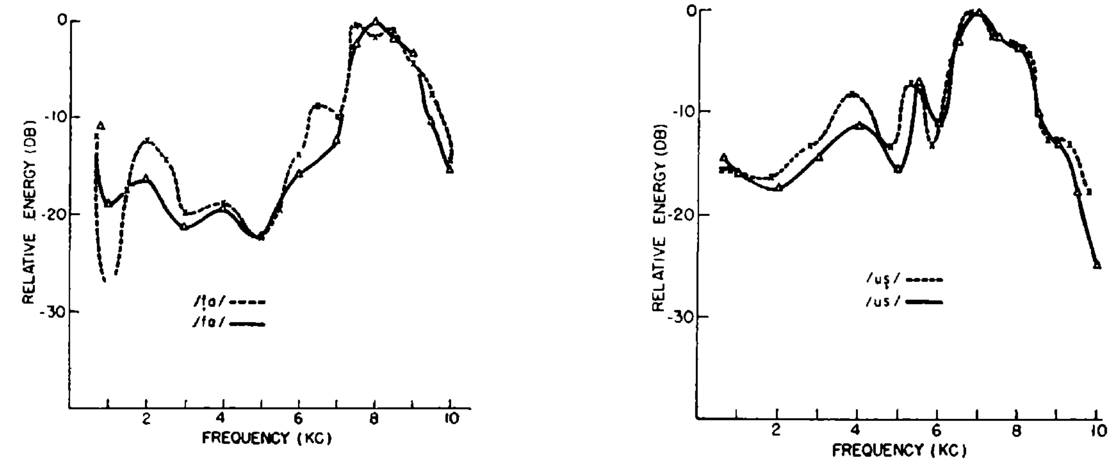


Fig. V-11. Energy density spectra of the continuants in the syllables /fa/, /f,a/ and /us/, /us,/. Subject J. The spectra of the sharped and plain continuants are practically indistinguishable.

energy from the source, since all of it will go to the first resonator. Consequently if a microphone were to be placed in front of the second resonator, little or no sound would be picked up at these frequencies.

These general considerations account fairly well for the measured data. Consider first the grave continuants; i.e., those produced with a constriction in a peripheral region of the oral cavity. In the case of {*f}, we have a very narrow constriction at the teeth, and it is almost certain that the effects of the cavities behind this constriction can be neglected (cf., Fig. V-11a). The spectrum of a labiodental is determined primarily by the effects of the two types of source. Low frequency energy in the labiodentals is produced by blowing air over the lips. Air flowing past the teeth generates turbulence with a high frequency maximum, which is greatly reinforced by the resonance of the short lip cavity. This maximum can be observed clearly in the spectra of speaker J. It is not always present in the spectra in Appendix II because it is frequently above 10 kc, where measurements were not made. The level of the lower frequency region relative to that of the upper frequencies depends on the force with which air flows past the different types of obstacle and varies a great deal from utterance to utterance.

/x/ is produced with a constriction at the rear of the velum. If this constriction is narrow enough, as it apparently is for speaker J, the cavities behind the constriction can be neglected, and the vocal tract will then react as a tube closed at one end about 8 or 9 cm long. If the constriction is not narrow enough, the sound produced will have a spectrum quite similar to that of a vowel, except that some of its formants may be suppressed by the action of the zeros due to the location of the exciting source in the middle, rather than at the end, of the tract. The turbulent source of primary interest in /x/ is the one having a low frequency maximum.

To account for the spectra of acute continuants, it is necessary to consider the effects of the zeros as well as of the resonances. The geometry of the vocal tract determines uniquely the resonances of the signal; the location of the source can affect only the position of the zero. If it is assumed that in both /s/ and /š/ the source is located at the teeth, a zero is to be expected at frequencies where the cavity behind the teeth is at resonance. This cavity extends from the teeth to the rear of the tongue constriction; i.e., to the place where the vocal tract begins to broaden out. For /s/ this cavity can be considered a uniform tube about 2.5 cm long, which would give resonances at odd multiples of a quarter wave length. Accordingly the lowest zero, and the only one of interest here, is to be expected at 3.5 kc. In /š/ the back cavity is longer – perhaps 4 cm – but since it is considerably broader in front than further back, it will resonate at frequencies somewhat lower than a uniform tube of equal length. This is consistent with the experimentally observed minima in /š/ spectra, which are found at 1.5 kc and at 4.5 kc.

The resonances are determined by the vocal tract configurations. In /s/ we can consider the cavity in front of the teeth as practically uncoupled from the cavity behind. Resonances are, therefore, to be expected at the quarter wave length of the front cavity and at the half wave length of the back cavity. If the former is assumed to be 1 cm long, and the latter, as before, 2.5 cm, resonances should appear at 8 kc and 7 kc respectively. Since these two resonances are so close in frequency, they will appear as a single maximum of high intensity. The prominence of this maximum is further accentuated by the low level of the frequencies around 3.5 kc due to the back cavity zero. The source exciting these resonances is, of course, the high frequency teeth source. A certain amount of energy in the low frequencies may at times be observed. (Cf., Appendix II, Fig. 9, speaker J /os/ and Appendix II, Fig. 11, speaker J /as/.) It is due to the low frequency turbulence produced by the air flowing past the lips.

For /š/ it is necessary to consider the entire cavity from the lips to the *front* of the tongue constriction as a single unit. This cavity would have an effective length of about 3.5 cm, which would yield a resonance at 2.5 kc. A second resonance in /š/, which is generally observed at about 6 kc, can be accounted for as the second mode of the same cavity.

4.21 To separate the compact strident phonemes from the noncompact the following set of criteria was developed.

(1) Locate the peak of the spectrum. If the peak is at 6000 cps or higher and also exceeds all lower maxima by at least 4 db, the phoneme in question is noncompact. If the peak is below 6000 dbs, proceed to step 2.

(2) (a) If the peak is between 2000 and 4000 cps, measure the average power level in a 1000 cps band centered at the peak.

(b) If the peak is below 2000 cps, measure the average power between 2000 and 3000 cps.

(c) If the peak is above 4000 cps, measure the average power level between 3000 and 4000 cps.

(3) Measure the average power level between 500 and 1500 cps. If the sound is voiced, subtract 10 db in order to eliminate the contribution of the voicing component.

(4) Subtract measurement 3 from measurement 2. If the difference is small (between -2 db and $+13$ db in our measurements), the phoneme is noncompact. If the difference is larger, the phoneme is compact.

This procedure was tested on 224 syllables as spoken by our subjects K and J, and "correct" judgments were obtained in 209 cases; i.e., in about 93% of the measured phonemes.

This measurement procedure is a specification of compactness given in *Preliminaries*, where it was described as a concentration of energy in the central part of the spectrum. Step (1) eliminates from consideration as compact all sounds which have concentrations of energy in the upper frequencies. Steps (2) and (3) define the region between 500 and 4000 cps as a central region. Step (4) compares energy in two 1000 cps bands and judges as compact that sound in which one of these greatly exceeds the other: it defines compactness.

4.22 The distinction between grave and acute continuants and affricates is also based on spectral properties. The procedures employed to establish this distinction were therefore of the same general nature as those of sec. 4.21. In order to eliminate tape noise, hum, etc., the samples were passed through a 200 cps high pass filter having an attenuation of 24 db per octave. The energy below 1000 cps and the total energy in the sound was then measured. In the case of voiced phonemes 15 db were subtracted from the first value. Next the 1000 cps low pass value was subtracted from the total energy, and it was found that in the case of noncompact continuants and affricates, the difference was less than 20 db, if these were grave (labial), and more than 20 db, if they were acute (dental).

This procedure was tested on 56 samples of noncompact continuants and of /c/ spoken by subjects K and J, and "correct" identifications obtained in 53 cases. All "errors" were samples of /z/ spoken by subject K.

Compact continuants are grave (velar) if the level of the 1000 cps band measured in step 2 above is lower than that of the band between 500 and 1500 cps. If the former band exceeds the 500 to 1500 cps band, the continuant in question is acute (palatal).

This procedure was tested on 86 samples of compact continuants and of /č/, spoken by subjects J and K, and "correct" identifications were obtained in 83 cases.

The role of the formant transitions in the portion of the vowel immediately adjacent to the consonants is discussed in section 9.1 below. In the case of the normally strident phonemes, the transition contributes very little towards the identification of the features of compactness and gravity; the relevant information resides primarily in the consonantal noise. This conclusion is in agreement with the results obtained in perceptual tests.[20]

4.3 *Mellow Stops.* Mellow stops are produced by a complex of articulatory

[20] Cf., Katherine S. Harris, "Cues for the Identification of the Fricatives of American English" (abstract of paper presented to the June 1954 meeting of the Acoustical Society of America), *JASA*, *26*, 932 (1954), where it is shown that except for the distinction between /f/ and /θ/ the vowel transitions play no important role in the perception of continuants.

movements among which it is customary to distinguish three phases: a rapid closing movement, a period of closure of the oral cavity, and an abrupt release. Of these three phases only the period of closure is a necessary cue for stop perception. Since the closure is common to all stops it does not serve to distinguish different classes of stops. These are distinguished by their individual closure and/or release-movements. It is the acoustical correlates of these closure- and release-movements that produce the distinctive cues that are of interest here.

When the stop is adjacent to vowels the closure- and release-movements ordinarily produce rapid transitions in the vowel formants. The release is normally accompanied by a short burst, the so-called *stop explosion.* When the explosion is absent – i.e., when we are dealing with an *imploded* stop – the formant transitions provide the only cue for the identification of the stop. On the other hand, when the stop is adjacent to phonemes other than vowels, the corresponding formant transitions are missing, and the relevant information is then conveyed by the stop explosion.

The problem of the relative importance of these two major cues, the burst and the transition, in the identification of the different stops has been the subject of considerable debate. The formant transitions are determined by the geometrical configurations which the vocal tract successively assumes as it passes from the vowel to the closure and/or from the closure to the vowel. In Russian, therefore, the transitions are dependent on two factors, the point of articulation of the stop and the feature of sharping (palatalization). In the body of data that was investigated the formant transitions were most closely correlated with the feature of sharping.[21] They did not show as systematic a correlation with the point of articulation; i.e., with the features compact-noncompact and grave-acute. This may in part be due to the fact that the corpus contained only exploded stops in which the stop burst alone is sufficient to provide the necessary information. In contrast to the transitions, the stop explosions which were examined were clearly correlated with the features of compactness and gravity. The following discussion focusses on the spectral properties of the stop bursts.[22]

To obtain spectra of the stop bursts it was necessary to isolate these from the rest of the syllable. A comparison of a sonagram with a full-wave-rectified display on a cathode ray oscilloscope made possible a fairly accurate placing of the gate. Since the stop burst is often very short – less than 20 msecs – the end of the gate was set at 20 msecs from the beginning of the burst or at the beginning of the periodic oscillation that marked the onset of the vowel, whichever was shorter. The pass band of the wave analyzer was adjusted to be substantially flat (-3 db) for 120 cps (see Fig. V-2, R=50 K). Spectra were then produced by measuring points at intervals of 250 cps between 250 cps and 5000 cps in the manner described in sec. 2.2 above. To reduce

[21] See below, sec. 9.1.

[22] As far as is known spectra of stop bursts had never been systematically measured prior to the present study. Since the conclusion of the work reported here, we have also studied the stop bursts of English; see Halle, Hughes and Radley, *op. cit.*

the effect of tape noise and hum, all measurements were taken with a fixed high pass filter inserted immediately after the gate circuit. Its critical frequency was 200 cps, below which there was an attenuation of about 24 db per octave. Sample spectra of stops produced by the above method are given in Appendix II.

Of the normally mellow stops of Russian: /k/, /k,/ and /g/ are compact; /p/, /p,/, /b/, /b,/, /t/, /t,/, /d/, /d,/ are noncompact. The opposition grave vs. acute is distinctive only for the latter; the first four of the phonemes listed are grave, the last four acute.

4.31 The procedure for separating compact from noncompact stops differs in some details from that used to separate compact from noncompact continuants (see sec. 4.2 above). Both procedures, however, have so much in common that with further work it should prove possible to combine them into a single operation. The procedure for stops consisted of the following steps:

(1) Locate the peak of the spectrum,
  (a) If the peak is above 3500 cps, the phoneme is noncompact,
  (b) If the peak is below 3500 cps, proceed to step 2.
(2) If the peak is between 1000 and 3500 cps,
  (a) Measure average level in a 1000 cps band centered at the peak,
  (b) Measure the average level of the entire spectrum,
  (c) Subtract measurements 2a from 2b.

Large differences indicate compact stops; small differences indicate noncompact stops. (In the cases investigated, if the peak exceeded the average by 8 db or more, the stop was called compact.)

(3) If the peak is below 1000 cps, measure the area under the curve. A large area indicates that the phoneme is noncompact; a small area indicates that the phoneme is compact. It is to be noted that this criterion is substantially similar to criterion (2); it was, however, impossible to reduce these two criteria to one.

These three criteria were tested on 114 stops in syllables spoken by subjects K and J. "Correct" identification were obtained in 104 cases; i.e., in 92% of the cases. There were no "errors" due to criterion (1) (applied to 9 cases); there were six "errors" due to criterion (2) (applied to 28 cases); and there were two "errors" due to criterion (3) (applied to 75 cases). The relatively large number of "errors" due to criterion (2) can in part be explained by the fact that three of the sounds were poorly recorded or produced.

4.32 As in the cases discussed in sec. 4.22 of this chapter, it was impossible to identify the feature of gravity in consonants by inspection of the spectra (cf. Fig., V-1 and Appendix). It does not seem likely that this failure is due to the fact that we limited our measurements to frequencies below 5000 cps, since it is known that the identifiability of stops is not affected as the pass band is limited to 0–5000 cps.[23]

[23] G. A. Miller and P. E. Nicely report in "An Analysis of Perceptual Confusions among Some English Consonants," *JASA*, *27*, 338–352 (1955), that as the cut-off frequency was lowered from 6500 cps to 5000 cps there was no significant decrease in the percentage of correct identifications of stops. Cf., their Tables VI and XII.

We had observed that [t] could be transformed into [p] if frequencies above 1000 cps were effectively eliminated; i.e., attenuated by 36 db per octave. This led us to compare the total energy in the stop burst with the energy in a high frequency band. Specifically we compared the total energy in the sound with that above 3000 cps. If the total energy exceeded the latter by 8 db or more, the phoneme was classed as grave; if the difference was less, the phoneme was classed as acute.

Some difficulty was experienced with the voiced stops of female subject K. Because the fundamental pitch of her voice is naturally higher, the 200 cps filter did not effectively eliminate the contribution of the voicing component. For this reason 10 db was subtracted from the total energy measurement, thereby bringing the values for voiced stops of subject K in line with the rest.

This procedure was tested on 116 sounds spoken by subjects K and J, and "correct" identifications obtained in 112 cases.

4.4 *The Nasal Consonants.* Russian possesses the following nasal consonants: /m/, /m,/, /n/ and /n,/. All nasal consonants are noncompact, hence the distinctive feature of primary interest here is grave vs. acute. /m/ and /m,/ are grave and /n/ and /n,/ are acute.

A syllable consisting of a nasal consonant and a vowel – e.g., /ma/ – has two clearly separated parts: one, the nasal part which corresponds to the time during which the mouth is closed and the sound radiates primarily from the nares; and the second, the vocalic part, where the mouth is open and the nasal passage, usually closed (cf., Figs. V-8, V-9 and V-12). Its structure is thus quite similar to a syllable consisting of a mellow stop followed by a vowel, the only difference being that in the latter syllable the period of oral closure is occupied either by a silence (in the voiceless case) or by the "vocal cord buzz" (in the voiced case). We should thus expect certain similarities in the behavior of mellow stops and nasals.

If in the syllable /ma/ we gate out the nasal part and the initial part (transition) of the vowel, we perceive a vowel. As we lengthen the gated portion by moving our gate toward the beginning of the syllable, we pass through a stage where there is a fairly distinct /ba/ impression, and then finally, as more and more of the nasal portion of the syllable is included, we obtain a clear /ma/ impression. The cue for nasality is thus contained primarily in the period of oral closure, while the cue for the other features that are relevant in the identification of a nasal consonant – in particular grave-acute – may reside not only in the period of oral closure, but also in the vowel transtition. In cases where there is no adjacent vowel – e.g., /mar,'izm/ "Marrism" as opposed to /ukar,'izn/ "reproaches" (gen. pl.) – the differentiating cue can, of course, lie only in the nasal portion.

A characteristic of spectra of the nasal portion is the presence of an anti-resonance, or "zero", due to the fact that during the period of oral closure the vocal cavity is bifurcated above the oral pharynx into the closed oral cavity, which forms a sort of side branch, and the nasal cavity, which is open to the outer air. In this connection, House has recently remarked:

The frequency position of the anti-resonance depends upon the length of the side branch, being lower in frequency for the longer side branches. Since the articulation of /m/ is bilabial the side branch formed by the closed oral cavity is long compared to the cavity formed by post-dental or velar closures. The anti-resonance that is characteristic of the bilabial nasal consonant is in the vicinity of 1000 cps, while the post-dental articulation creates an anti-resonance near 3500.[24]

The acoustic effects during and after the release of the oral closure must be the same as those in the case of the mellow stop consonants. The effects of the nasal passage are much attenuated by the opening of the oral cavity, since the dissipation in the nasal cavity is much greater than that of the straight passage through the mouth and lips.

We had records of nasal consonants for four speakers. With the available measuring techniques it was unfortunately not possible to arrive at definitive conclusions about the feature of gravity in nasal consonants.

The most easily interpreted were the records obtained from speaker D. By and large the samples from this speaker confirmed the theoretical description given above. During the period of oral closure there were two clearly defined formants, one at approximately 230 cps and the other between 2000 and 2500 cps (see Figs. V-8 and V-12). All plain /m/ in addition had a formant at about 800 cps. This formant was absent in sharped /m,/ as well as in the acute nasals.[25]

The transitions of the second and third formant also showed fairly consistent differences. In general, the transition in /n/ and /n,/ terminated at higher frequencies than those in /m/ and /m,/. This was realized in a number of ways: the acute nasal had an upward shift in the second and third formants while the grave nasal had no noticeable shift; or the acute nasal lacked any noticeable shift while the grave nasal had a downward shift; or the acute nasal had an upward shift and the grave nasal, a downward shift. An interesting difference was observed in the grave vowels adjacent to sharped nasals. Here next to the grave /m,/ there is a short interval during which there is no change in the formant position, which is followed by a downward shift. In the grave vowels adjacent to /n,/ the transition is immediate. (See Fig. V-12.)[26]

The records for the female speaker, K, were more difficult to interpret. In these, the main difference between acute and grave nasals seemed to lie in the fact that the latter had only one very strong low formant, while the former had a second formant at about 2500 cps. In a few cases of /m/, a formant at about 1000 cps could be observed. The transitions were, in general, similar to those of speaker D, strikingly so in the case of grave vowels adjacent to sharped nasals.

[24] A. S. House, "Analog Studies of Nasal Consonants," *Journal of Speech and Hearing Disorders, 22*, 199 (1957). See also S. Hattori, K. Yamamoto, O. Fujimura, "Nasalization of Vowels in Relation to Nasals," *JASA, 30*, 267–274 (1958).

[25] According to Fant this extra formant is due to a peculiarity of D's pronunciation of /m/ which is articulated with a special constriction in the back of the oral cavity. See the radiographs of /m/ in C. G. M. Fant's forthcoming *Acoustic Theory of Speech Production* ('s-Gravenhage, in press).

[26] See sec. 9, below.

The data from the remaining two speakers (J and I) provided little information, primarily because certain peculiarities of their voices made it difficult to obtain good sonagrams from records made by these speakers, and hence it was impossible to study the formant transitions. For these speakers, moreover, differences among the spectra of the nasal portion of nasal consonants were so unsystematic that no general characterization can be given.

It thus appears that in the case of the feature of gravity in the nasal consonants, we are dealing with a multiplicity of cues, none of which can clearly be said to be necessary. The evidence seems to suggest that the grave nasals give greater prominence to lower frequency regions than do the acute nasals, *ceteris paribus*.

5. *Strident vs. Mellow*

The perceptual counterpart of stridency, a high degree of noisiness, is produced by a random distribution of acoustical energy. Limitations on stridency can be achieved by imposing an order either in the frequency domain (formants) or in the time domain (short bursts of sound). For Russian both manners are relevant: on the one hand, the nasals are distinguished from strident phonemes by possessing a clear formant structure; on the other hand, the mellow stops (imploded as well as exploded) are opposed to strident phonemes as sounds limited in time, to noises of indefinite duration. This organization is apparent in the sonagram where the nasals are represented by horizontal bars (formants) (cf., Figs. V-8 and V-12), and the mellow stops, by narrow vertical lines – often only a single line (cf., FIgs. V-13a, b; V-14a, c; V-15; V-16a). Such an organization is absent in the strident phonemes. (Cf., Figs. V-13c, d and V-16b.)

There has already been occasion to remark on the difficulty of implementing a measurement procedure that would identify "formant structure",[27] which is one of the two major cues for the absence of stridency. The situation with regard to a measurement scheme for the second oue for mellowness – i.e. the "short burst" character of the mellow stop – is hardly in a more satisfactory state. An appropriate measure for this property might involve an estimate of the phase relations of their frequency components. This idea is suggested by the following remark of Licklider:

> The various frequency components of the white noise are assigned their phase angles at random; the frequency components of the single pulse all reach their maximum amplitude at the time t = O, and they cancel one another at all other times. As a result, we hear the white noise as *sshhh* and the single pulse as *pt*.[28]

It is to be hoped that the question of the significance of phase in the perception of speech and of sound, in general, will be subjected to a thorough examination. Enough

[27] See above, sec. 3.

[28] J. C. R. Licklider, "Basic Correlates of the Auditory Stimulus," in S. S. Stevens, ed., *Handbook of Experimental Psychology* (New York, 1951), p. 1026.

evidence has been accumulated to indicate that the traditional view of phase being irrelevant in the perception of sound can no longer be maintained.[29] It appears probable that psycho-acoustics as a whole will gain a great many new insights if investigations in this area were to be carried out.

A random distribution of acoustical energy is produced by a turbulent air stream. Turbulence is ordinarily generated in the human vocal tract by forcing air through a narrow constriction. The efficiency of conversion of the energy contained in the air flow into sound is dependent upon the radius of curvature of the obstacle generating the turbulence as well as on the angle at which the turbulent airstream strikes the obstacle. The smaller the radius of curvature – i.e., the sharper the obstacle – and the closer the angle of incidence of the air stream is to 90°, the more efficient the conversion into sound, the more noise produced for the same effort. These considerations explain why /s/ and /š/ are more strident than /x/ and /f/, for the former have an additional obstacle against which the air stream strikes at almost 90°.

The differences between strident and nonstrident interrupted consonants reside in the temporal properties of the excitation. In the nonstrident interrupted consonants (stops) the excitation consists of a short burst of noise, while in the strident interrupted consonants (affricates), the excitation consists of a noise which, except for its rapid onset, differs but little from the strident continuants. It is easy to convert a tape recorded strident continuant into its cognate affricate simply by eliminating – either by electrically erasing or by splicing out – the initial part of the strident continuant. When the duration of the noise is further reduced, the cognate mellow stop is often perceived. The results of this experiment can be explained in terms of phase relations among different components as was done in the beginning of this section.

6. *Nasal vs. Nonnasal*

The four nasal consonants of Russian are the only consonants having formant structure. As has been remarked above the cue for nasality lies in the period of oral closure. The nasal portion differs from the vowels and the glide in that its formants have an abrupt, instead of a gradual transition to the formants of adjacent vowels. In common with liquids and glides, the nasal portion has a lower overall level than the adjacent vowels. It differs from the liquids (and vowels, too) in that it has more of its energy concentrated in the region below 300 cps.

In a pilot experiment we measured the amount of energy below and above 300 cps in nasals and liquids, and we found that in liquids these varied between -10 db and $+1$ db, while in nasals these varied between -17.5 db and -8.5 db. The overlap would probably disappear if proper weighting procedures were to be applied.

Additional information about nasals is given in secs. 3 and 4.4 above. Sonagrams of nasal consonants are given in Figs. V-8, V-9 and V-12.

[29] *Ibid.*, pp. 1024 ff.

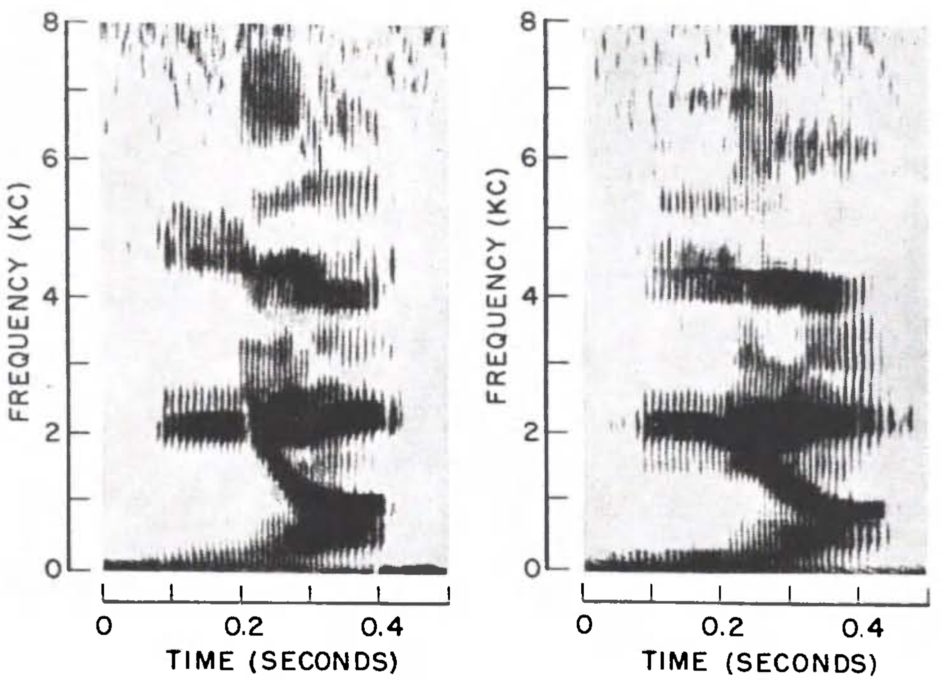


Fig. V-12. Sonagrams of the syllables /n,o/ and /m,o/. Subject D. The transition next to /n,/ is immediate, while next to /m,/ there is a short stationary interval. Compare the formant transitions here with those next to plain nasals in Fig. V-8.

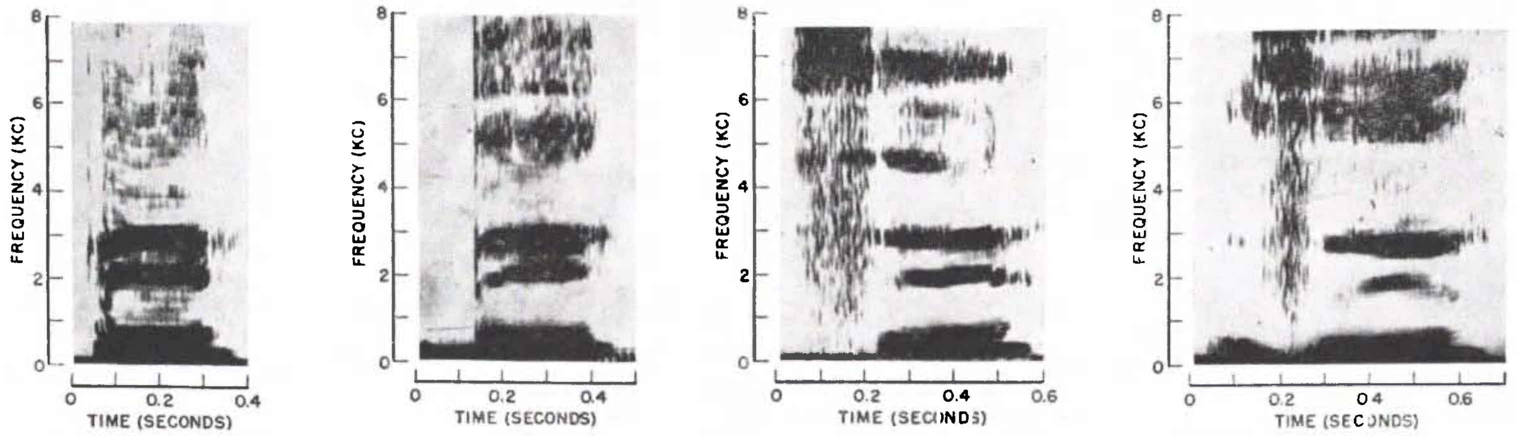


Fig. V-13. Sonagrams of the syllables /pe/, /be/, /se/, and /ze/. Subject K. The voicing component in /b/ and /z/ can be seen on the lower part of the sonagram; it is absent in /p/ and /s/. Compare the sharply defined vertical line of the stop explosions with the irregularly striated high-frequency area in the continuants.

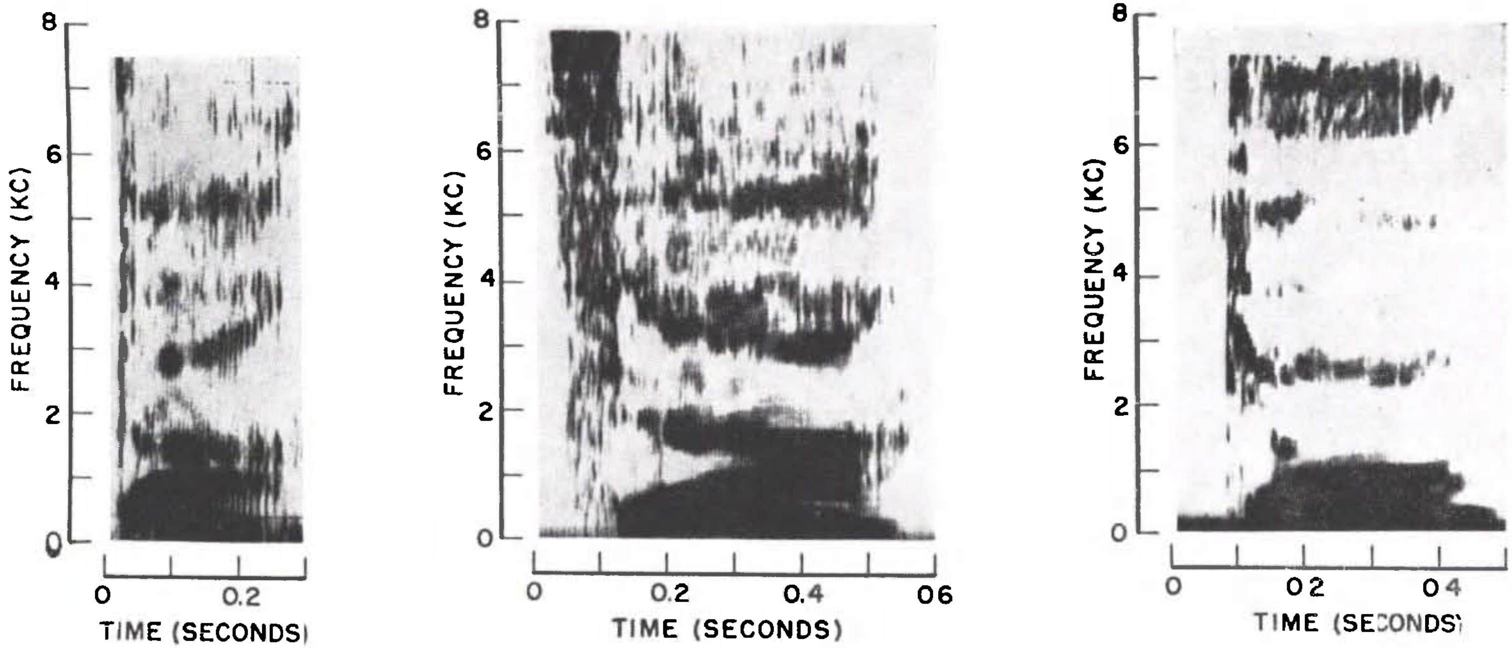


Fig. V-14. Sonagrams of syllables /ta/, /t,a/, Subject J, and /t,o/, Subject K. The plain /t/ has very little noise, whereas the sharped /t,/ has a considerable amount of noise.

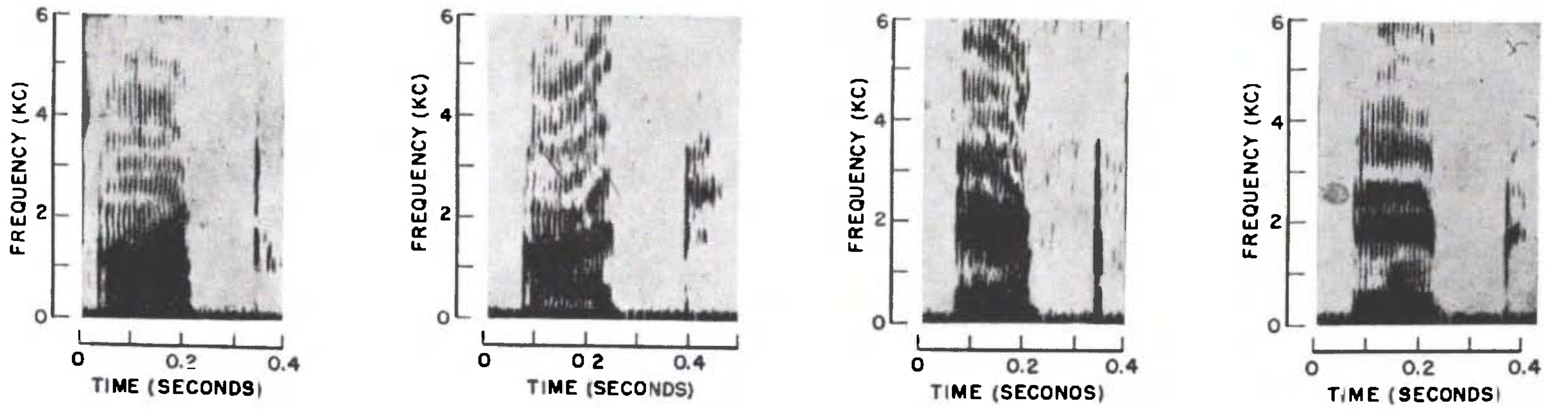


Fig. V-15. Sonagrams of syllables /aɔ/, /ap,/, /ep/, and /ɛp,/. Subject D. The transitions in /a/ are very positive before sharped /p,/ and practically zero before the plain /p/. On the contrary, in the acute /e/ they are very negative before the plain /p/ and practically zero before the sharped /p,/.

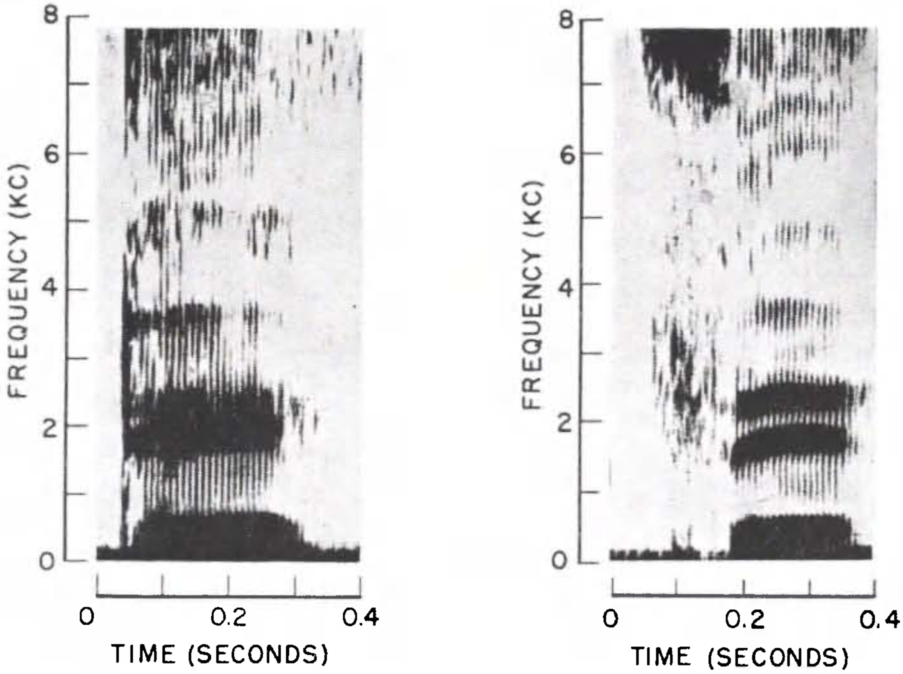


Fig. V-16. Sonagrams of syllables /te/ and /ce/. Subject D. The mellow /t/ is represented by a sharp vertical line showing that all of its energy is released at once. In the strident /c/ the energy is released over a long interval.

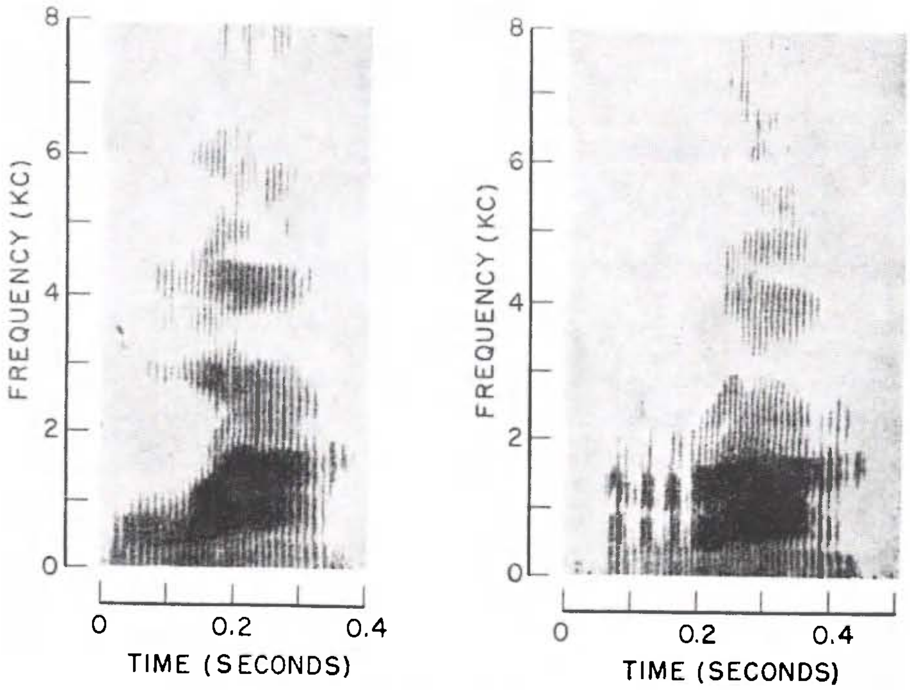


Fig. V-17. Sonagrams of the syllables /la/ and /ra/. Subject D. The /l/ is continuant. The /r/ is interrupted as shown by its intermittent silences.

7. *Continuant vs. Interrupted*

The manner in which the sound sets on is of significance in distinguishing between such minimal pairs as /k'or/ "peels" (gen. pl.) and /x'or/ "chorus" Continuants like /x/ have gradual onsets, while interrupted phonemes like /k/ are characterized by rapid changes of intensity in their onsets. Interrupted phonemes can be readily identified in the sonagram by the vertical white areas – the absence of energy in all frequencies above the voicing components – which are preceded and/or followed by vertical black lines. (See Figs. V-1, V-13, V-14, V-15, V-16.)

The feature of interrupted-ness is distinctive primarily among the compact consonants, where it distinguishes the stops /k/, /k,/, /g/ and /č/ from the fricative consonants /x/, /š/, and /ž/. Among the diffuse consonants it is distinctive only in differentiating the stop /c/ from the continuants /s/, /s,/, /z/ and /z,/, and among the liquids it distinguishes the interrupted /r/ and /r,/ from the continuant /l/ and /l,/. In addition to its distinctive function the feature of interrupted-ness plays an important role as a redundant feature for the mellow stops, which are all interrupted.

7.1 *Consonants: Continuants vs. Stops.* Stops sounds are produced by a rapid closing and/or opening of the oral cavity while the nasal cavity is closed. The articulatory gesture characteristic of all stops is, therefore, a closed oral cavity followed and/or preceded by rapid articulatory movements. This gesture has the following acoustical consequences: while the oral cavity is closed no sound radiates from the lips and hence no acoustical energy is emitted except for the low frequency vocal cord vibration, which is radiated from the face and neck of the speaker.[30] The "silence" in all except the "voicing" frequencies (below 400 cps) can be very clearly observed on sonagrams of stop consonants, where it appears as a vertical white band. (Cf., Fig. V-1.) The rapid articulatory movements following a "silence" are ordinarily accompanied by a short burst of sound, the stop explosion. If the stop is adjacent to vowels the rapid articulatory movements cause fast transitions in the vowel formants. In the latter case, the explosion may be omitted altogether. In order to determine whether or not a given sound is a stop we must be able to identify a "silence" preceded by rapid formant transitions and/or followed by a burst and by rapid formant transitions.

No particular difficulty is associated with identifying a "silence" In contrast to this the problem of what constitutes "rapid formant transitions" could not be tackled here since an examination of data gathered from a fair-sized corpus failed to yield useable results. It seems likely that no real solution can be found until information is made available concerning the perception of signals produced by time varying resonance circuits.[31]

[30] It is to be noted that the "silence" is an indispensable part of the stop. If the "silence" interval is filled by sound, a stop is not perceived. This is well illustrated in the recent experiments of Malécot on the perception of nasal consonants, which were synthesized exactly like the cognate stops except that the "silence" was filled with special "nasal" formants. Cf., Malécot, *op. cit.*

[31] For further discussion of this point see Halle, Hughes and Radley, *op. cit.*, 116–117, and A. M.

The remaining stop cue is the burst. In order to identify the stop burst it is necessary to investigate the temporal properties of the sound following upon a "silence". This is done most easily by examining the envelope of the rectified wave form. Since minor amplitude variations are of no interest here, the envelope was smoothed. It was proposed to use an oscillographic trace of this envelope as the basic data. Difficulties arose, however, with voiced consonants because the periodic vocal cord vibrations were so strong as to obscure the temporal properties of the high frequency components. High pass filtering at 500 cps with a filter with very steep skirts suppressed the component sufficiently in the voiced stops. In voiced continuants, however, the low frequency component could not be eliminated by this simple expedient, and a more elaborate scheme was attempted, which, however, will not be described here, since I am uncertain of its relevance.[32]

The envelopes were then subjected to the following tests in order to determine whether they were stops or continuants:

(1) The time at which the wave reached 20% of its ultimate peak value was noted.

(a) If within 150 msecs before this point there was a silence which in turn was preceded by a sound, the phoneme was judged a stop.

(b) If there was no silence in this space, the phoneme was judged a continuant.

(c) If no sound preceded the silence, we proceeded to part (2).

Criteria (1a) and (1b) are sufficient for most stops and fricatives in non-initial position. The silence which precedes the explosion of a stop was never found to be longer than 150 msecs; and in connected discourse it is usually much less.[33] There is, of course, no intervening silence between continuants and phonemes adjacent to them, unless a stop follows the continuant.

(2) We measured the length of the time during which the wave remained at a level of 70% of the peak. Minor maxima were eliminated by fairly heavy smoothing (time constant of 40 msecs).

(a) If the time was less than 40 msecs, the phoneme was judged a stop.

(b) If the time was more than 40 msecs, the phoneme was judged a continuant.

Criterion (2) distinguishes phonemes which, in the rectified and smoothed presentation used, have the appearance of steep hills or spikes from those with gently sloping sides. The former are classed as stops; the latter, as continuants.

7.2 *Liquids.* In the case of the liquids the feature of interrupted-ness is implemented as follows: by vibrating the tongue, the mouth is opened and closed at a rate which varies from 10 to 30 times a second. This causes fluctuations in the intensity of the sound; in some cases there is acutally a complete silence lasting 15 to 20 msecs. A

Liberman *et al.*, "Tempo of Frequency Change as a Cue for Distinguishing Classes of Speech Sounds," *Journal of Experimental Psychology*, *52*, 127–137 (1956).

[32] Cf., M. Halle, *The Russian Consonants* (unpublished Harvard Ph.D. dissertation, 1955), pp. 135–139.

[33] In 68 English sentences spoken by four different subjects, J.-P. A. Radley found the silences to vary from 15 to 140 msecs, with the greatest number between 30 and 100 msecs. (Personal communication from J.-P. A. Radley).

single vibration is sufficient to produce a good [r] impression although some speakers have two or three vibrations. The continuant liquids have no comparable fluctuations in intensity. (See Figs. V-5, V-6 and V-17.)

8. Voiced vs. Unvoiced

The feature of voicing in consonants is one of the simplest to recognize: in the sonagram it is represented by a heavy dark line in the low frequency region (Fig. V-13). A rough impression of the intensity of this component can be gained from the fact that the region below 300 cps is about 10 db stronger in voiced than in unvoiced consonants compared to the total energy in the phoneme.

It is well known that voiced consonants are of lower intensity than unvoiced consonants.[34] No extensive measurements of intensity could be carried out, because we were not clear in our own minds as to how to conduct such measurements. It is evident that absolute values are not significant here since one can speak a voiced phoneme louder than an unvoiced one. Attempts to compare the intensity of the consonants to that of the adjacent vowel led to no conclusive results. The problem of intensity in speech, which, without doubt, plays a significant role in perception, needs to be studied in a fundamental way before a whole series of phonetic problems can be solved.

9. Sharped vs. Plain

The differences in articulation of the Russian sharped and plain consonants were characterized with great exactness by Broch in the following terms:

> While one articulatory group, the "soft" sounds, moves the tongue forward and thereby raises its upper part towards the front of the hard palate [cf. English "front"]; a second group is distinguished by the fact that it moves the tongue mass backward and raises its dorsum at different heights towards the soft palate [cf. the English expression "back"].[35]

X-ray studies of the pronunciation of these consonants have shown that the most striking changes in vocal tract configuration correlated with these different modes of articulation are a widening vs. a narrowing of the pharynx.[36] In some cases the plain consonants are not only velarized, but labialized as well. Since, however, labialization is not universal, it may be asserted, that the major modifications producing this distinction are those in the pharynx, i.e., behind the primary constriction in the vocal tract. This immediately raises the question as to what extent modifications in the cavity configuration behind the point of articulation can affect the acoustical output of the vocal tract.

[34] See, e.g., H. Fletcher, *Speech and Hearing* (New York, 1929), Table IX, p. 73, where every unvoiced consonant, except /θ/, has greater intensity than its voiced cognate.
[35] O. Broch, *Slavische Phonetik* (Heidelberg, 1911), p. 224.
[36] Cf., C. G. M. Fant, *Theory*. Also H. Koneczna and W. Zawadowski, *Obrazy rentgenograficzne glosek rosyjskich* (Warsaw, 1956).

In general, effects of the back cavity can only be neglected when the cavities in front of the constriction are virtually uncoupled from those behind. This obtains when there is total occlusion in the vocal cavity as, e.g., in the oral closure of stops and nasal consonants. As was mentioned above in sec. 4.2, there is reason to believe that the effects of the cavities behind the tongue constriction can be neglected also in the continuants. In view of this we should expect to observe no striking differences between spectra of the fricative noise of sharped and plain continuants. Relatively minor effects are due to the longer tongue constriction in sharped continuants as compared with the constriction in plain continuants. The back cavities begin to play a very important role, however, as soon as the occlusion or constriction is widened. Differences are, therefore, to be expected in the spectra of sharped and plain stop bursts, in the spectra of the period of oral closure in sharped and plain nasal consonants, and above all in the formant transitions.

Our data bear out these expectations. There were no consistent differences between the spectra of the fricative noise in sharped and plain continuants. Cf., Fig. V-11 and Appendix II.

In the burst spectra of /k/ and /k,/ there were clear differences. /k,/ had an energy maximum (peak) above 2000 cps, and /k/ below that frequency. Table V-4 presents the data for subjects K and J.

In the case of the bursts of the stops /t/ /d/ and /t,/ /d,/ the most striking differences lay in the fact that the latter were usually pronounced with a considerable amount of stridency (affrication). There were, however, instances of /t,/ with little affrication (Fig. V-14c). The burst spectra showed no consistent differences of these stops; nor could we find consistent cues for the distinction between /p/ and /p,/ and /b/ and /b,/ in the bursts spectra of the latter. (Cf., Appendix II.)

In the nasal consonants there were some differences for subject D in the spectra of the period of oral closure. The plain /m/ had a clear formant at about 800 cps, which was absent in the sharped /m,/.[37] The plain /n/ differed from the sharped /n,/ in that it had a stronger formant at 2000 cps. (This is somewhat surprising since we were inclined to think that the sharped phonemes have more energy than the plain phonemes in the upper frequencies, particularly above 2000 cps; i.e., in the region of the second formant of [i]). (Cf., Figs. V-8 and V-12.) For the three other speakers no consistent differences were observed.

In the liquids the spectra showed clear differences for subjects D and K. The data were difficult to evaluate in the case of the two remaining male speakers, J and I. The differences observed were the following: the sharped /r,/ usually had a second formant at about 2000 cps and a third formant above 2500 cps. The plain /r/ had a second formant at about 1400 cps and a considerably attenuated third formant located below 2500 cps. In some instances of plain /r,/, the third formant was altogether lacking. Similarly the sharped /l,/ had a second formant about 1800 to

[37] See fn. 25 above.

Adjacent to	/k/initial		/k,/initial		/k/final	
	K	J	K	J	K	J
/'u/	750	500	2250	3250	1000	9250
/'e/	675	500	2250	2750	1000	1000
/'a/	1500	1500	2750	?	1500	1500
/'e/			3000	2750	?	1500
/'i/			3250	3000	1500	1250
/'is/					1750	1250–1500[38]

Table V-4. Frequency position of the spectral peak of the bursts in /k/ and /k,/.

2000 cps and a third formant above 2300 cps, while the plain /l/ had a second formant below 1000 cps and no strong third formant.

9.1 *Formant Transitions.* The resonance (formant) frequencies of a cavity system are determined uniquely by its geometrical properties. Changes in the geometry, such as occur constantly during the production of connected discourse, are directly reflected in the acoustical signal as changes in the formant frequencies. The latter changes, which have been called *formant transitions*, contain important cues for the perception of speech sounds, particularly of stop and nasal consonants. In experiments with English speaking subjects it has been found that in consonants adjacent to vowels, transitions in the second, and, to a lesser extent, in the third formant[39] contain cues essential for the identification of the features of compactness and tonality; i.e., of the so-called "point of articulation".[40]

As has been remarked above, the Russian data collected for the present study do not show such a direct relation between formant transitions and these two features. This difference between English and Russian is probably due to the different role played in the two languages by variations in the shape of the pharynx. Whereas in English, the different configurations of the pharynx are closely correlated with different points of articulation; in Russian for most points of articulation there are two distinctively different pharyngeal configurations: a widened pharynx for sharped phonemes, and a narrowed or neutral pharynx for plain phonemes. The effects on the formant transitions of the two types of pharynx configuration appear to be greater than – and hence capable of obscuring – the effects due to different points of articulation. This explanation finds support in the results obtained by Fant in his calculations of the formant frequencies associated with different consonantal articulations (see Table V-5).

[38] The empty spaces in the table are due to the fact that Russian does not possess the corresponding syllables. The question marks indicate that it was impossible to interpret the data (absence of clearly defined maximum).

[39] The first formant is of little interest since it is primarily controlled by the area of the minimal constriction in the vocal tract. In the case of almost all sounds except vowels this area is so small that it can be assumed that the first formant goes down to zero frequency.

[40] Cf., for instance, F. S. Cooper *et al.*, "The Role of Consonant-Vowel Transitions in the Perception of the Stop and Nasal Consonants," *Psychological Monographs*, *68*, 8, 1–13 (1954).

Phoneme	Plain		Sharp	
	F2	F3	F2	F3
/x/	1350	2350		
/k/	1200	2800		
/k,/			2150	3200
/č/			1850	2750
/š/	1100	2000		
/s/ /t/ /n/	1400	2200		
/s,/ /t,/ /n,/			1900	3000
/f/ /p/ /m/	850	2350		
/f,/ /p,/ /m,/			1900	2400

Table V-5. Formant frequencies of consonantal articulations. Calculations made by Fant from radiographs of subject D.[41]

Fant concludes: "... all soft consonants have F2 positions higher than 1700 cps and all hard consonants have F2 positions lower than 1400 cps. For a single pair the difference in F2 position is of the order of 500 to 800 cps."[42]

The formant positions given above are the terminal (beginning or end) positions towards which the vowel formants would tend in a syllable beginning or ending with the corresponding phoneme. These frequencies are in agreement with the following observations made on the basis of a study of sonagrams of about a thousand Russian utterances.

Sharped phonemes are associated with more positive F2 and F3 transitions to the adjacent vowel, than are their plain cognates. This is actualized differently in the front vowels ([i] and [e]) than in the other vowels.[43] In the front vowels a zero transition is correlated with the presence of the sharping in the adjacent phoneme, while a negative transition is correlated with the absence of sharping in the adjacent phoneme. In the other vowels a sharply positive transition is correlated with the presence of sharping in the adjacent phoneme, and a zero transition, with its absence. Cf., Figs. V-8, V-12 and V-15.

These facts are in complete agreement with Broch's articulatory description quoted at the beginning of this section. Since in front vowels the pharynx is widened, no transitional effects are observed when a sharped phoneme is adjacent to a front vowel. The same absence of transitional effects is observed when plain consonants, articulated with a constricted pharynx (velarized), are adjacent to vowels articulated

[41] C. G. M. Fant, *Theory*

[42] *Ibid.*

[43] Following the usage established in the writings of the researchers at Haskins Laboratories, a transition is termed "positive" if its terminal [beginning or end] point is higher in frequency than the steady state position of the formant in the adjacent vowel; "negative" if it is lower in frequency; and "zero" if it does not differ substantially.

with a constricted pharynx. Striking effects are observed when there are drastic changes in the width of the pharynx.[44] The transitional effects explain why foreigners often perceive Russian sharped consonants as being followed by [j], and plain consonants as being followed by a [w]. They also confirm again the remarkable accuracy of the observations of A. Thomson, who wrote:

> ... in addition, a third fact characterizes the soft consonants, namely their falling transitional sound after the [i]-like explosion and their rising transitional sound before the [i]-like implosion. In our auditory perception the most important cue for the soft consonants are these transitional sounds, which only in position next to [i], and in part, next to [e], are less audible and, therefore, have less significance....[45]

Analogous effects have been observed in languages in which consonants are distinguished by the feature flat vs. natural; e.g., in Arabic, where emphatic consonants produced with a narrowed pharynx are opposed to nonemphatic consonants produced with a neutral pharynx. As is to be expected, in position next to emphatic consonants vowels exhibit more negative transitions than in position next to nonemphatic consonants; see, e.g., the sonagrams reproduced in *Preliminaries*, Fig. 7, p. 50.

[45] A. Thomson, "Über die weichen Konsonanten," *Zeitschrift für slavische Philologie*, *8*, 95–96 (1931). See also the remarks on the same subject in his *Obščee jazykovedenie* (Odessa, 1910), pp. 197–204, and in his "Fonetičeskie ètjudy," *Russkij filologičeskij vestnik*, *54*, 231 (1905).

CONTEXTUAL VARIANTS OF THE RUSSIAN VOWELS

by

LAWRENCE G. JONES

CONTEXTUAL VARIANTS OF THE RUSSIAN VOWELS

by

Lawrence G. Jones

1. Introduction

One of the most striking characteristics of the Russian vowels is their variation according to context. This characteristic is noted by a number of investigators, some of whom tend to exaggerate the amount of variation but few of whom tend to underestimate it. There is disagreement regarding the phonemic status of certain variants, especially the phonemic status of the Russian vowel whose phonetic manifestation is usually symbolized as [ɨ]. Some feel that it is distinct from the phoneme /i/, while others, including the present writers, consider it only to be a variant of /i/. Because of this and other related problems, it will be well to consider certain factors involved in the relationship between phonetics and phonemics before carrying out a detailed description of the vowel variants. In particular we want to discuss the inevitability of having a phonemic hypothesis before carrying out a phonetic analysis whether it be in articulatory or acoustic terms.

2. Phonetic and Phonemic Analysis

There seem to be two schools of thought about the criteria to be used in making a phonemic analysis. One school maintains that physical criteria should be used and the other says that they should not be used.[1] We maintain that if physical criteria, by which we mean either articulatory or acoustic data, are not admitted then even the preliminary segmentation of the speech is impossible. One is limited to the statement that if there are recurrent signs in speech and they are not infinite in number, then some must be considered the same sign in that language system, and others different. And this would hold for all languages. But the moment one asks for the properties by which one classes the signs as the same or different, one either does this in terms of physical properties, or one does not do it at all. Even the case of dead languages, known only through written texts, must rely on the very concrete examination of quite

[1] For a brief discussion of this "algebraic view" of phonemic solution, see R. Jakobson and M. Halle, *Fundamentals of Language* (= *Janua Linguarum*, No. I) ('s-Gravenhage, 1956), 15.

physical ink marks, although this is a rather trivial example. Even the most inadequate phonetic transcriptions are based on the identification of varying physical features heard or recorded in the speech waveform.

As a matter of fact, the task of making a phonetic or phonemic analysis might be a great deal simpler if it were possible to forget the physical data, since dealing with them raises a number of problems, chief among them being how to obtain the data, how to interpret them and the relation of phonemic analysis to the interpretation. To illustrate this, consider an experiment in which no phonemic criteria are brought into play at all. Suppose someone had tape recordings made of several people each speaking about two hundred samples of speech. The people are not told what kind of samples are wanted. No one keeps any kind of notes on what has been recorded. Then a rather deaf person who can just about tell the difference between sound and silence succeeds in making spectrograms of all the data. He knows approximately what vowels look like on a spectrogram and he measures the frequency position of what appears to be the two lowest formants of the vowel (although he is not absolutely sure) at what appears to be the steady state of the vowel (although he is not always sure of this either).

For each sample he now has two numbers and wants to classify the vowels into groups. Perhaps he makes a diagram like the one below, p. 162, with the first formant plotted on one scale and the second formant on the other. Thus he has a single dot on his chart for every vowel in his data. When he has finished he will have a group of dots all over the diagram. There are two extreme possibilities of clustering of dots on this diagram, one highly improbable and the other highly probable. In the highly improbable case, the dots will cluster into separate regions, each neatly spaced from the rest. If this were to happen, what conclusions can he draw? From a phonemic point of view he can say nothing, since it is possible that two separate regions might represent the same phoneme. This would be highly probable if the language were Russian and the only samples were /put,/, /mi/ and /b,it,/, repeated over and over again. (Remember that in this hypothetical experiment no one has told the speakers that they must use a wide range of contexts, nor that they should not repeat samples.) In phonetic terms there would be a cluster of dots for [u], another for [ɨ], and another for [i], but according to our phonemic analysis, only two phonemes represented. Thus the only conclusion could be that in this experiment the sounds tend to cluster.

In the other, highly probable case, where the dots are spread over the entire diagram, the analyst faces new problems. He might notice that dots tend to be more dense in certain areas and lighter in others. He might want to set threshold values where the dots appear lightest so that he can say that the majority of one kind of sound appears in an area inside or outside a certain threshold value. But he has no way of knowing the success of his threshold values, since the conclusions are the same as those given above for neatly separated groups. The point of this illustration is to show that some phonemic hypothesis must be included in the experimental design if one wishes to derive phonemic conclusions as a result of the experiment.

Suppose now, that we change this experiment by asking the speakers to make lists of words illustrating all of the possible combinations of sounds in their language. (Procedures could easily be devised to make sure that all sounds were included and all contexts represented.) The fact that the speakers use this language to communicate with one another implies that they recognize certain sounds as systematically alike and other as systematically different. But we hasten to add that these speakers are not necessarily conscious of what a linguist would call phonemic distinctions. They are merely aware of recurrent physical tokens of the same types of sounds.

Now when they record their samples, a list of their samples is kept, along with a record of their choice of "sames" and "differents", so that the analyst can classify them in a variety of ways. Now when he plots his data he is at least able to group the data according to what the speakers consider to be the same and what they consider to be different. If the data separates into neat groups he can tell whether or not each group represents a different sound. In the Russian example mentioned above, he would now know that there had been a change in context among the samples and would ask for more samples.

Where the data do not separate into neat groups he is on safer ground regarding the choice of thresholds, although, as we shall see in a moment, this new experimental design is far from satisfactory. But at least he can determine his thresholds so that he gets the maximum number of sounds in one area. Suppose, for instance, that a large group of [e] dots occur in a region below 400 cps on the first formant scale, and a large group of [a] dots occur in a region above 800 cps on the same scale, but there are several [e] dots and [a] dots in the intervening region. Then he sets his threshold at some point between 400 and 700 cps so that as many of the ambiguous dots will be included in with the rest of their brothers.

Chances are that the setting of thresholds will not be this simple. The analyst is sure that he has a group of sounds which are phonetically different according to his speakers, and he has measurements which illustrate tentative differences. But it will probably turn out that some thresholds are much easier to set than others. In this case he can examine the contexts in which the sounds occur and test them for redundancy or predictability. As we shall see, this phonemic analysis all by itself, unrelated to phonetic data, can yield strange results.

If the language is Russian, he will find that certain vowels are always preceded by a palatalized consonant, while others are always preceded by an unpalatalized consonant. For lack of further evidence, this situation is ambiguous, since either the consonant or the vowel could be redundant in this case. If he decides that the vowels are distinctive, but palatalization in the consonants is not, further examination of contexts would show him that he cannot rule out palatalized consonants completely since they occur in opposition to non-palatalized consonants in absolute final position. On the other hand he might consider the palatalization of consonants distinctive in all positions and the vowel differences redundant. This would be a neater and simpler solution.

Now he can go back to his measurements and try to set his thresholds on the basis of his phonemic categories. One of the two arrangements should lead to a better set of thresholds. In the case of our own analysis discussed below, it is easier to set thresholds for five vowels rather than approximately 10 as would be the case where palatalization is considered redundant.

The point of this illustration is to show that certain phonemic hypothesis fit the phonetic facts better than others.

3. The Russian Vowel Variants

In the following discussion of the Russian vowel variants our acoustic analysis is compared with the articulatory analysis made by Trofimov and Jones.[2] There are a number of reasons for limiting our choice to their study. In the first place, our analysis treats the vowel system as a whole and examines the systematic trends in the changes of vowel quality as a function of contextual change. Therefore we must compare it with another such analysis which treats variations within the whole system. Other acoustic analysis of Russian vowels are either based on crude or outmoded types of recording; or are so fragmentary or restricted in nature considering the vowels only as they are spoken in isolation, or give examples of only one or two vowels in a restricted number of consonantal contexts.

In the second place, modern acoustic analysis affords a wealth of detail which is usually unparalleled in articulatory studies. In this respect the work by Trofimov and Jones is most useful since it lists more variants per vowel phoneme than do most other studies.

Finally, and perhaps most important, their study can most easily be translated into acoustic terms. They use the IPA Cardinal Vowel Quadrilateral as the basis for classification, and the cardinal vowels of this quadrilateral have recently been synthesized acoustically at the Haskins Laboratories.[3] Thus we have what appear to be the frequency positions of the formants for each of the cardinal vowels. The following diagram shows the formant frequency positions for several of the cardinal vowels as given in a report of the Haskins Laboratories.[4] A glance at this diagram leads to the following correlations:

(1) As the tongue position goes from high to low, the frequency of the first formant becomes higher. Thus if the terms "high" and "low" are used for both articulatory and acoustic descriptions, one can say that the articulatory position is the reverse of the acoustic description, since a low tongue position is correlated with a high first formant and vice versa.

[2] M. V. Trofimov and D. Jones, *The Pronunciation of Russian* (Cambridge, 1923).

[3] P. Delattre, A. M. Liberman, F. S. Cooper, and L. J. Gerstman, "An Experimental Study of the Acoustic Determinants of Vowel Color; Observations on One- and Two-Formant Vowels Synthesized from Spectrographic Patterns," *Word*, *8*, 195–210 (1952).

[4] The data are from Delattre *et al.*, *op. cit.*, Table I, 198.

(2) As the tongue goes from front to back, the frequency position of the second formant becomes lower. Thus front vowels have high second formants, while back vowels have low ones.

These conclusions are in agreement with the guarded conclusions of House and Stevens.[5] In their experiments with an electronic vocal tract, they examine the relations between changes in articulatory positions and changes in formant distribution. Their statement is more elaborate than those given above, and more germane to our treatment of the articulatory distinctive features of the Russian vowels. They find that the position of the first formant can be related, as above, to tongue height and/or lowering of the jaw. The more open the jaw, and/or the lower the tongue, the higher the first formant. Thus open vowels have high first formants, close vowels low ones. Secondly, they find that as the tongue goes from front to back and/or the lips are increasingly rounded, the second formant falls. Thus a low second formant is correlated with lip rounding as well as with a back position of the tongue.

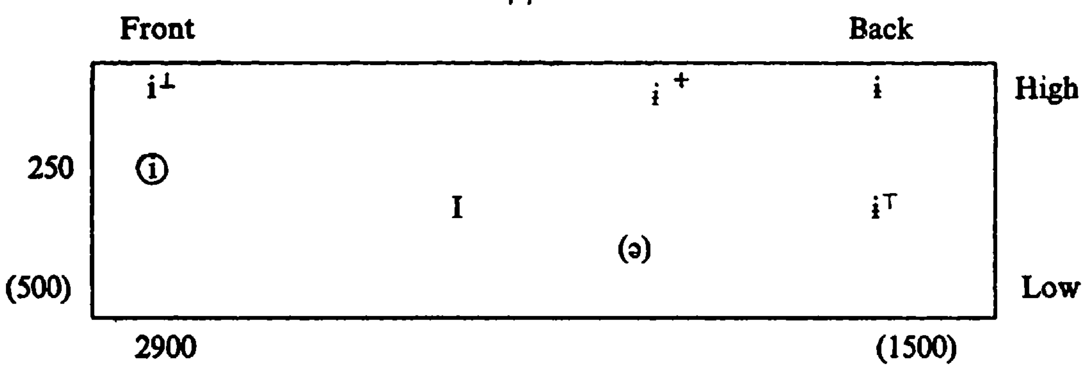
This is important for our discussion of the Russian vowels because we shall want to say that the open-close opposition, and the rounded-unrounded opposition is distinctive in Russian, but that front-back and high-low tongue positions are not. For instance /u/ is distinct from /i/ by the feature of rounding, but [ɨ] is not distinct from [i], their chief difference being a front-back tongue variation. Both are unrounded and hence distinct from any member of the /u/ phoneme.

For the sake of convenience, the variants are treated in groups, each group representing a different vowel phoneme. Each section contains a diagram of the articulatory area in which this set of variants occurs according to Trofimov and Jones. The formant frequency limits is given for each diagram and is based on the data given by the Haskins Laboratories. The circled vowel in each diagram represents a cardinal vowel and is the basis for determining the frequency limits of each diagram. In cases where diagrams do not terminate in a cardinal vowel, it was necessary to guess the approximate value of the frequency at that point. Such cases are given in parenthesis. The contexts in which the variants occur are based on the descriptions given by Trofimov and Jones, with certain reservations and additions which are noted.[6] Then samples of such contextual combinations from our data are discussed.

K. N. Stevens, and A. S. House, "Development of a Quantitative Description of Vowel Articulation," *Journal of the Acoustical Society of America*, *27*, 484–493 (1955).

[6] The articulatory and contextual descriptions are taken from Trofimov and Jones, *op. cit.*, Chapter VIII and examples in Chapter XXV.

/i/



[i] This is considered to be the principal implementation of the /i/ phoneme. It occurs in absolute initial position (#V...); preceded but not followed by a palatalized consonant (C,VC...).

[i]$^{\perp}$ occurs only when surrounded by palatalized consonants (C,VC,).

[ɨ]$^{+}$ occurs preceded by a non-palatalized and followed by a palatalized consonant (CVC,).

[ɨ] occurs preceded by a non-palatalized consonant and followed by a non-palatalized consonant or pause (CVC), (CV#).

[ɨ]$^{\top}$ = [ɨ] in unstressed position. (From the examples given, it would appear that this only appears in pretonic position.)

[ɪ] Trofimov and Jones consider this to be an unstressed variant of /e/. In our analysis, there is no such category since /e/ only occurs in accented positions.

[ə] Trofimov and Jones consider this to occur in unstressed syllables where cognate words have [ɨ]: [g'orə] "mountains" as opposed to [gar'ɨ] "of the mountain". This problem is taken up below in connection with certain unstressed variants of /a/.

/i/

Samples

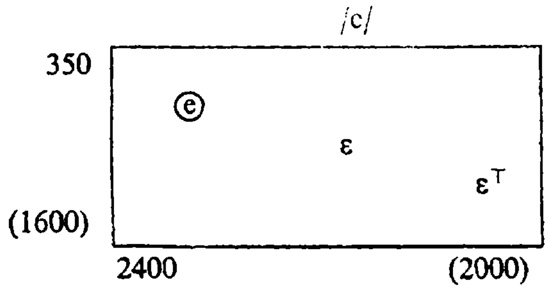
	[i]$^{\perp}$	[ɨ]$^{+}$	[ɨ]	[ɨ]$^{\top}$
F_1	400	300	200	300
F_2	2300	1600	600	1500
	(J /stat,ist,'ičiskaj/)	(J /m'il,u/)	(J /si/)	(J /pil,iv'oj/ first vowel)

[i]	[ɪ]	[ə]
150	200	400
2400	2100	1900
(J /s,i/)	(J /t,išin'a/, first vowel)	(J /v'iriti/, last vowel)

Discussion

As appears to be the general tendency with the variants of all the vowels when they are compared in phonemic groupings, the most striking variation from sample to sample among the /i/ variants is in the second formant. The place of these variants in the diagram above would closely pattern the articulatory features where it is the correlate of the second formant, namely, the front-back tongue position which is responsible for most of the difference among the variants. There is less variation in the first formant than in the second formant, the correlate of the first formant being either high-low tongue position or open-close. Hence it would be easier to set a threshold for the first formant to include this group of variants into one phoneme. To separate them in terms of separate phonemes according to their first formants would be almost impossible. This group of samples could easily be separated from one another in terms of their second formant, but as more contexts are added this becomes increasingly difficult, since the second formant of the [ɨ] type variant approaches that of the [i] variants when the [ɨ] is preceded and, or followed by acute consonants, as for instance [tɨt].

One characteristic of the [ɨ] variants is not shown by mere formant frequency readings. They tend to be quasi-diphthongal in nature, the second formant being quite low in frequency at the beginning of the vowel (some times as low as 1200 cps) and rising toward a typical [i] formant distribution at the end. Hence it is difficult to determine a steady state in such cases and the formant frequency readings given above and elsewhere for such vowels represent a point somewhere in the middle of the vowel duration.



(All forms of this vowel occur only in stressed position)

[e] occurs between palatalized consonants (C,VC,).

[ε] occurs preceded by a palatalized consonant or /j/ and followed by a non-palatalized consonant or a pause (C,VC), (C,V#), (jVC), (jV#).

[ε^T] occurs initially or when preceded by a non-palatalized consonant (#V...), (CV...).

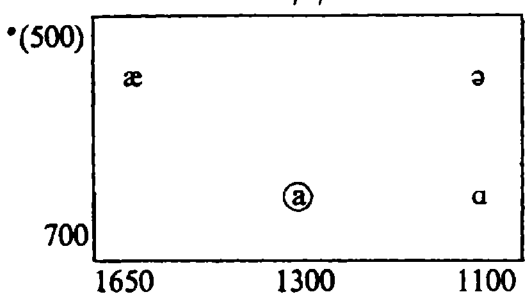
Samples

	[e]	[ɛ]	[ε^T]
F_1	450	400	450
F_2	2200	2025	1625
	(D /t,et,/)	(D /t,e/)	(D /ep/)

Discussion

Our data show that as in the case of the /i/ variants, the second formant reflects the backward movement of the tongue from one variant to the next, by becoming successively lower. On the other hand, the first formant remains quite constant. In VC positions this vowel tends to have the quality of an open or back [ε^T], while in VC positions, it is more like [ɛ]. Compare D's /ep,/ (F_1 — 400, F_2 — 2000 cps) with /ep/ above. In such cases there appears to be a tense-lax relationship between the two variants which is not clearly observable for any of the other vowels.

/a/



[a] This vowel appears to occur in stressed position and strong unstressed positions, that is in unstressed vowels which are either in absolute initial position or immediately pretonic. This variant is never preceded by a palatalized consonant (#V...), (CV...).

[æ] occurs in stressed position when preceded and followed by palatalized consonants.

[ɑ] According to Trofimov and Jones this is a retracted form of [a] and occurs only where the following consonant is /l/.

[ə] occurs in weak unstressed position, i.e., all positions other than absolute initial or pretonic, and may be preceded and/or followed by a non-palatalized consonant.

Samples

	[æ]	[a]	[ɑ]	[ə]
F_1	750	700	700	550
F_2	1700	1375	1500	1100
	(D /t,at,/)	(D /ta/)	(D /al/)	(D /papad'at,/, first vowel)
		750	650	
		1300	1200	
		(D/at/)	(D/ap/)	

Discussion

The second formant shows the expected relationship between [æ] and [a], but not for [ɑ] as described by Trofimov and Jones. That is, in the case of [ɑ] one would expect a lower second formant than for [a] but this is not usually the case, although it is true that the second formant tends to be lower when the vowel is followed by labials; cf., /ap/ as opposed to /at/ above.

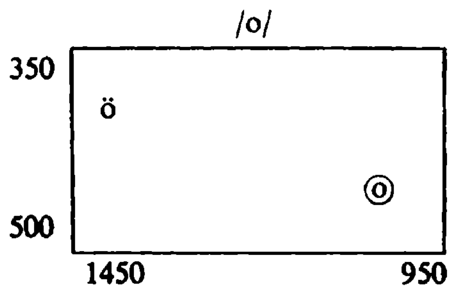
The [ə] variant presents a rather complex situation. The weak unstressed vowels are extremely short in duration and vary greatly according to their consonantal contexts. They range from an [ɨ] type formant pattern to the typical [ə] formant distribution given above. Compare for instance the readings for D in the first vowels of the following words:

/tatarv'a/:	450	/patam'u/:	425
	1625		1125
/tapar'i/	350	/pap,irat,/	650
	1600		1150

This range of variations occurs also in the final vowels of such words as /t'opat/ and /v'irit/:

/t'opat/	250	/v'irit/	350
	1250		1500

where, in the case of /t'opat/ an [ɨ] formant distribution is expected. Both formants are outside the range typical of the [ə] variant, and the second formant is completely outside the /a/ range.



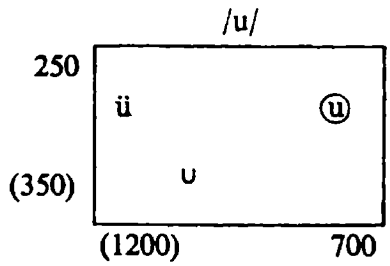
(This vowel, like /e/, occurs only in accented positions.)

[o] occurs when preceded by a non-palatalized consonant or pause (CV...), (#V...).

[ö] occurs when preceded by a palatalized consonant, or between palatalized consonants (C,V...), (C,VC,).

	[o]	[ö]
F_1	400	500
F_2	900	1200
	(D /t'opat/)	(D /t,'ot,am,i/)

The second formant follows the usual pattern: high when the vowel is preceded by a palatalized consonant, lower when it is not.



[u] occurs initially or when preceded by a non-palatalized consonant and followed by a non-palatalized consonant or pause.

[ü] when preceded and followed by a palatalized consonant (C,VC,).

[ʋ] follows the same distribution as [u] only is unstressed.

Samples

	[ü]	[ʋ]	[u]
F_1	350	250	250
F_2	1100	850	675
	(D /t,ut,/)	(D /tupav'at/)	(D /tu/)

The variation in the second formant is greater than it is in the first. The variation shows the front-back correlation posited by Trofimov and Jones.

The two formants of this vowel phoneme are extremely close in a low frequency region and it is often difficult to divide one from the other, that is, to tell where one ends and the next begins on the frequency scale.

4. General Remarks

The one striking characteristic which runs throughout this comparison of the vowel variants, is the range of variation of the second formant. If we find any basis for correlating varying formant distributions with the articulatory positions presented by Trofimov and Jones, it is usually in terms of the second formant, where the correlate of a front-back tongue position is expressed acoustically. The first formant within variants of a given phoneme does not appear to have any significant variation. Because of this lack of significant variation in the first formant, it would be next to impossible to set threshold values to separate all of the phonetic variants from one another in terms of the first formant. But the formant distribution of the variants is such that it is fairly simple to set thresholds as we have done, so that various degrees of opening are separated in terms of the first formant: /i/, /u/ separate from /e/, /o/, and both groups from /a/, while the second formant can be used to distinguish /e/ from /o/ and /i/ from /u/, where there is a wide range of difference between the second formants of these pairs, resulting in practically no errors.

Appendix I.

Formant Frequencies of Russian Vowels.

In the following pages are presented the results of formant frequency measurements on Russian vowels in different contexts. The measurement procedure is outlined in Chapter V, sec. 2.11. The data are arranged in two parts: A. accented vowels in CV and VC syllables and B. unaccented and accented vowels in words. The complete results for speakers D, J, and K are given in that order. See also Tables V-1, V-2 and V-3.

A. Formant Frequencies of Accented Vowels in CV (left hand side of page) and VC (right hand side of page) Syllables.

Speaker D.

Context			e	a	o	u		e	a	o	u
p—	F1	200	400	700	400	275	200	450	650	350	300
	F2	1475	1875	1250	700	550	2100	1625	1250	750	700
	F3	2125	2500	2200	2125	2150	2700	2250	2200	2200	—
p,—	F1	150	425	700	500	300	250	400	550	350	300
	F2	2150	1900	1375	1000	575	2150	2000	1475	1025	700
	F3	3000	2625	2250	2200	2500	3000	2600	2400	2275	—
b—	F1	275	525	725	500	250		×			
	F2	2150	1675	1125	750	550		×			
	F3	3000	2325	2300	2350	2200		×			
b,—	F1	175	450	650	500	250	×			×	
	F2	2100	1925	1300	1075	550	×			×	
	F3	2850	2725	2250	2100	2750	×			×	
t—	F1	250	350	700	450	250	200	450	750	500	200
	F2	1975	1800	1375	725	675	2100	1800	1300	1100	500
	F3	2250	2350	2200	2250	2300	2900	2400	2250	2250	—
t,—	F1	200	400	700	475	250	200	400	600	275	200
	F2	2200	2025	1450	950	650	2200	2000	1550	725	500
	F3	2950	2700	2100	2200	2200	3000	2600	2200	2150	—
d—	F1	250	350	700	400	150	×	×	×	×	×
	F2	1750	1800	1300	800	600	×	×	×	×	×
	F3	2475	2350	2200	2400	2100	×	×	×	×	×
d,—	F1	150	400	650	550	250	×	×	×	×	×
	F2	2150	1900	1375	950	650	×	×	×	×	×
	F3	2900	2625	2200	2150	2150	×	×	×	×	×
k—	F1	×	×	725	450	250		—			—
	F2	×	×	1325	750	500				—	—
	F3	×	×	2150	2100	2100			—	—	
k,—	F1	250	375	750	450	250	×	×	×	×	×
	F2	2125	1850	1600	1100	800	×	×	×	×	×
	F3	2925	2450	2350	2250	2150	×	×	×	×	×
g—	F1	×	×	700	500	250	×	×	×	×	×
	F2	×	×	1350	800	575	×	×	×	×	×
	F3	×	×	2150	2150	2100	×	×	×	×	×
g,—	F1	175	350	×	×	×	×	×	×	×	×
	F2	2150	1950	×	×	×	×	×	×	×	×
	F3	3100	2500	×	×	×	×	×	×	×	×
s—	F1	300	450	600	400	250	250	500	600	300	250
	F2	2050	2000	1200	700	550	2400	2000	1150	650	650
	F3	2500	2600	2200	2200	—	3200	2700	2200	2100	—
s,—	F1	300	500	650	450	250	200	400	700	300	200
	F2	2350	2050	1400	950	550	2200	1850	1200	700	400
	F3	3100	2650	2200	2200	—	2800	2400	2100	2000	—

Speaker D.

Context			e	a	o	u		e	a	o	u
z—	F1	350	550	800	400	200			×		×
	F2	1100	1950	1150	1200	700			×		×
	F3	2400	2150	2400	2300	2600			×		×
z,—	F1	300	400	750	500	300			×		×
	F2	2350	2200	1550	1000	550			×		×
	F3	3050	2800	2300	2300	2200			×		×
f—	F1	200	500	600	600	350	300	500	700	350	200
	F2	1600	1800	1300	1000	750	2250	1800	1200	750	550
	F3	2200	2400	2200	2750	2200	3000	2500	2200	2400	—
f,—	F1	250	600	700	700	250	200	400	700	400	200
	F2	2100	2100	1400	1100	650	2200	1900	1250	900	500
	F3	2700	2700	2200	2250	—	2900	2500	2200	2200	—
v—	F1	300	450	750	350	200	×	×	×	×	×
	F2	1700	1800	1200	750	600	×	×	×	×	×
	F3	2200	2300	2200	2200		×	×	×	×	×
v,—	F1	200	400	700	450	250	×	×	×	×	×
	F2	2100	1900	1450	950	550	×	×	×	×	×
	F3	2900	2400	2150	2200	2200	×	×	×	×	×
š—	F1	200	650	750	400	175	250	400	650	400	250
	F2	1600	1900	1300	900	500	2200	1750	1200	800	500
	F3	2200	2400	2300	2200	—	3000	2400	2200	2200	—
ž—	F1	200	500	900	400	200	×	×	×	×	×
	F2	1900	1800	1500	900	650	×	×	×	—	×
	F3	2300	2250	2400	2200	2200	×	×	×	×	×
x—	F1	200	500	600	300	250	250	500	750	400	200
	F2	2100	1700	1300	750	650	2200	1800	1300	800	600
	F3	2900	2200	2250	2000	—	3000	2500	2200	2200	—
c—	F1	200	400	700	400	150	200	400	600	350	150
	F2	1900	1800	1300	800	450	2200	1800	1150	750	400
	F3	2250	2450	2200	2100	—	2800	2500	2200	2250	—
č—	F1	200	300	700	350	250	200	350	700	400	250
	F2	2100	1800	1300	800	500	2200	1700	1250	850	550
	F3	3000	2200	2100	2000	—	2900	2350	2200	1900	—
šč—	F1	250	500	800	400	250	350	450	800	400	200
	F2	2200	1850	1500	850	650	2250	2100	1300	750	550
	F3	2700	2500	2300	2000		2900	2700	2300	2200	—
r—	F1	300	700	800	500	300	300	700	800	600	300
	F2	1900	2000	1550	1000	600	2250	1800	1450	950	700
	F3	2300	2500	2300	1800	—	2800	2500	2300	2100	—
r,—	F1	250	500	800	650	300	300	600	800	700	250
	F2	2250	2200	1700	1100	750	2250	2100	1500	1100	600
	F3	2700	2600	2450	2000	2400	3000	2700	2400	2400	—
l—	F1	300	650	800	700	300	250	700	800	600	200
	F2	2100	1850	1400	1100	700	2200	1800	1400	900	650
	F3	2400	2600	2400	1800	2600	2900	2550	2600	2400	2400

Speaker D.

Context			e	a	o	u		e	a	o	u
l,—	F1	300	600	800	600	250	250	500	800	600	300
	F2	2200	1900	1600	1100	700	2200	2100	1350	900	700
	F3	2900	2800	2350	2250	2400	3000	2600	2400	2400	
m—	F1	200	650	850	550	250	250	650	650	500	250
	F2	1250	2100	1400	900	600	2100	1850	1200	900	700
	F3	2250	2700	2300	2300	2250	2800	2400	2400	2400	2400
m,—	F1	250	550	750	650	250	300	550	900	600	300
	F2	2300	2200	1950	1000	600	2250	2100	1500	1000	600
	F3		2800	2500	2300	1350	3000	2600	2400	2650	2400
n—	F1	200	700	750	400	150	300	650	750	600	250
	F2	2100	2000	1500	900	450	2200	1950	1250	1000	600
	F3	2400	2500	2400	2100		2900	2650	2300	2400	2300
n,—	F1	250	600	800	700	250	300	600	800	600	300
	F2	2300	2200	1700	1100	650	2200	2200	1400	900	650
	F3	2900	2600	2300	2250	2350	2800	2800	2400	2300	2300
j—	F1	250	450	800		250	250	350	400	350	250
	F2	2200	2000	1350		600	2200	2000	1200	750	550
	F3	2700	2500	2250		2100	3000	2600	2200	—	—
j—j	F1		400	800	500	250				×	×
	F2		2100	1600	900	650				×	×
	F3		2800	2200	2100	2150				×	×
	F1	300	450	—		350		×	×	×	×
	F2	2050	1900			750		×	×	×	×
	F3	2150	2500			2300		×	×	×	×
t,—t,	F1		450	750		350	×	×	×	×	×
	F2		2200	1700		1100	×	×	×	×	×
	F3		2800	2600		2350	×	×	×	×	×

Speaker J.

Context			e	a	o	u		e	a	o	u
p—	F1	300	600	750	350	250	200	500	800	500	350
	F2	1700	1950	1250	550	550	2250	1800	1300	800	750
	F3	2300	2350	2200	2550	2100	2450	2300	2200	1900	1900
p,—	F1	175	450	800	450	250	200	500	900	450	150
	F2	2100	1900	1450	750	750	2100	1800	1300	800	450
	F3	2800	2400	2200	1900	2000	2500	2200	2000	1900	1900
b—	F1	300	600	750	300	250		×	×	×	×
	F2	1950	2050	1225	550	500		×	×	×	×
	F3	2350	2400	2150	2400	2000		×	×	×	×
b,—	F1	200	600	775	350	250		×	×	×	×
	F2	2450	1800	1400	1000	500		×	×	×	×
	F3	3500	2300	2300	1950	1950		×	×	×	×
	F1	250	500	750	400	250	200	575	800	600	300
	F2	2100	1800	1400	600	550	2000	1800	1300	1000	700
	F3	2450	2450	2250	1800	2400	2400	2300	2300	2000	2200

Speaker J.

Context			e	a	o	u		e	a	o	u
t,—	F1	225	500	700	400	250	200	650	750	500	150
	F2	2200	1900	1525	725	550	2200	1925	1200	1000	550
	F3	2700	2150	2225	2300	2450	3400	2475	1900	2000	2200
d—	F1	200	375	650	325	200	×	×	×	×	×
	F2	1900	2100	1500	825	500	×	×	×	×	×
	F3	2300	2400	2300	1950	2400	×	×	×	×	×
d,—	F1	200	375	700	300	250	×	×	×	×	×
	F2	2250	2200	1350	650	500	×	×	×	×	×
	F3	3150	2700	2800	2200	2200	×	×	×	×	×
k—	F1	×	×	800	400	250	—	—	—	—	—
	F2	×	×	1350	600	600	—	—	—	—	—
	F3	×	×	2000	2600	2200	—	—	—	—	—
k,—	F1	250	400	675	350	250	×	×	×	×	×
	F2	2200	2150	1400	800	550	×	×	×	×	×
	F3	2850	2900	2100	2150	2000	×	×	×	×	×
g—	F1	×	×	700	450	250	×	×	×	×	×
	F2	×	×	1300	700	550	×	×	×	×	×
	F3	×	×	2150	1900	2200	×	×	×	×	×
g,—	F1	200	400	×	×	×	×	×	×	×	×
	F2	2200	2000	×	×	×	×	×	×	×	×
	F3	3275	2500	×	×	×	×	×	×	×	×
s—	F1	200	600	800	600	250	200	600	850	600	250
	F2	1600	1600	1750	1100	550	2400	1800	1500	1000	650
	F3	2200	2100	2350	2100	2400	2900	2200	2300	1900	2300
s,—	F1	150	400	750	400	350	100	400	800	400	150
	F2	2400	1800	1700	1100	650	2500	1800	1700	1000	500
	F3	2900	2500	2700	2000	2700	3000	2400	2500	1800	2300
z—	F1	250	500	800	600	200	×	×	×	×	×
	F2	2000	1750	1500	1000	650	×	×	×	×	×
	F3	2400	2000	2200	2100	2400	×	×	×	×	×
z,—	F1	150	450	750	350	150	×	×	×	×	×
	F2	2500	1800	1600	1100	700	×	×	×	×	×
	F3	2900	2200	2200	2200	2400	×	×	×	×	×
f—	F1	200	400	800	400	150	150	500	800	400	150
	F2	1800	1900	1350	800	600	2100	1800	1400	1000	500
	F3	2200	2300	2200	2100	2200	2800	2400	2200	2300	2700
f,—	F1	150	500	750	400	200	250	500	900	500	150
	F2	2400	1750	1800	1100	450	2100	1800	1400	800	450
	F3	2900	2100	2400	1900	—	—	2200	2300	2000	2400
v—	F1	250	500	900	350	300	×	×	×	×	×
	F2	1800	1850	1500	800	650	×	×	×	×	×
	F3	2200	2300	2350	1900	2300	×	×	×	×	×
v,—	F1	150	400	900	350	250	×	×	×	×	×
	F2	—	1900	1500	1050	500	×	×	×	×	×
	F3	—	2300	2400	2300	—	×	×	×	×	×

Speaker J.

Context			e	a	o	u		e	a	o	u
š—	F1	250	500	800	500	300	200	500	750	475	250
	F2	1900	1700	1400	950	650	2150	1800	1250	1100	600
	F3	2200	2100	2400	1800	2000	2600	2200	2200	1900	2400
ž—	F1	250	500	800	500	300	×	×	×	×	×
	F2	1500	1750	1400	1000	750	×	×	×	×	×
	F3	2000	2250	2100	2000	2200	×	×	×	×	×
x—	F1	200	500	900	400	250	225	500	800	500	250
	F2	2200	1900	1500	900	600	2200	1800	1400	950	700
	F3		2300	2400	2100	2200	2700	2250	2300	2000	2600
c—	F1	200	400	900	450	200	200	500	800	500	200
	F2	1900	1800	1500	950	700	2100	1700	1400	1000	500
	F3	2200	2200	2100	1900	2000	2500	2000	2100	2000	2500
č—	F1	300	450	900	500	200	150	500	900	400	250
	F2	1800	1900	1500	1000	650	2100	1800	1400	900	600
	F3	2000	2600	2100	2000	2400	—	2150	2000	2000	2200
šč—	F1	200	600	850	500	200	300	600	900		150
	F2	2200	1800	1600	1000	400	1800	1650	1400		400
	F3	2400	2300	2250	1900		2300	2050	2400	—	2400
	F1		600	1050	600	250	250	600	900	450	200
	F2		2000	1700	1000	500	2100	1800	1400	850	475
	F3		2500	2400	1900		2600	2300	2300	1900	2800
r,—	F1	200	550	800	550	200	250	600	900	400	200
	F2	2300	2100	1600	1000	400	2200	1900	1500	1400	400
	F3	2800	2300	2300	1900	2200	2700	2600	2200	2300	2700
l—	F1	250	450	800	500	300	250	650	700	500	250
	F2	1800	1800	1600	900	750	2200	1800	1500	1100	600
	F3	2200	2250	2300	2000	2300	2800	2400	2300	1800	200
l,—	F1	250		800	650	250	250	550	900	600	350
	F2	2000		1500	1200	550	2200	1800	1500	900	850
	F3			2400	2000	—	3100	2200	2200	2600	2500
m—	F1	300	800	900	700	300	400	650	800	600	300
	F2	1900	2000	1600	1100	750	2300	1800	1700	1000	775
	F3	2200	2600	2500	2200	2350	2600	2400	2500	2000	2250
m,—	F1	250	700	900	700	300	350	600	900	500	350
	F2	2150	1800	1600	1200	600	2400	2200	1500	1150	450
	F3	2500	2200	2600	2500	2500	2600	2800	2400	2400	2100
n—	F1	300	700	800	400	300	250	600	800	600	300
	F2	2400	2000	1600	750	700	2400	1800	1400	1100	600
	F3	2700	2600	2500	1900	2400	2800	2600	2500	1800	2700
n,—	F1	300	600	900	700	300	200	650	900	600	250
	F2	2300	2000	1700	1100	750	2250	1800	1500	1100	500
	F3	2500	2400	2400	1800	2400	—	2600	2800	1900	2800
j—	F1	500	600	900	600	350	250	600	900	600	250
	F2	2300	1800	1600	1000	900	2200	1650	1500	1100	600
	F3	—	—	—	—	2200	—	2100	2400	2000	2200

Speaker J.

Context			e	a	o	u		e	a	o	u
j—j	F1	—	650	600	750	350	×	×	×	×	×
	F2	—	2100	1600	1250	850	×	×	×	×	×
	F3	—	2800	2350	1800	—	×	×	×	×	×
t—t	F1	350	600	900	600	300	×	×	×	×	×
	F2	1700	1800	1700	1100	—	×	×	×	×	×
	F3	2200	2400	2600	1900		×	×	×	×	×
t,—t,	F1	350	700	900	650	300	×	×	×	×	×
	F2	2150	1900	1700	1200	900	×	×	×	×	×
	F3	—	2400	2500	2100	1500	×	×	×	×	×

Speaker K.

Context			e	a	o	u		e	a	o	u
p—	F1	375	700	850	450	400	250	725	1000	600	300
	F2	2350	2200	1350	950	900	2600	1900	1500	1000	700
	F3	2800	3000	2375	2100	2800	3300	2750	2250	2000	2900
p,—	F1	200	500	875	450	300	250	650	900	650	300
	F2	2700	2200	1625	1000	800	2200	2300	1600	1100	600
	F3	3300	2900	2625	2600	2950	3600	2800	2750	1950	2600
b—	F1	350	600	800	450	350	×	×	×	×	×
	F2	2300	2150	1500	850	800	×	×	×	×	×
	F3	2700	2975	2550	2700	2800	×	×	×	×	×
b,—	F1	250	650	825	600	350	×	×	×	×	×
	F2	2650	2250	1625	1050	900	×	×	×	×	×
	F3	3400	3150	2650	2650	3000	×	×	×	×	×
t—	F1	350	450	900	450	350	275	750	1100	650	300
	F2	2400	2050	1550	950	800	2500	2000	1750	1325	700
	F3	2900	2750	2675	2850	2950	3450	2600	2500	2350	3100
t,—	F1	350	475	800	450	400	250	550	850	625	325
	F2	2700	2450	1500	950	900	2750	2450	1800	1525	900
	F3	3300	3250	2300	2700	2800	3500	3300	2750	2825	2750
d—	F1	300	500	825	500	300	×	×	×	×	×
	F2	2200	2000	1450	1050	825	×	×	×	×	×
	F3	2800	3100	2600	2400	3000	×	×	×	×	×
d,—	F1	300	650	750	450	350	×	×	×	×	×
	F2	2650	2200	1475	1000	850	×	×	×	×	×
	F3	3100	2900	2300	2600	3000	×	×	×	×	×
k—	F1	×	×	850	450	400				—	
	F2	×	×	1450	900	700				—	
	F3	×	×	2200	2350	2900			—	—	
k,—	F1	300	375	950	525	300	×	×	×	×	×
	F2	2750	2200	1650	950	950	×	×	×	×	×
	F3	3500	3000	2600	2500	2750	×	×	×	×	×
g—	F1	×	×	650	450	400	×	×	×	×	×
	F2	×	×	1400	950	700	×	×	×	×	×
	F3	×	×	2500	2350	2900	×	×	×	×	×

Speaker K.

Context			e	a	o	u		e	a	o	u
g,—	F1	250	450			×	×	×	×	×	×
	F2	2650	2300			×	×	×	×	×	×
	F3	3350	2900			×	×	×	×	×	×
s—	F1	500	700	800	600	350	300	500	750	600	300
	F2	2200	2200	1500	1100	650	2450	1900	1250	950	650
	F3	2800	2700	2200	2000		2700	2750	2000	2000	—
s,—	F1	350	500	800	400	300	300	475	750	600	300
	F2	2650	2200	1650	800	600	2700	2100	1250	950	650
	F3	3200	2600	2300	2400	—	3600	2700	2000	2000	
z—	F1	500	600	800	600	400	×	×	×	×	×
	F2	2300	2000	1600	1100	600	×	×	×	×	×
	F3	2500	2600	2150	2000	—	×	×	×	×	×
z,—	F1	300	600	900	700	350	×	×	×	×	×
	F2	2200	2300	1800	1200	600	×	×	×	×	×
	F3	—	2700	2700	2400	—	×	×	×	×	×
f—	F1	500	600	800	500	300	300	650	900	550	350
	F2	2000	2200	1500	1000	600	2500	2100	1400	850	650
	F3	2600	3000	2600	2200		3200	2900	2400	1950	
f,—	F1	350	700	850	600	300	300	500	800	600	300
	F2	2600	2200	1650	1050	650	2700	2200	1600	1000	750
	F3	3200	2800	2300	2500		3500	2800	2400	2100	2900
v—	F1	400	600	900	400	300	×	×	×	×	×
	F2	1800	2200	1500	950	600	×	×	×	×	×
	F3	2500	2700	2500	1900		×	×	×	×	×
v,—	F1	550	850	850	500	300	×	×	×	×	×
	F2	2200	2200	1700	900	500	×	×	×	×	×
	F3	2650	2650	2300	2600	—	×	×	×	×	×
š—	F1	400	600	800	400	—	300	600	1000	500	250
	F2	2100	1900	1500	800		2500	1900	1600	900	500
	F3	2500	2500	2500	2100		3000	2400	2800	2100	—
ž—	F1	400	700	900	600	350	×	×	×	×	×
	F2	1800	2100	1600	1000	750	×	×	×	×	×
	F3	2600	2600	2700	2400	—	×	×	×	×	×
x—	F1	250	650	800	500	250	250	600	850	600	250
	F2	2600	2400	1300	1000	650	2600	2200	1400	1000	600
	F3	3200	3200	2250	2200		3000	3000	2600	2300	2300
c—	F1	600	750	800	800	400	300	800	750	400	500
	F2	2400	1900	1650	1300	700	2800	2300	1500	700	850
	F3	3100	2200		2600	2200	3600	3000	2800	2000	—
č—	F1	350	750	850	800	400	400	600	850	600	500
	F2	2600	1900	1500	1300	700	2600	1900	1500	900	800
	F3	3100	2200	2300	2200	—	2850	2400	2600	2000	3100
šč—	F1	300	650	750	700		250	550	800	500	250
	F2	2400	2100	1500	1100		2500	2200	1300	900	650
	F3	2800	2600	2500	2600	—	3200	3000	2500	2200	—

Speaker K.

Context			e	a	o	u		e	a	o	u
r—	F1	300	600	1000	600	500	250	700	1000	500	300
	F2	2100	2000	1600	1100	800	2250	2200	1600	1100	650
	F3	2800	2800	2600	2800	—	3250	3100	2300	2300	3000
r,—	F1	300	600	1050	700	300	300	700	1100	600	250
	F2	2800	2200	1750	1200	650	2800	2350	1500	1100	600
	F3	3200	2900	2850	2400	2900	3200	3150	2500	2400	2900
l—	F1	350	750	850	600	300	300	600	900	550	350
	F2	2300	2200	1350	1000	700	2650	2250	1600	1100	850
	F3	3100	3100	2800	2250	3100	3250	3200	2500	2350	3150
l,—	F1	325	700	1000	650	350	350	650	900	500	300
	F2	2800	2150	1600	1200	700	2600	2500	1600	1000	750
	F3	3200	3100	2800	2700	2600	3200	3200	2350	2400	
m—	F1	400	500	800	500	500	250	600	800	500	400
	F2	1800	2250	1400	900	900	2500	2200	1500	900	650
	F3	2700	3100	2400	2300	—	3000	3000	2250	2100	—
m,—	F1	300	500	800	650	400	250	475	800	500	250
	F2	2550	2250	1700	1000	700	2800	2300	1400	1000	550
	F3	2750	2650	2300	—	—	—	3200	2250	2300	2500
n—	F1	350	500	700	400	275	500	650	900	600	350
	F2	2200	2000	1500	950	700	2600	2200	1500	900	650
	F3	2700	—	2500	2200	—	3400	3100	2300	2300	
n,—	F1	250	650	700	600	250	300	600	900	400	300
	F2	2500	2150	1600	1100	550	2700	2300	1500	800	650
	F3	3200	3000	2400	2700	2500	3200	3200	2300	2200	—
j—	F1	250	500	850	400	250	250	400	750	400	250
	F2	2600	2400	1150	900	550	2650	2200	1350	900	500
	F3	3000	3100	2200	2300	—	3100	2900	2200	2200	—
t—t	F1	350	550	800	375	300	×	×	×	×	×
	F2	2100	1900	1600	800	700	×	×	×	×	×
	F3	2800	2800	2250	2200	—	×	×	×	×	×
t,—t,	F1	300	500	750	500	200	×	×	×	×	×
	F2	2400	2200	1500	1200	500	×	×	×	×	×
	F3	3000	3000	2200	2000	2500	×	×	×	×	×

B. Formant Frequencies of Unaccented and Accented Vowels in Isolated Words.

Speaker D.

	2a	1a	A	a1	a2
1. 'etat			e	a	
			550	500	
			2000	1600	
			2600	2200	
2. t'opat			o	a	
			400	250	
			900	1250	
			2200	2200	
3. k'opat			o	a	
			350	400	
			800	1100	
			2200	—	
4. k'ol,im			o	i	
			450	450	
			950	1700	
			2400	2300	
5. k'ol,ut			o	u	
			450	200	
			875	400	
			2350	2000	
6. g'or,a			o	a	
			350	350	
			700	1750	
			2100	2250	
7. s'onca			o	a	
			500	350	
			900	1450	
			2100	2250	
8. v'inut			i	u	
			250	300	
			1700	900	
			2200	2500	
9. v'inut,			i	u	
			250	300	
			1600	1100	
			2400	2400	
10. v'irit			i	i	
			250	350	
			1650	1500	
			2400	2400	
11. v'irit,			i	i	
			250	400	
			1600	1600	
			2400	2500	

Speaker D.

	2a	1a	A	a1	a2
12. v'orat			o	a	
			450	500	
			850	1250	
			2000	2000	
13. v,'er,im			e	i	
			400	450	
			2050	1750	
			2700	2600	
14. m'il,u			i	u	
			350	200	
			1600	400	
			2350	2100	
15. m,'et,it			e	i	
			400	250	
			2200	2200	
			2700	2500	
16. m'et,it,			e	i	
			350	200	
			2200	2100	
			2700	2600	
17. gd'al,am			a	a	
			700	450	
			1600	1750	
			2500	2600	
18. sp'ol,am			o	a	
			400	450	
			800	1700	
			2400	2500	
19. zv,'er,am			e	a	
			450	400	
			2200	1600	
			2650	2400	
20. zv,'er,a			e	a	
			250	600	
			2200	1700	
			2800	2250	
21. fsk'or,i			o	i	
			475	400	
			850	1900	
			2400	2400	
22. 'et,im,i			e	i	i
			350	250	200
			2200	2100	2300

Speaker D.

	2a	1a	A	a1	a2
23. t'ot,am,i			o	a	i
			500	300	300
			1200	1900	2100
			2200	2500	2900

Speaker D.

	2a	1a	A	a1	a2
24. v'iriti			i	i	i
			300	200	200
			1500	1500	1900
			2200	2200	2200

	2a	1a	A	a1	a2
25. p,ič'atat,		i	a	a	
		350	750	400	
		2000	1500	1700	
			2400	2400	
26. akal'otak	a	a	o	a	
	625	500	450	450	
	1200	1400	750	1500	
	2300	2000	—	2400	
27. apas,'en,a	a	a	e	a	
	650	500	350	400	
	1250	1100	2100	1900	
	2400	2300	2800	2450	

	2a	1a	A	a1	a2
28. pupav,'ina	u	a	i	a	
	250	550	250	600	
	650	1100	2150	1650	
	2300	2250	3000	2350	
29. kat,ir,'ina	a	i	i	a	
	400	300	250	600	
	1800	2150	2300	1500	
	2400	2600	3000	2400	
30. stat,ist,''	a	i	i	i	i/a(?)
čiskaj	350	250	275	400	300
	1800	2175	2300	1800	2100
	2550	2700	2900	2400	2900

	2a	1a	A	a1	a2
31. patam'u	a	a	u		
	425	700	350		
	1125	1500	550		
	2425	2350	2400		
32. puzan'i	u	a	i		
	200	500	300		
	600	1400	1950		
	2400	2400	2350		
33. tapar'i	a	a	i		
	350	750	350		
	1600	1400	1800		
	2450	2500	2150		
34. kakav'a	a	a	a		
	400	650	650		
	1500	1500	1350		
	2050	2250	2350		
35. tatarv'a	a	a	a		
	450	500	700		
	1625	1500	1100		
	2300	2500	2250		
36. t'ir,im'a	i	i	a		
	250	250	600		
	2000	2000	1300		
	2600	2600	2200		
37. t,ut,un'a	u	u	a		
	200	300	600		
	600	600	1300		
	2400	2500	2400		

	2a	1a	A	a1	a2
38. t,išin'a	i	i	a		
	200	200	600		
	2100	1600	1400		
	2500	2400	2300		
39. palav'oj,	a	a	o		
	—	—	—		
	1000	1150	650		
	2250	2525	2350		
40. tilav'oj	i	a	o		
	300	550	350		
	1600	1200	650		
	2250	2600	2300		
41. pil,iv'oj	i	i	o	j	
	250	250	350	300	
	1200	1950	850	1800	
	2300	2450	2250	2000	
42. pal,iv'oj	a	i	a	j	
	500	300	350	350	
	1000	1900	850	1750	
	2250	2475	2350	1950	
43. atam'an	a	a	a		
	750	700	700		
	1300	1450	1350		
	2300	2300	2300		
44. papad'at,	a	a	a		
	550	650	650		
	1100	1300	1600		
	2350	2375	2450		

Speaker D.

	2a	1a	A	a1	a2
45. tupav'at	u	a	a		
	250	600	650		
	850	1150	1400		
	2250	2275	2350		
46. pat,ir'at,	a	i	a		
	450	250	650		
	1250	2000	1500		
	2300	2400	2200		

Speaker D.

	2a	1a	A	a1	a2
47. pap,ir'at,	a	i	a		
	650	500	800		
	1150	2200	1600		
	2400	2600	2400		
48. pul,im,'ot	u	i	o		
	250	250	450		
	800	2000	700		
	2300	2650	2250		

Speaker J.

1. 'etat	e	a
	600	400
	1900	1600
	2900	2350
2. t'opat	o	a
	700	500
	1100	1500
	2000	2400
3. k'opat	o	a
	600	500
	1200	1500
		2250
4. k'ol,im	o	i
	600	300
	1200	2200
	2200	2700
5. k'ol,ut	o	u
	450	250
	1100	
	—	
6. g'or,a	o	a
	600	400
	1200	1750
	—	
7. s'onca	o	a
	750	400
	1200	1400
	2400	—
8. v'inut	i	ú
	300	300
	1750	1100
	2700	2400
9. v'inut,	i	u
	300	300
	1800	1000
	2300	2400

Speaker J.

10. v'irit	i	i
	300	400
	1750	1600
	2400	2250
11. v'irit,	i	i
	300	400
	1800	1750
	2400	2200
12. v'orat	o	a
	750	500
	1250	1250
	2250	1900
13. v,'er,im	e	i
	500	400
	2000	2000
	2800	2750
14. m'il,u	i	u
	300	300
	1600	1000
	2300	2250
15. m,'et,it	e	i
	400	300
	2200	2000
	2700	2600
16. m',et,it,	e	i
	400	250
	2250	2100
	2750	3000
17. gd'al,am	a	a
	800	400
	1600	1750
	2400	2500
18. sp'ol,am	o	a
	400	400
	900	1750
	2400	2400

Speaker J.

	2a	1a	A	a1	a2
19. zv,'er,am			e	a	
			600	500	
			1800	1750	
			2600	2250	
20. zv,'er,a			e	a	
			600	400	
			1800	1750	
			2250	—	
21. fsk'or,i			o	i	
			750	450	
			1250	2000	
				2500	
22. 'et,im,i			e	i	i
			400	300	250
			2000	2350	—
			2750	2700	2500
23. t,'ot,am,i			o	a	i
			600	400	250
			1300	1750	1400
			2300	2550	1900
24. v'iriti			i	i	i
			300	400	400
			1700	1700	1900
			2600	2250	2200
25. p,ič'atat,		i	a	a	
		300	800	400	
		2200	1400	1800	
		2900	2250	2600	
26. akal'otak	a	a	o	a	
	800	675	450	500	
	1550	1300	950	1400	
	2300	—	1900	2400	
27. apas'en,a	a	a	e	a	
	850	750	350	200	
	1425	1300	2200	2000	
	—	2200	2900		
28. pupav,'ina	u	a	i	a	
	350	750	350	400	
	750	1550	2400	825	
	2375	1900	3000	1575	
29. kat,ir,'ina	a	i	i	a	
	200	400	350	400	
	2200	2350	2400	1450	
		—	2900	2250	
30. stat,ist,'' čiskaj	a	i	i		a
	350	400	250		450
	2000	2300	2300		1800
	2800	3000	2100	—	2100
31. patam'u	a	a	u		
	150	700	250		
	1400	1450	650		
	—	2300	2450		
32. puzan'i	u	a	i		
	300	700	300		
	1000	1500	2250		
	2600		2900		
33. tapar'i	a	a	i		
	500	700	350		
	1500	1100	2200		
	2500	2400	2850		
34. kakav'a	a	a	a		
	500	600	650		
	1600	1250	1400		
	2350	1625	2000		
35. tatarv'a	a	a	a		
	400	600	750		
	1450	1350	1350		
	2550	—	2150		
36. t,ir,im'a	i	i	a		
	300	300	800		
	2250	2250	1500		
		—			
37. t,ut,un'a	u	u	a		
	250	300	800		
	400	700	1600		
	2400	2400	2250		
38. t,išin'a	i	i	a		
	300	300	800		
	2250	1800	1400		
	2750	2500	2700		
39. palav'oj	a	a	o	j	
	500	600	400	300	
	1000	1250	900	2000	
	2900	—	2200	2200	
40. tilav'oj	i	a	o	j	
	300	750	400	300	
	1800	1250	750	2000	
	2400	—	—	2200	

Speaker J.

	2a	1a	A	a1	a2
41. pil,iv'oj	i	i	o	j	
	300	300	500	300	
	1500	2400	1100	2000	
	2250	2900	2000	—	
42. pal,iv'oj	a	i	o	j	
	400	300	400	300	
	1500	2200	1100	2000	
	2250	3000	2000	—	
43. atam'an	a	a	a		
	800	750	800		
	1550	1350	1400		
	2300	2300	2700		
44. papad'at,	a	a	a		
	400	600	600		
	1100	1200	1700		
		2200	2250		

Speaker J.

	2a	1a	A	a1	a2
45. tupav'at	u	a	a		
	350	650	650		
	850	1350	1350		
	2700	2150	2150		
46. pat,ir'at,	a	i	a		
	400	250	700		
	1375	2375	1650		
	2400	3150	2150		
47. pap,ir'at,	a	i	a		
	450	250	650		
	1400	2300	1500		
	2400	2950	2200		
48. pul,im,'ot	u	i	o		
	300	250	750		
	1000	2200	1250		
	2500	3000	2200		

Speaker K.

1. 'etat	e	a
	750	600
	2000	1700
	3100	3200
2. t'opat	o	a
	600	600
	1200	1300
	3250	3200
3. k'opat	o	a
	600	600
	1100	1500
	3200	2700
4. k'ol,im	o	i
	600	400
	1250	2200
	2900	3400
5. k'ol,ut	o	u
	600	400
	1200	1200
	2800	3000
6. g'or,a	o	à
	600	500
	1250	2200
	2900	2900
7. s'onca	o	a
	750	450
	1250	1700
	2500	3100

Speaker K.

8. v'inut	i	u
	400	600
	2000	1000
	2800	2500
9. v'inut,	i	u
	600	600
	2300	1200
	2800	3600
10. v'irit	i	i
	500	750
	2500	2250
	3200	3250
11. v'irit,	i	i
	400	500
	2000	1800
	2900	3200
12. v'orat	o	a
	700	700
	1250	1500
	3200	2750
13. v,'er,im	e	i
	400	500
	2500	2600
	3200	3200
14. m'il,u	i	u
	400	600
	2250	900
	2750	2500

Speaker K.

	2a	1a	A	a1	a2
15. m,'et,it			e	i	
			500	400	
			2600	2250	
			3500	3350	
16. m,'et,it,			e	i	
			500	600	
			2500	2500	
			3600	3500	
17. gd'al,am			a	a	
			1000	500	
			1600	2600	
			2750	3200	
18. sp'ol,am			o	a	
			600	500	
			1200	2200	
			3000	3100	
19. zv,'er,am			e	a	
			400	600	
			2400	2250	
			3200	3200	

Speaker K.

	2a	1a	A	a1	a2
20. zv,'er,a			e	a	
			500	500	
			2500	2000	
			3200	2900	
21. fsk'or,i			o	i	
			700	500	
			1250	2400	
			2800	3200	
22. 'et,im,i			e	i	i
			500	350	400
			2300	2500	2750
			2750	3400	3400
23. t'ot,am,i			o	a	i
			800	600	250
			1250	2250	2300
			2900	3000	3400
24. v'iriti			i	i	i
			400	350	600
			2100	1750	2300
			3000	3100	3100

	2a	1a	A	a1	a2
25. p,ič'atat,		i	a	a	
		400	900	500	
		2500	1600	1800	
		3200	2900	3400	
26. akal'otak	a	a	o	a	
	1000	800	500	600	
	1750	1600	1000	1750	
	2600	2600	—	2700	
27. apas,'en,ja	a	a	e	a	
	1200	1000	500	500	
	1700	1750	2600	2700	
	3100	2750	3400	3600	
28. pupav,'ina	u	a	i	a	
	400	800	400	500	
	1050	1600	2700	1200	
	3200	2900	3500	1750	
29. kat,ir,'ina	a	i	i	a	
	400	300	400	500	
	2250	2700	2700	1000	
	3000	3400	3400	3250	
30. stat,ist,'i-	a	i	i	i	a
čiskaj	400	850	300	500	650
	2000	2500	2600	2200	2750
	3100	3200	3400	3200	3200

	2a	1a	A	a1	a2
31. patam'u	a	a	u		
	500	800	250		
	1700	1600	600		
	3200	3200	3400		
32. puzan'i	u	a	i		
	400	900	400		
	1100	1600	2800		
	3200	3200	3200		
33. tapar'i	a	a	i		
	500	1000	600		
	1750	1700	2600		
	3200	3000	3200		
34. kakav'a	a	a	a		
	600	800	900		
	1600	1600	1600		
	2650	2400	2600		
35. tatarv'a	a	a	a		
	500	850	800		
	1600	1650	1500		
	3200	3200	2700		
36. t,ir,im'a	i	i	a		
	500	500	1000		
	2500	2600	1500		
	3250	3250	3000		

Speaker K.

	2a	1a	A	a1	a2
37. t,ut,un'a	u	u	a		
	300	350	1100		
	600	700	1650		
	2900	3250	3000		
38. t,išin'a	i	i	a		
	500	450	1100		
	2400	2100	1600		
	3200	2800	3000		
39. palav'oj	a	a	o	j	
	500	750	400	400	
	1250	1300	1000	2500	
	3200	3250	3200	3000	
40. tilav'oj	i	a	o	j	
	400	900	500	300	
	1000	1400	1100	2600	
	3050	3100	2800	3100	
41. pil,iv'oj	i	i	o	j	
	300	300	400	300	
	2500	2500	1000	2500	
	2700	3200	2900	3200	
42. pal,iv'oj	a	i	o	j	
	500	300	400	300	
	1600	2400	1000	2500	
	2650	3200	3000	2500	

Speaker K.

	2a	1a	A	a1	a2
43. atam'an	a	a	a		
	1000	900	1100		
	1750	1600	1700		
	2200	2200	2000		
44. papad'at,	a	a	a		
	500	900	900		
	1250	1600	1750		
	2900	2100	3200		
45. tupav'at	u	a	a		
	400	900	800		
	600	1300	1500		
		2800	2900		
46. pat,ir'at,	a	i	a		
	500	500	900		
	1750	2600	1750		
	3000	3400	3000		
47. pap,ir'at,	a	i	a		
	500	500	900		
	1600	2700	1750		
	2900	3400	2700		
48. pul,im,'ot	u	i	o		
	400	300	900		
	1000	2500	1300		
	2800	3200	3200		

Appendix II.

Energy Density Spectra of Stop Bursts (Figs. 1-6) *and Continuant Consonants* (Figs. 7-13).

In each figure, samples spoken by subject K are in the upper row, those spoken by the subject J are in the lower row. Syllables from which the samples were taken are indicated on the graphs.

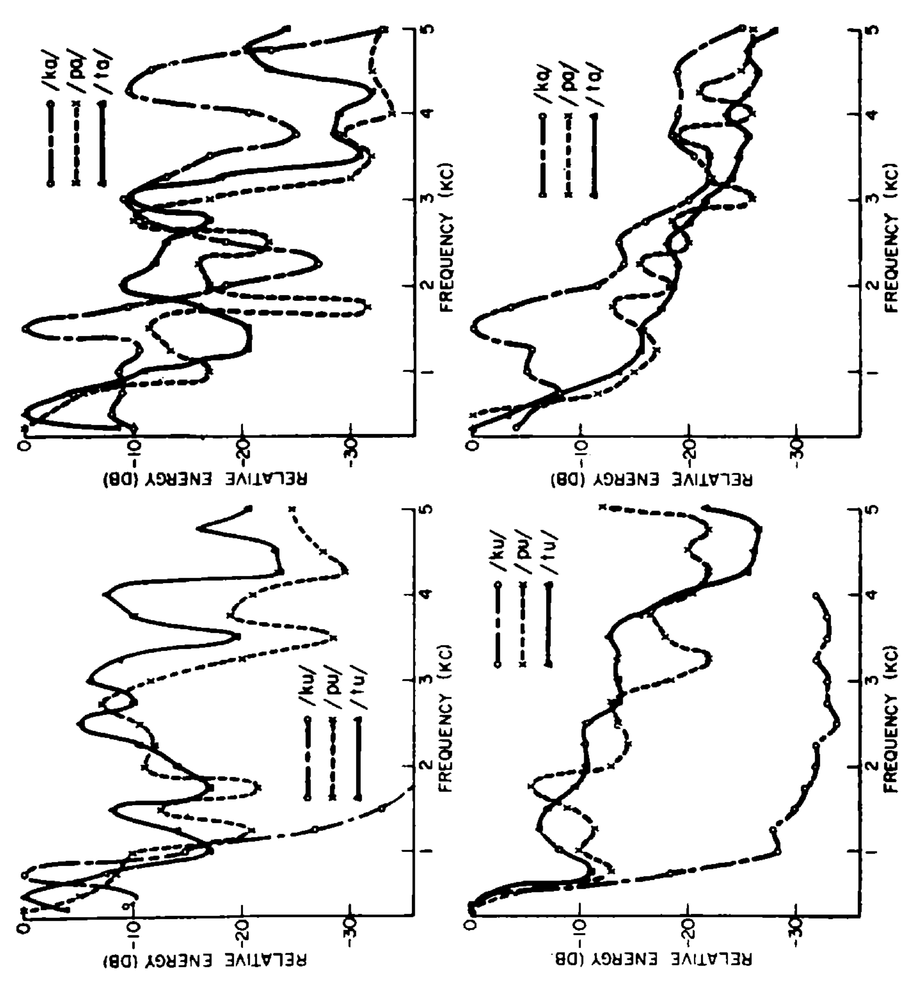


Fig.

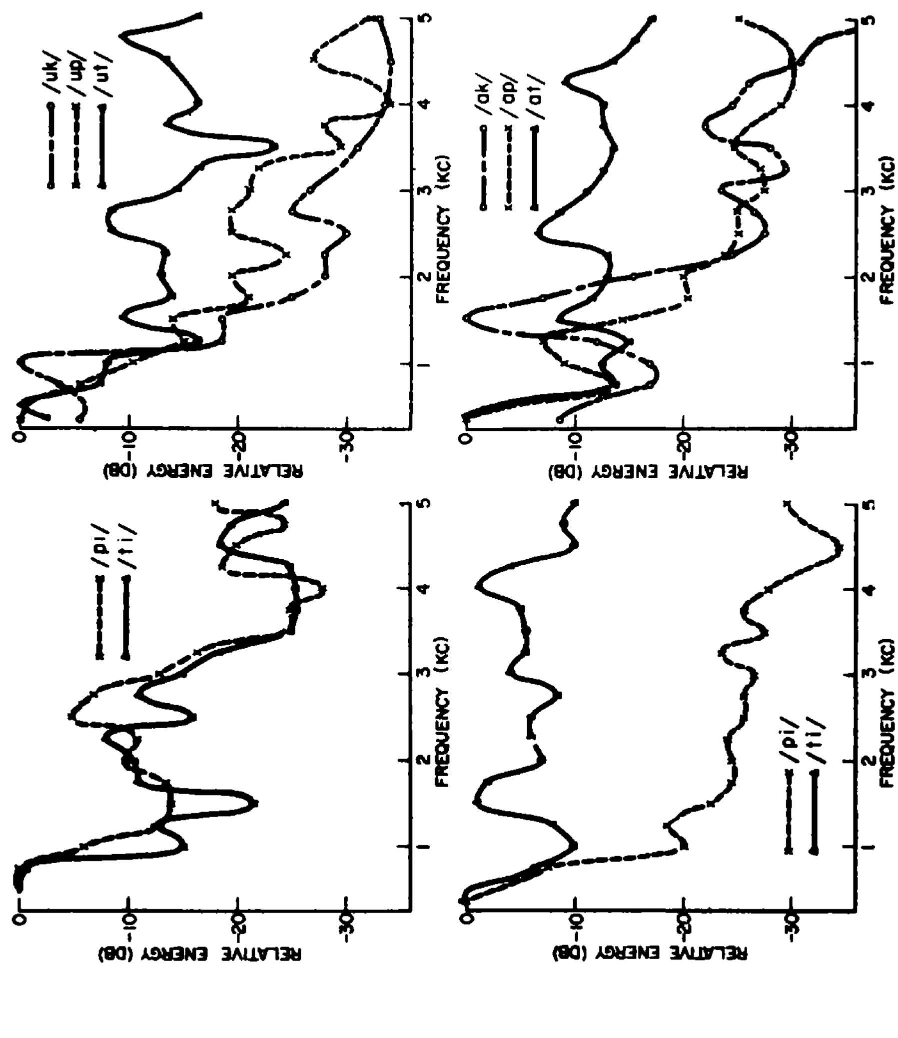


Fig. 2

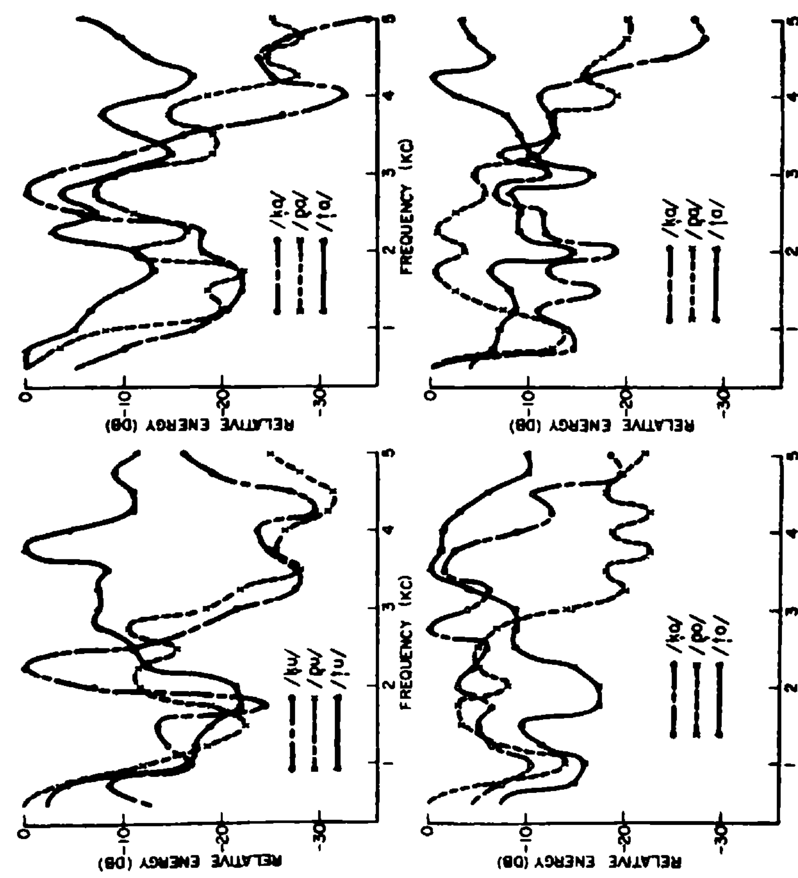

RELATIVE ENERGY (DB)
FREQUENCY (KC)
/ku/
/pu/
/tu/
/ka/
/pa/
/ta/
0
-10
-20
-30
1
2
3
4
5

Fig. 3

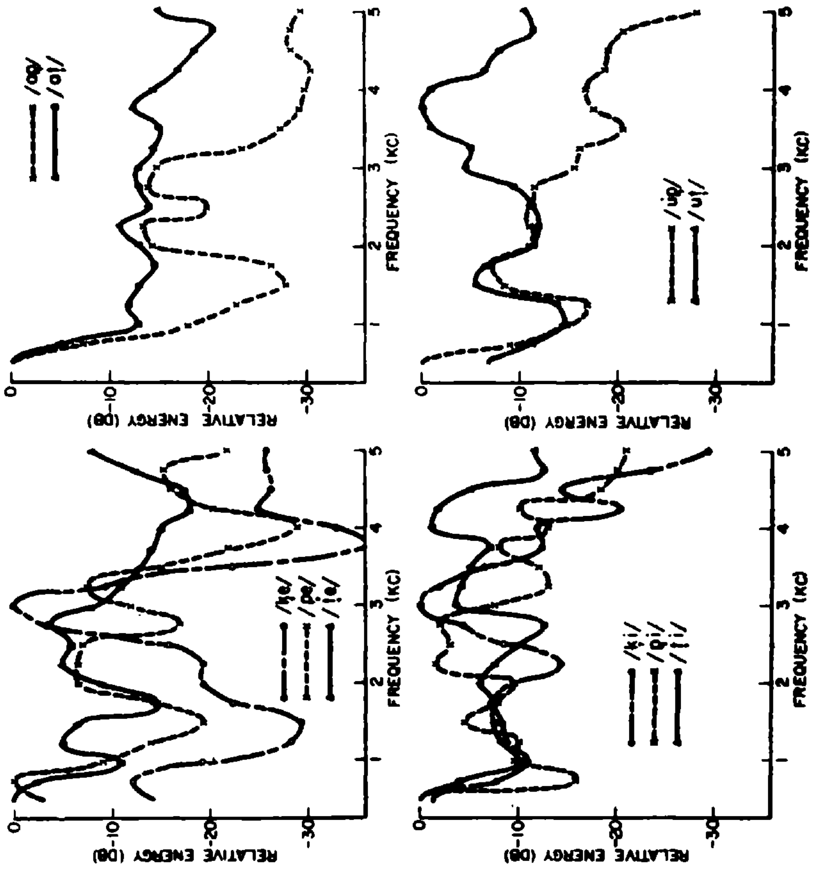


Fig. 4

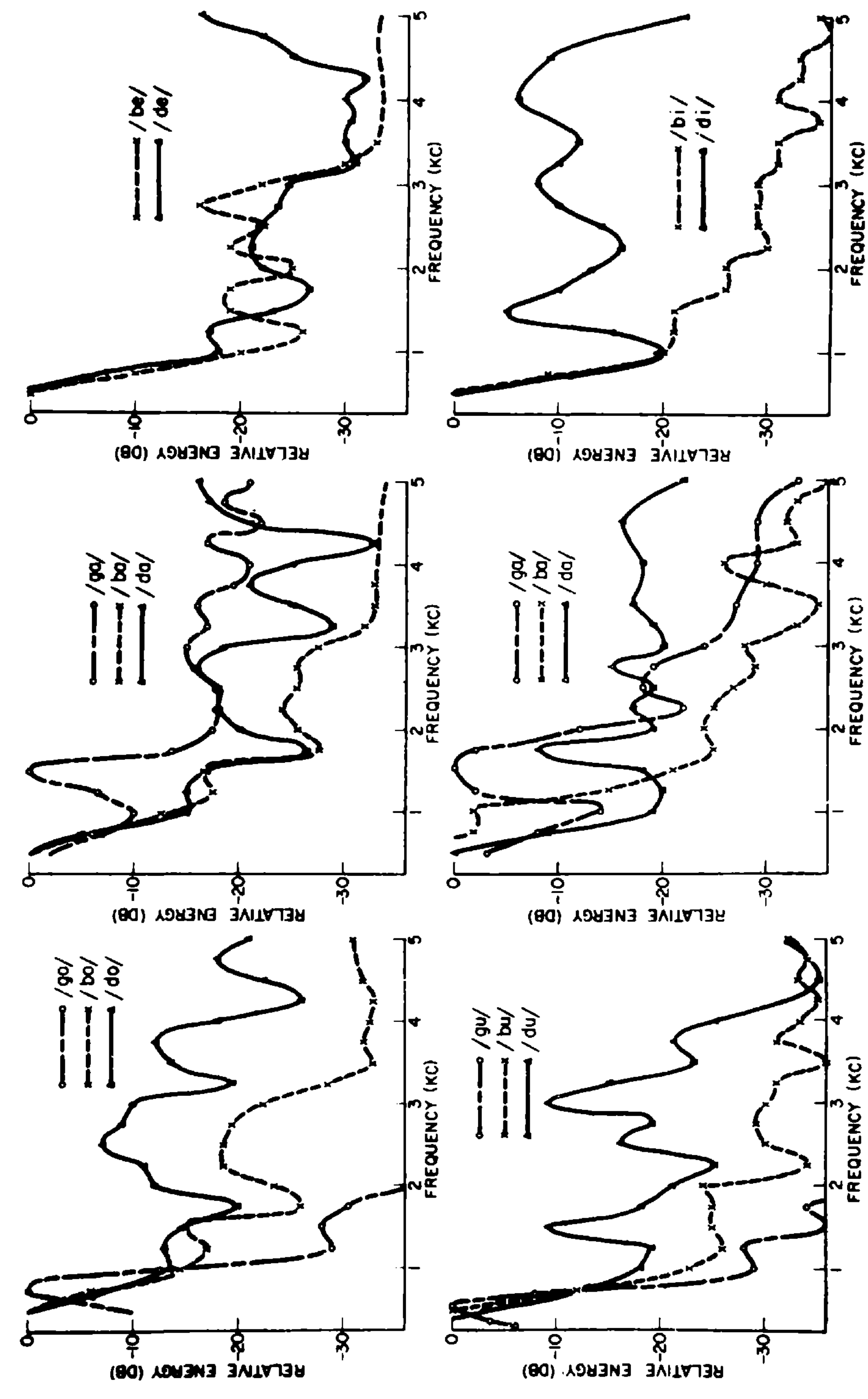

/go/
/bo/
/do/
/ga/
/ba/
/da/
/be/
/de/
/gu/
/bu/
/du/
/bi/
/di/
RELATIVE ENERGY (DB)
FREQUENCY (KC)

Fig. 5

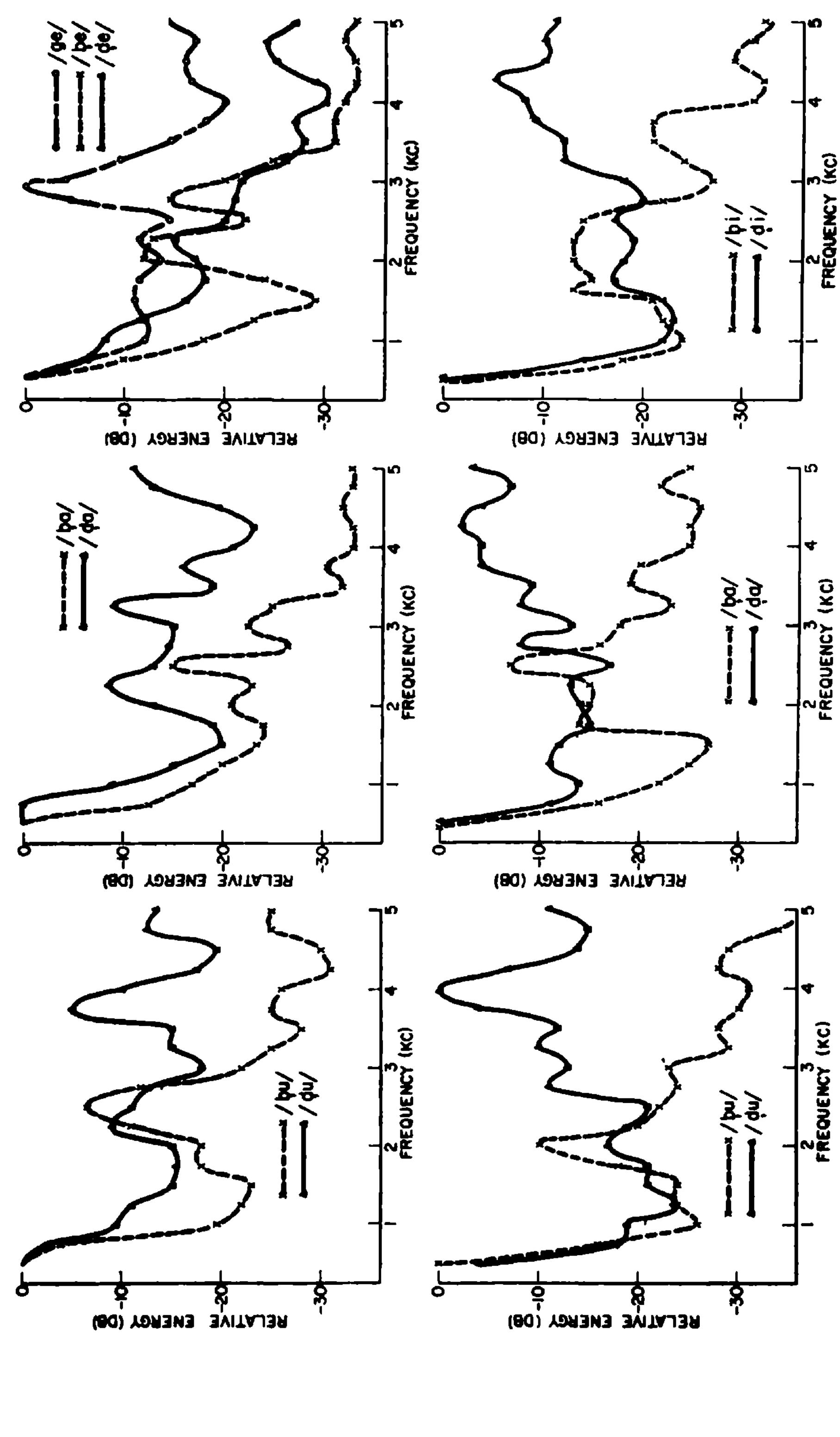


Fig. 6

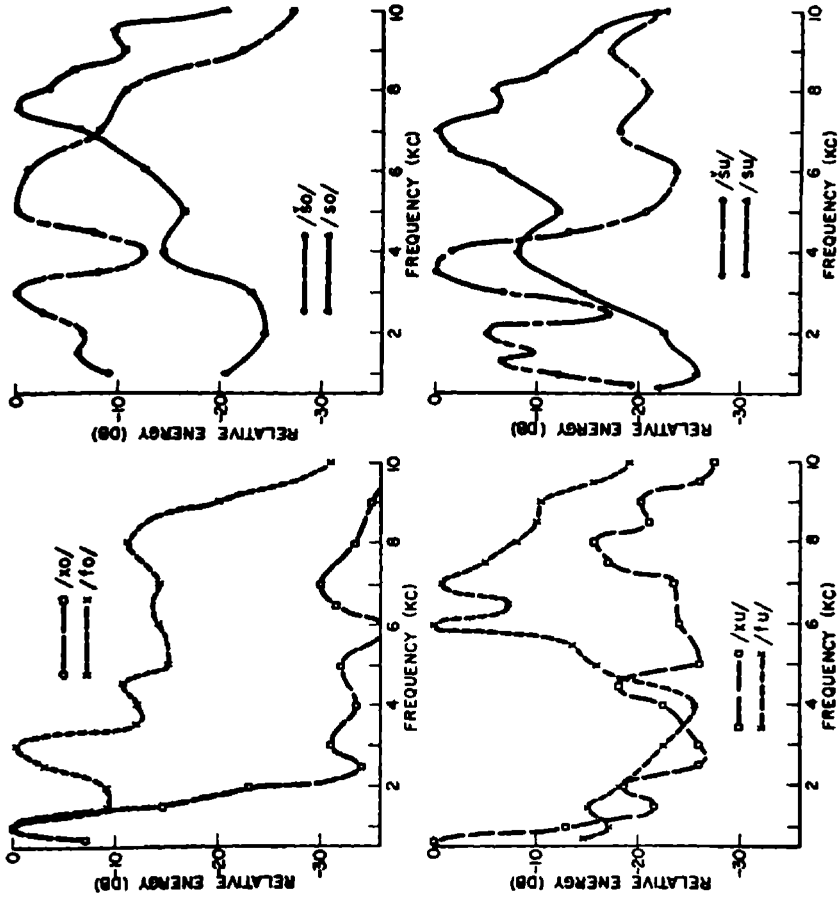


Fig. 7

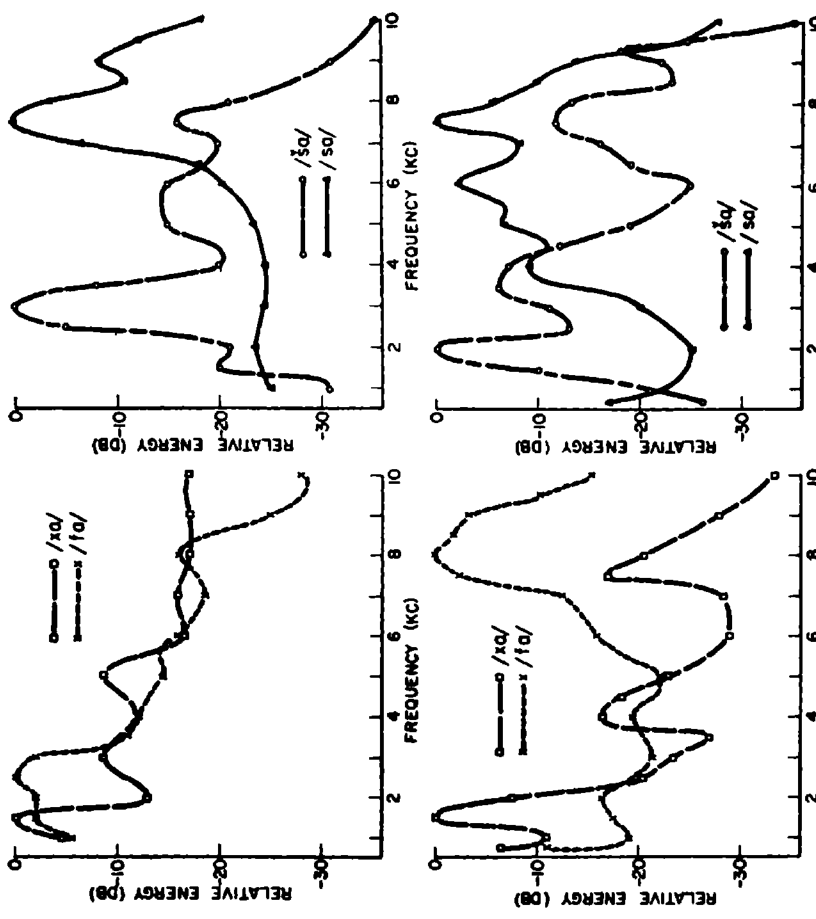


Fig. 8

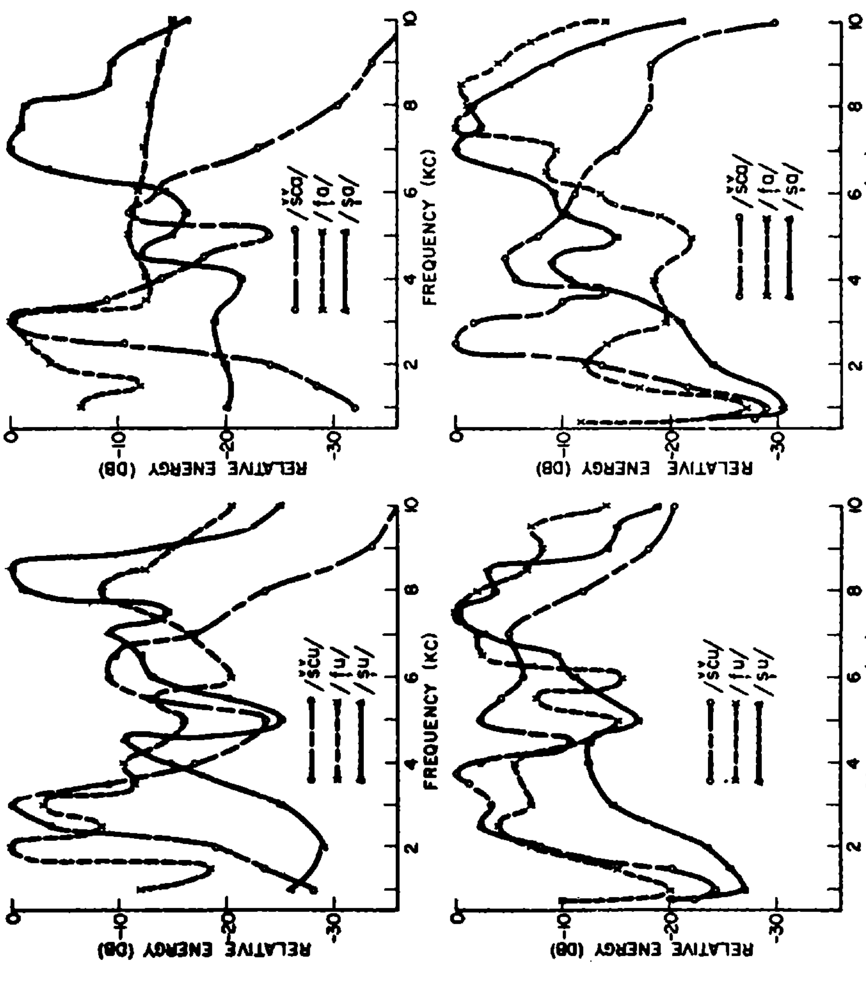


Fig. 9

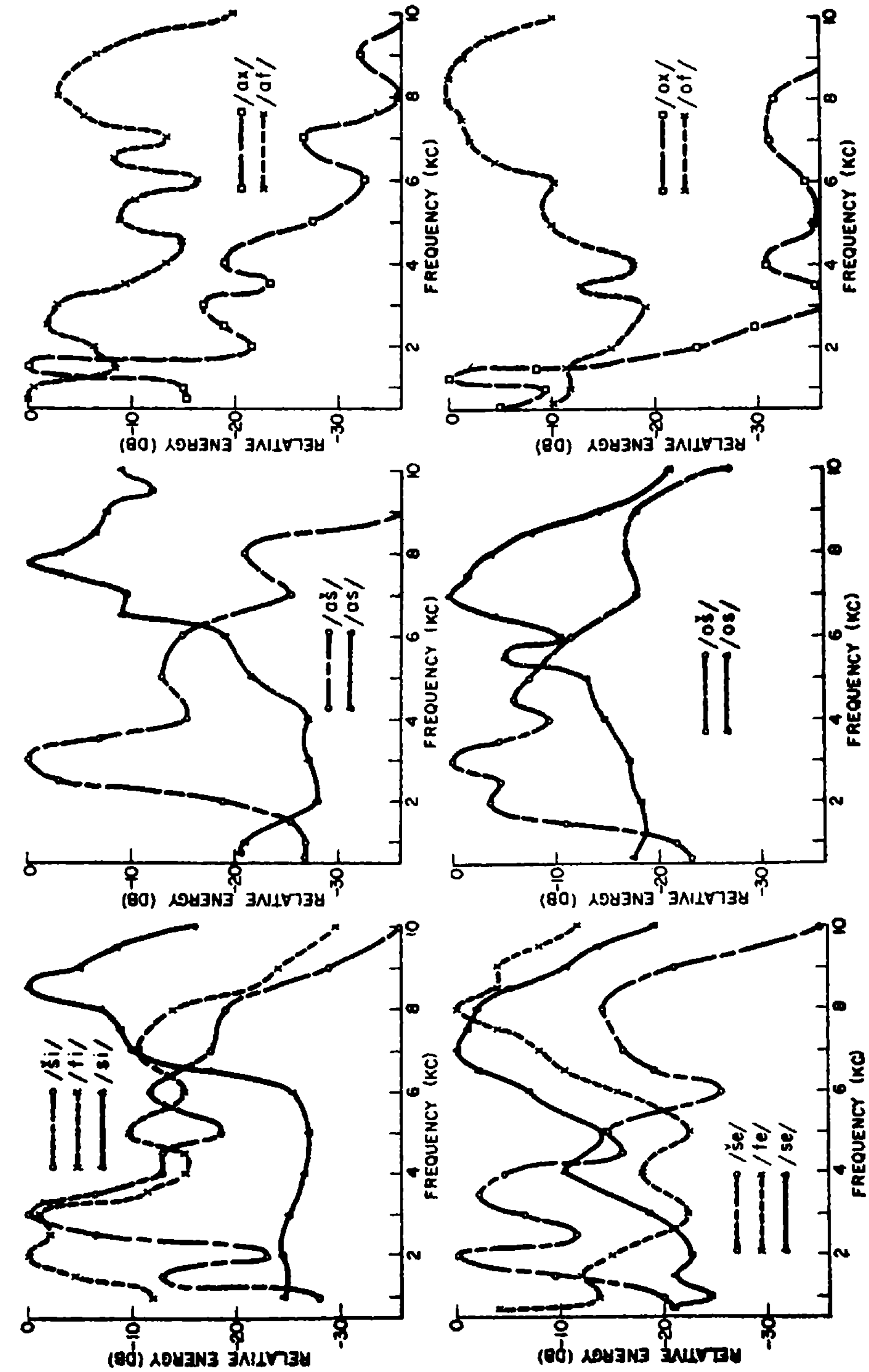

RELATIVE ENERGY (DB)
FREQUENCY (KC)
/ax/
/af/
/aš/
/as/
/ox/
/of/
/oš/
/os/
/ši/
/ti/
/si/
/še/
/te/
/se/

Fig. 10

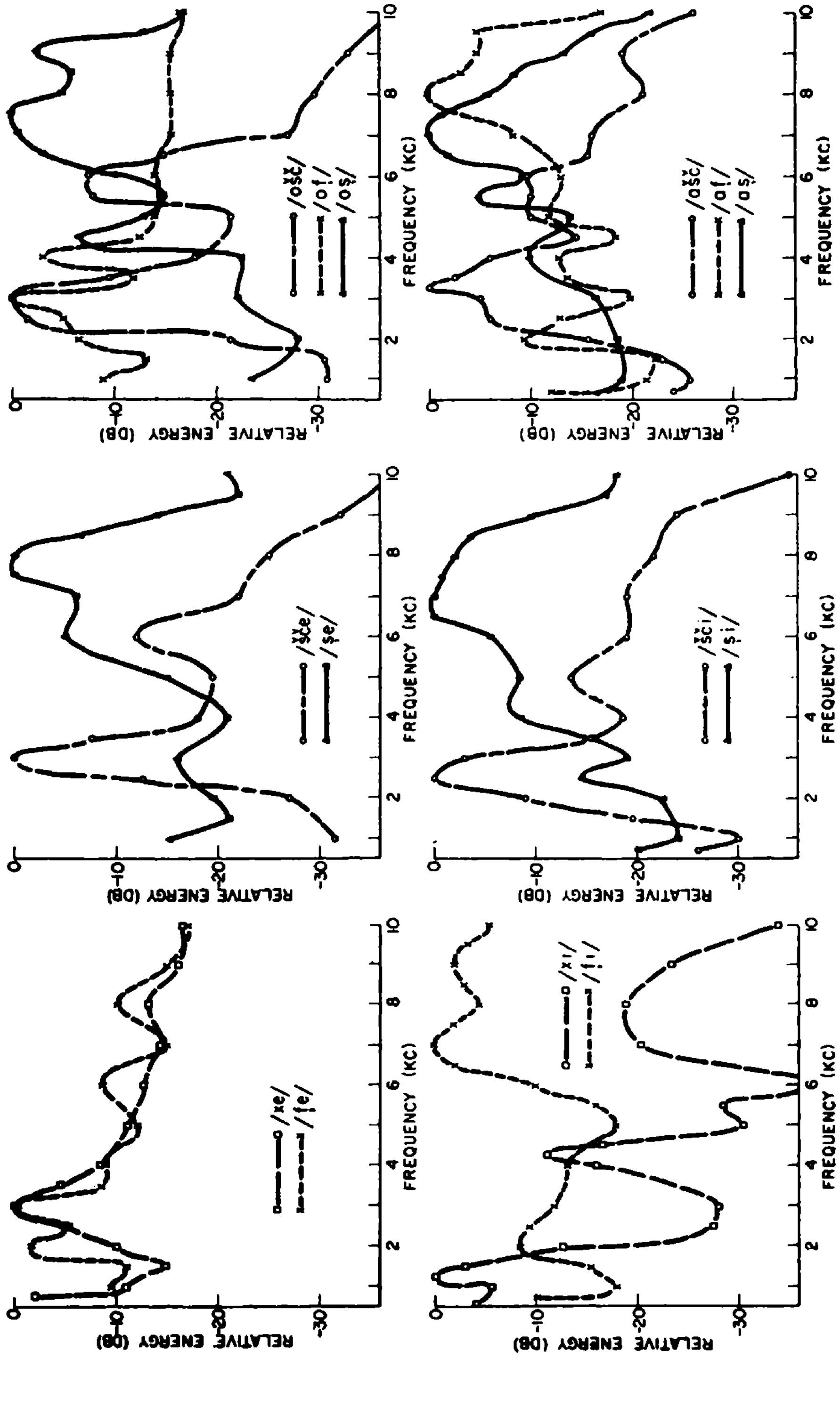


Fig. 11

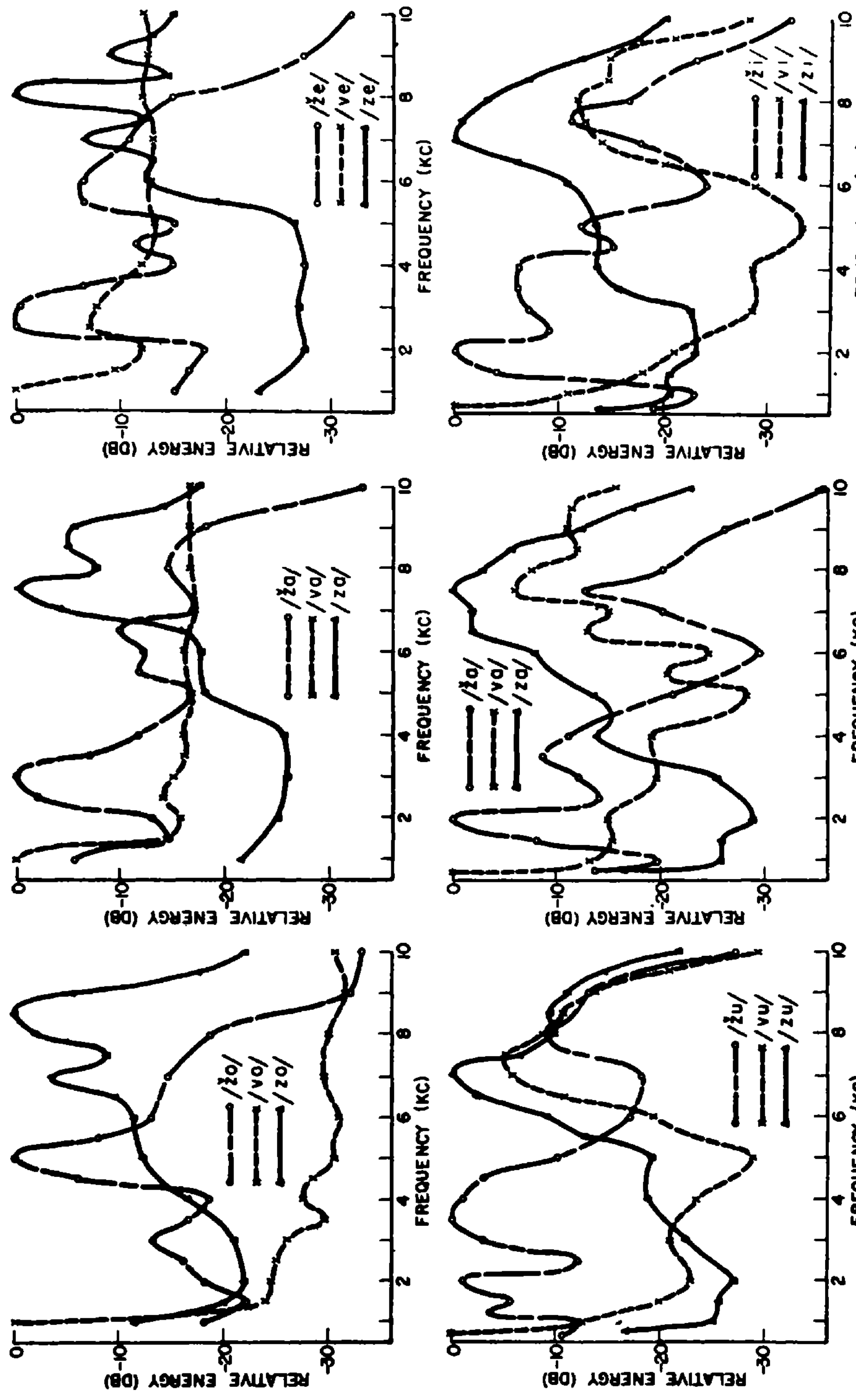

/žo/
/vo/
/zo/
/ža/
/va/
/za/
/že/
/ve/
/ze/
/žu/
/vu/
/zu/
/ža/
/va/
/za/
/ži/
/vi/
/zi/
FREQUENCY (KC)
RELATIVE ENERGY (DB)

Fig. 12

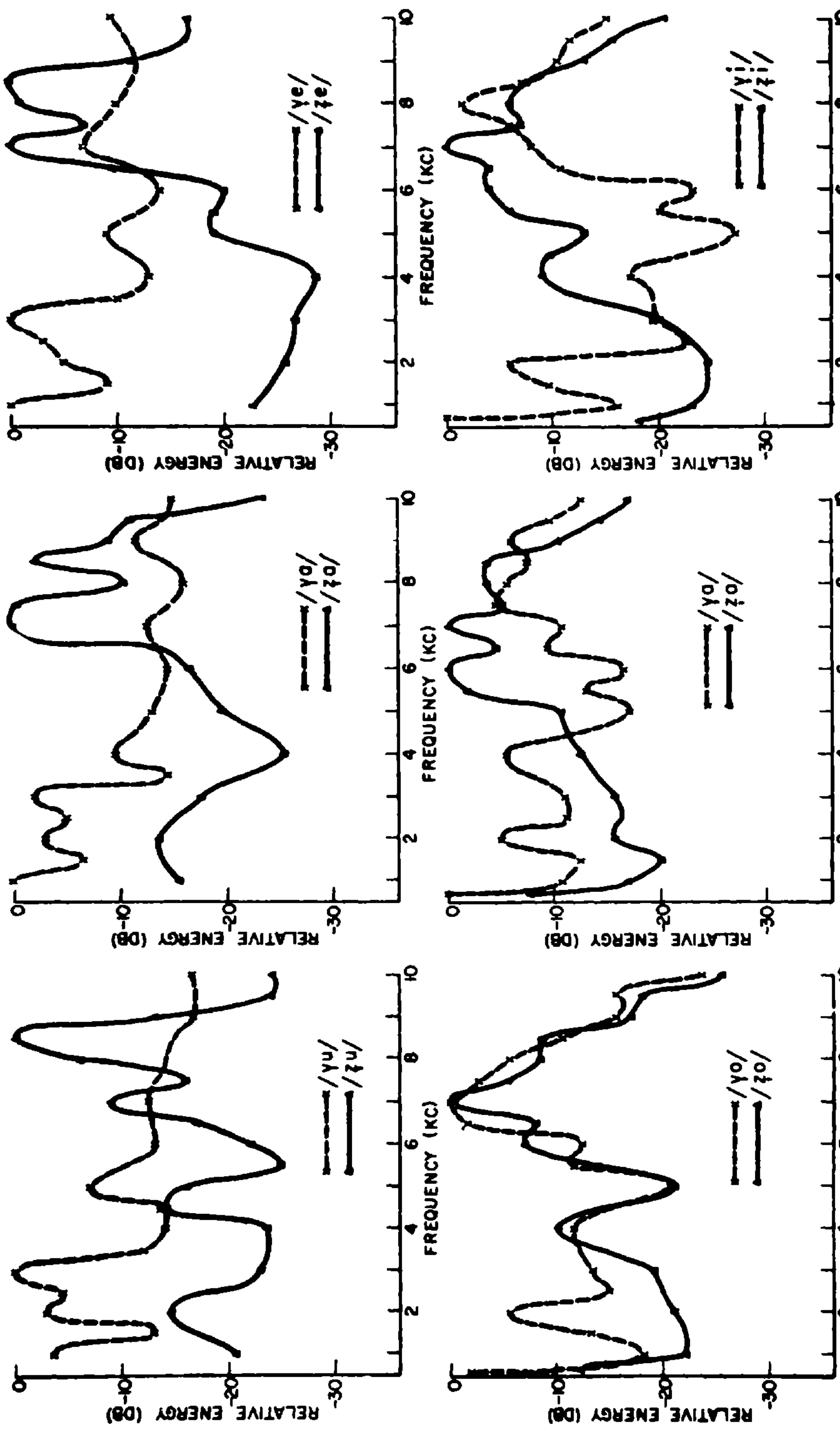


Fig. 13

BIBLIOGRAPHY

Allen, W. S., *Phonetics in Ancient India* (London, 1953).

Avanesov, R. I. "Voprosy fonetičeskoj sistemy russkix govorov literaturnogo jazyka," *Izvestija AN SSSR, Otd. lit. i jaz., 6*, 211–228 (1947).

——, "O dolgix šipjaščix v russkom jazyke," *Doklady i soobščenija filologičeskogo fakul'teta Moskovskogo gosudarstvennogo universiteta*, No. 6 (1948), 23–30.

——, *Russkoe literaturnoe proiznošenie*, (Moscow, 1950).

——, "O kačestve zadnenebnoj frikativnoj soglasnoj pered glasnymi perednego rjada v russkom jazyke," *Doklady i soobščenija instituta jazykoznanija AN SSSR*, 1952, No. 2, 34–41.

——, "O slogorazdele i stroenii sloga v russkom jazyke," *Voprosy jazykoznanija*, 1954, No. 6, 88–101.

——, *Udarenie v sovremennom russkom literaturnom jazyke* (Moscow, 1955).

——, "O trex tipax naučno-lingvističeskix transkripcij," *Slavia, 25*, 347–371 (1956).

——, *Fonetika sovremennogo russkogo literaturnogo jazyka* (Moscow, 1956).

Avanesov, R. I. and S. I. Ožegov, *Russkoe literaturnoe udarenie i proiznošenie: Opyt slovarja-spravočnika* (Moscow, 1955).

Avanesov, R. I. and V. N. Sidorov, *Očerk grammatiki russkogo literaturnogo jazyka*, I (Moscow, 1945).

Bargielowna, M., "Grupy fonemów społgłoskowych współczesnej polszczyzny kulturalnej," *Bulletin de la Société polonaise de linguistique, 10*, 1–25 (1950).

Backhaus, H., "Über die Bedeutung der Ausgleichsvorgänge in der Akustik," *Zeitschrift der technischen Physik, 13*, 31–47 (1932).

Baudouin de Courtenay, J. A., *Otryvok iz lekcij po fonetike i morfologii russkogo jazyka* (Voronež, 1882).

——, *Ob otnošenii russkogo pis'ma k russkomu jazyku* (St. Petersburg, 1912).

Bernštejn, S., "Protiv idealizma v fonetike," *Izvestija Akademii Nauk SSSR, Otd. lit. i jaz., 11*, 6, 541–559 (1952).

Bloch, B., "A Set of Postulates for Phonemic Analysis," *Language, 24*, 3–46 (1948).

Bloomfield, L., *Language* (New York, 1933).

Bogorodickij, V. A., *Fonetika russkogo jazyka* (Kazan', 1930).

——, *Očerki po jazykoznaniju i russkomu jazyku* (Moscow, 1939).

Bolinger, D. L. and Gerstman, L. J., "Disjuncture as a Cue to Constructs," *Word, 13*, 246–256 (1957).

Boyanus, S. C., *A Manual of Russian Pronunciation* (London, 1946).

Broch, O., "Ob isčeznovenii mežduglasnogo i, j," *Sbornik statej v čest' F. F. Fortunatova* (Warsaw, 1902), pp. 134–140.

——, *Slavische Phonetik* (Heidelberg, 1911).

Bulaxovskij, L. A., "Deètimologizacija v russkom jazyke," *Trudy Instituta russkogo jazyka, 1*, 147–209 (1949).

Bykov, Ju. S., "Sovetskie raboty po teorii razborčivosti reči," *Izvestija Akademii Nauk SSSR, serija fizičeskaja, 13*, 6, 728–739 (1949).

——, "Spektral'nye xarakteristiki russkoj reči," *Trudy komissii po akustike, 7*, 36–39 (1953).

——, "Formantnye xarakteristiki russkoj reči," *Trudy komissii po akustike, 7*, 40–42 (1953).

——, "Issledovanie pikovyx urovnej russkoj reči," *Trudy komissii po akustike, 7*, 43–47 (1953).

Carmody, F. J., *X-Ray Studies of Speech Articulation = University of California Publications in Modern Philology, 20*, 4, i-viii and 187–238 (1937).

——, *An X-Ray Study of Pharyngeal Articulation = University of California Publications in Modern Philology, 21*, 5, 377–384 (1941).

Chao, Y. R., "The Non-Uniqueness of Phonemic Solutions of Phonetic Systems," *Bulletin of the Institute of History and Philology, Academia Sinica, 4*, part 4 (1934).

——, Review of *Preliminaries to Speech Analysis, Romance Philology, 8*, 40–46 (1954).

Cherry, E. C., Halle, M., and Jakobson, R., "Toward the Logical Description of Languages in their Phonemic Aspect," *Language, 29*, 24–46 (1953).

Chiba, T., and Kajiyama, M., *The Vowel, its Nature and Structure* (Tokyo, 1941; reprint 1959).

Chomsky, N., Review of R. Jakobson and M. Halle, *Fundamentals of Language, I.J.A.L., 23*, 234–242 (1957).

——, Review of C. F. Hockett, *Manual of Phonology, I.J.A.L., 23*, 223–234 (1957).

——, *Syntactic Structures* ('s-Gravenhage, 1957).

Cooper, F. S., et al, "The Interconversion of Audible and Visible Patterns as a Basis for Research in the Perception of Speech," *Proceedings of the National Academy of Sciences, 37*, 318–325 (1951).

——, "Some Experiments on the Perception of Synthetic Speech Sounds," *JASA, 24*, 597–606 (1952).

——, "The Role of Consonant-Vowel Transitions in the Perception of the Stop and Nasal Consonants," *Psychological Monographs, 68*, 8, 1–13 (1954).

David, E. E. Jr., "Signal Theory in Speech Transmission," *IRE Transactions on Circuit Theory*, CT-3, 232–244 (1956).

Delattre, P., "The Physiological Interpretation of Sound Spectrograms," *PMLA, 66*, 864–875.

Delattre, P., Liberman, A. M., Cooper, F. S., and Gerstman, L. J., "An Experimental Study of the Acoustic Determinants of Vowel Color: Observations on One- and Two-Formant Vowels Synthesized from Spectrographic Patterns," *Word, 8*, 195–210 (1952).

Delattre, P. C., Liberman, A. M., and Cooper, F. S., "Acoustic Loci and Transitional Cues for Consonants," *JASA, 27*, 769–773 (1955).

Dieth, E., *Vademekum der Phonetik* (Bern, 1950).

Dudley, H. "On the Carrier Nature of Speech," *Bell System Technical Journal, 19*, 495–515 (1940).

Dudley, H., Riesz, R. R., and Watkins, S. S. A., "A Synthetic Speaker (Voder)", *Journal of the Franklin Institute, 227*, 739–764 (1937).

Dunn, H. K., "The Calculation of Vowel Resonances and an Electrical Vocal Tract," *JASA, 22*, 740–753 (1950).

Durnovo, N. N., *Očerk istorii russkogo jazyka* (Moscow, 1924; reprint 's-Gravenhage, 1959).

Durnovo, N. N. and Ušakov, D. N., "Opyt fonetičeskoj transkripcii russkogo literaturnogo proiznošenija," *Slavia, 5*, 342–347 (1926).

Fant, C. G. M., *Analys av de svenska vokal-ljuden = L. M. Ericsson protokoll H/P-1035* (1948).

——, *Analys av de svenska konsonant-ljuden = L. M. Ericsson protokoll H/P-1064* (1949).

——, *Transmission Properties of the Vocal Tract with Application to the Acoustic Specification of Phonemes = MIT Acoustics Laboratory Technical Report* No. 12 (January 1952).

——, "On the Predictability of Formant Levels and Spectrum Envelopes from Formant Frequencies," *For Roman Jakobson* (The Hague, 1956), pp. 109–121.

——, Modern Instruments and Methods for Acoustic Studies of Speech, Report Presented to the Eighth International Congress of Linguists Oslo August 1957.

——, *Acoustical Theory of Speech Production* ('s-Gravenhage, in press).

Farnsworth, D. W., "High Speed Motion Picture of the Human Vocal Cords," *Bell Laboratories Record, 18,* 203 (1940).

Fischer-Jørgensen, E., "The Phonetic Basis for Identification of Phonemic Entities," *JASA, 24,* 611–617 (1952).

——, "Acoustic Analysis of Stop Consonants," *Miscellanea Phonetica, 2,* 42–59 (1954).

Flanagan, J. L., "A Difference Limen for Vowel Formant Frequency," *JASA, 27,* 613–617 (1957).

——, "Difference Limen for Formant Amplitude," *Journal of Speech and Hearing Disorders, 22,* 206–212 (1957).

Fletcher, H., *Speech and Hearing* (New York, 1929).

——, *Speech and Hearing in Communication* (New York, 1953).

Fröschels, E., Hajek, L., and Weiss, D., "Untersuchungsmethoden der Stimme und Sprache," in *E. Abderhaldens Handbuch der biologischen Arbeitsmethoden,* V, 7-2 (Leipzig, 1937), 1383–1540.

Gemelli, A., and Pastori, G., *L'analisi elettro-acustica del linguaggio*: (Milan, 1934) - Pubblicazione; Università del S. Cuore Series VI (biological sciences) Vol. II.

Gierke, H. von, "Über die mit dem Mund hervorgebrachten Pfeiftöne," *Pflügers Archiv, 249,* 307–312 (1947).

Gleason, H. A., *An Introduction to Descriptive Linguistics* (New York, 1955).

——, *Workbook in Descriptive Linguistics* (New York, 1955).

Grassmann, H., "Über die physikalische Natur der Sprachlaute," *Annalen der Physik und Chemie,* series 3, I, 623–639 (1877).

Grützner, P., "Physiologie der Stimme und Sprache," in L. Herrmann, ed., *Handbuch der Physiologie,* Volume I, part 2 (Leipzig, 1879).

Gvozdev, A., *O fonologičeskix sredstvax russkogo jazyka* (Moscow-Leningrad, 1949).

Hála, B., *Akustická podstáta samohlásek = Rozpravy České akademie věd a umění,* class III No. 78 (1940).

——, *Úvod do fonetiky* (Prague, 1948).

——, "Une contribution à l'éclaircissement de la nature phonétique des affriquées," *Zeitschrift für Phonetik, 6,* 77–93 (1952).

Hála, B. and Polland, B., *Artikuláce českých zvukü v roentgenovych obrazech* (Prague, 1926).

Hála, B. and Sovák, M., *Hlas, reč, sluch* (Prague, 1947).

Halle, M., "Why and How Do We Study the Sounds of Speech," *Report of the Fifth Annual Round Table Meeting on Linguistics and Language Teaching*: The Institute of Languages and Linguistics, School of Foreign Service, Georgetown University, Washington, 73–80 (1954).

"The Strategy of Phonemics," *Word, 10,* 197–209 (1954).

——, Review of C. F. Hockett, *Manual of Phonology, JASA, 28,* 509–511 (1956).

"In Defense of the Number Two," *Studies Presented to J. Whatmough* ('s-Gravenhage, 1957), pp. 65–72.

Halle, M., Hughes, G. W., and Radley, J.-P. A., "Acoustic Properties of Stop Consonants," *JASA, 29,* 107–116 (1957).

Harris, K. S., Hoffman, H. S., Liberman, A. M., Delattre, P. C. and Cooper, F. S., "Effect of Third Formant Transition on the Perception of Voiced Stop Consonants," *JASA, 30,* 122–126 (1958).

Harris, Z. S., *Methods of Structural Linguistics* (Chicago, 1951).

Hattori, S., Yamamoto, K., and Fujimura, O., "Nasalization of Vowels in Relation to Nasals, *JASA, 30,* 267–274 (1958).

Haugen, E., "The Syllable in Linguistic Description," *For Roman Jakobson* ('s-Gravenhage, 1956), pp. 213–221.

——, "Syllabification in Kutenai," *I.J.A.L., 22,* 196–201 (1956).

Heffner, R-M. S., *General Phonetics* (Madison, 1949).

Hegedüs, L., "Die Natur der ungarischen Affrikaten," *Archives néerlandaises de phonétique expérimentale, 15*, 97–102 (1939).

Helmholtz, H. V., *Die Lehre von den Tonempfindungen* (Braunschweig, 1913).

Hempel, Carl G., *Fundamentals of Concept Formation in Empirical Science International Encyclopedia of Unified Science*, Vol. II, Number 7 (Chicago, 1952).

Hockett, C. F., "A System of Descriptive Phonology," *Language, 18*, 1–22 (1942).

——, "Two Fundamental Problems in Phonemics," *Studies in Linguistics, 7*, 29–51 (1949).

——, "Two Models of Grammatical Description," *Word, 10*, 210–234 (1954).

——, *A Manual of Phonology* (Bloomington, 1955).

Hooykaas, W., "Luftverbrauch bei Plosiven und Frikativen," *Archives néerlandises de phonétique expérimentale, 13*, 51–68 (1936).

House, A. S., "Analog Studies of Nasal Consonants," *Journal of Speech and Hearing Disorders, 22*, 190–204 (1957).

House, A. S. and Fairbanks, G., "The Influence of Consonant Environment upon the Secondary Acoustical Characteristics of Vowels," *JASA, 25*, 105–113 (1953).

House, A. S. and Stevens, K. N., "Analog Studies of the Nasalization of Vowels," *Journal of Speech and Hearing Disorders, 21*, 218–232 (1956).

Hughes, G. W., and Halle, M., "Spectral Properties of Fricative Consonants," *JASA, 28*, 303–310 (1956).

Isačenko, A., *Fonetika spisovnej ruštiny* (Bratislava, 1947).

——, Review of W. Steinitz, *Russische Lautlehre, Zeitschrift für Phonetik, 8*, 411–6 (1955).

Jakobson, R., "Observations sur le classement phonologique des consonnes," *Proceedings of the Third International Congress of Phonetic Sciences* (Ghent, 1939), pp. 34–41.

——, *Kindersprache, Aphasie und allgemeine Lautgesetze = Språkvetenskapliga sällskapets i Uppsala förhandligar* (1940–1942).

——, Review of G. L. Trager, *Introduction to Russian, The Slavonic and East European Review, 22*, 120–133 (1944).

——, "On the Identification of Phonemic Entities," *Travaux du Cercle linguistique de Copenhague, 5*, 205–213 (1954).

——, "Die Verteilung der stimmhaften und stimmlosen Geräuschlaute im Russischen, *Festschrift für Max Vasmer* (Berlin, 1956), pp. 199–202.

——, *Shifters, Verbal Categories, and the Russian Verb* (Harvard University, 1957).

Jakobson, R., Fant, C. G. M., and Halle, M., *Preliminaries to Speech Analysis = M. I. T. Acoustics Laboratory Technical Report* No. 13 (May, 1952).

Jakobson, R., and Halle, M., *Fundamentals of Language* ('s-Gravenhage, 1956).

Jespersen, O., *Lehrbuch der Phonetik* (Leipzig, 1904).

Jones, D., *The Phoneme, Its Nature and Use* (Cambridge, 1950).

Jones, D. and Trofimov, M. V., *The Pronunciation of Russian* (Cambridge, 1923).

Jones, L. G., *Acoustic Patterns of the Russian Vowels* (Ph.D. dissertation, Harvard University., 1952).

Joos, M., *Acoustic Phonetics = Language Monograph #23* (Baltimore, 1948).

——, "Description of Language Design," *JASA, 22*, 701–708 (1950).

Joos, M., ed., *Readings in Linguistics* (Washington, 1957).

Kaiser, L., "A Few Remarks Concerning Occlusives in Russian," *Archives néerlandaises de phonétique expérimentale, 4*, 77–80 (1929).

Kaiser, L., ed., *Manual of Phonetics* (Amsterdam, 1957).

Katzenstein, J., "Methoden zur Erforschung der Tätigkeit des Kehlkopfes sowie der Stimme und Sprache," *E. Abderhaldens Handbuch der biologischen Arbeitsmethoden*, V, 7–1 (Berlin–Vienna, 1930), pp. 261–418.

Kempelen, W. de, *Le mécanisme de la parole* (Vienna, 1791).

Klagstad, H. L. Jr., *Vowel-Zero Alternations in Contemporary Standard Russian* (Ph.D. dissertation, Harvard University, 1954).

Koneczna, H., and Zawadowski, W., *Obrazy rentgenograficzne glosek rosyjskich* (Warsaw, 1956).

Košutic, R., *Gramatika ruskog jezika, 1* (Petrograd, 1919).

Kuryłowicz, J., "Contributions à la théorie de la syllabe," *Bulletin de la Société polonaise de linguistique, 8*, 80–114 (1948).

——, "Uwagi o polskich grupach spółgloskowych," *Bulletin de la Société polonaise de linguistique, 11*, 54–69 (1952).

Ladefoged, P., and Broadbent, D. E., "Information Conveyed by Vowels," *JASA, 29*, 98–104 (1957).

Liberman, A. M., et al, "The Role of Selected Stimulus Variables in the Perception of Unvoiced Stop Consonants," *American Journal of Psychology, 65*, 497–516 (1952).

Liberman, A. M., Delattre, P. C., Gerstman, L. J., and Cooper, F. S., "Tempo of Frequency Changes as a Cue for Distinguishing Classes of Speech Sounds," *Journal of Experimental Psychology, 52*, 127–137 (1956).

Licklider, J. C. R., "The Manner in Which and the Extent to Which Speech Can Be Distorted and Remain Intelligible," H. V. Foerster, ed., *Cybernetics: Transactions of the Seventh Conference* (New York, 1950), pp. 58–122.

Licklider, J. C. R., Bindra, D., and Pollack, I., "The Intelligibility of Rectangular Speech Waves," *American Journal of Psychology, 61*, 1–20 (1948).

Licklider, J. C. R., and Miller, G. A., "The Perception of Speech," in S. S. Stevens, ed., *Handbook of Experimental Psychology* (New York, 1951), pp. 1040–1074.

Lisker, I., "Minimal Cues for Separating /w,r,l,y/ in.Intervocalic Position," *Word, 13*, 256–268 (1957).

Lloyd, R. J. "On Consonant Sounds," *Proceedings of the Royal Society of Edinburgh, 22*, 224–263 (1898).

Lotz, J., "Speech and Language," *JASA, 22*, 712–717 (1950).

——, "The Structure of Human Speech," *Transactions of the New York Academy of Sciences*, Series II, Vol. 16, No. 7, 373–384 (1954).

MacKay, D., "The Nomenclature of Information Theory," H. von Foerster, ed., *Cybernetics: Transactions of the Eighth Conference* (New York, 1951), pp. 222–235.

Malécot, A., "Acoustic Cues for Nasal Consonants," *Language, 32*, 274–284 (1956).

Malmberg, B., "The Phonetic Basis for Syllable Division," *Studia Linguistica, 9*, 80–87 (1955).

Mandelbrot, B., "Structure formelle des textes et communication," *Word, 10*, 1–27 (1954).

——, *Contribution à la théorie mathématique des jeux de communication = Publications de l'Institut de statistique de l'Université de Paris, 2*, 1–2 (1953).

Martinet, A., "Neutralisation et archiphonème," *Travaux du Cercle linguistique de Prague, 6*, 46–57 (1936).

Mathes, R. C., and Miller, R. L., "Phase Effects in Monaural Perception," *JASA, 19*, 780–797 (1947).

Menzerath, P., and Lacerda, A. de, *Koartikulation, Steuerung und Lautabgrenzung* (Berlin–Bonn, 1933).

Meyer-Eppler, W., "Untersuchungen zur Schallstruktur der stimmhaften und stimmlosen Geräuschlaute," *Zeitschrift für Phonetik, 7*, 89–104 (1953).

——, "Zum Erzeugungsmechanismus der Geräuschlaute," *Zeitschrift für Phonetik, 7*, 196–212 (1953).

Milewski, T., "The Concept of the Word in the Languages of North American Natives," *Lingua Posnaniensis, 3*, 248–267 (1951).

Miller, D. C., *The Science of Musical Sounds* (New York, 1916).

Miller, G. A., *Language and Communication* (New York, 1951).

——, "The Perception of Speech," *For Roman Jakobson* ('s-Gravenhage 1956), pp. 353–360.

——, "Some Effects of Intermittent Silence," *American Journal of Psychology, 70*, 311–313 (1957).

Miller, G. A., and Nicely, P. E., "An Analysis of Perceptual Confusions among Some English Consonants," *JASA*, *27*, 338–352 (1955).

Miller, G. A., and Wiener, F. M., *Transmission and Reception of Sounds under Combat Conditions = Summary Technical Report of Division 17, NDRC*, Vol. 3 (Washington, 1946).

Mjasnikov, L. L., "Ob'ektivnoe raspoznavanie zvukov reči," *Žurnal texničeskoj fiziki*, No. 3, 109–115 (1943).

——, "Zvuki reči i ix ob'ektivnoe raspoznavanie," *Vestnik Leningradskogo universiteta*, No. 3, 3–13 (1946).

——, "Fizičeskie issledovanija zvukov russkoj reči," *Izvestija Akademii Nauk SSSR, Serija fizičeskaja*, *13*, No. 6, 723–727 (1949).

O'Connor, J. D., and Trim, J. L. M., "Vowel, Consonant and Syllable – a Phonological Definition," *Word*, *9*, 103–122 (1953).

O'Connor, J. D., Gerstman, L. J., et al., "Acoustic Cues for the Perception of Initial /w,j,r,l/ in English," *Word*, *13*, 24–43 (1957).

Paget, R., *Human Speech* (London–New York, 1930).

Panconcelli-Calzia, G., *Die experimentelle Phonetik in ihrer Anwendung auf die Sprachwissenschaft* (Berlin, 1924).

Perrett, W., *Some Questions of Phonetic Theory* (Cambridge, 1919).

Peškovskij, A., "Desjat' tysjač zvukov russkogo jazyka," *Sbornik statej* (Leningrad, 1925), pp. 167–191.

——, *Russkij sintaksis v naučnom osveščenii* (Moscow, 1956).

Peterson, G. E., and Barney, H. L., "Control Methods Used in a Study of the Vowels," *JASA*, *24*, 175–184 (1952).

Petrovskij, S., *Nabljudenija nad prodolžitel'nost'ju i vysotoj proiznošenija zvukov v slove* (Kazan', 1903).

Pike, K. L., *Phonetics* (Ann Arbor, 1943).

——, "Grammatical Prerequisites to Phonemic Analysis," *Word*, *3*, 155–172 (1947).

——, *Phonemics* (Ann Arbor, 1947).

——, "Operational Phonemics in Reference to Linguistic Relativity," *JASA*, *24*, 618–25 (1952).

Polivanov, E. D., *Konspekt lekcij po vvedeniju v jazykoznanie i obščej fonetike* (Petrograd, 1916).

Pollak, I. and Ficks, L., "Information of Elementary Multidimensional Auditory Displays," *JASA*, *26*, 155–159 (1954).

Potter, R. K., Kopp, G. A., and Green, H. C., *Visible Speech* (New York 1947).

Potter, R. K. and Peterson, G. E., "The Representation of Vowels and their Movements," *JASA*, *24*, 528–535 (1948).

Potter, R. K., and Steinberg, J. C. "Toward the Specification of Speech," *JASA*, *22*, 807–819 (1950).

Purkině, J. E., *Badania w przedmiocie fizyologii mowy ludzkiej* (Cracow, 1836); also published in *Kwartalnik naukowy 2*, 121ff. and *3*, 104ff. (1835).

Radley, J.-P. A., *The Role of Formant Transitions in the Identification of English Stops* (S.M. Thesis M. I. T., 1956).

Reformatskij, A. A., "O sootnošenii fonetiki i grammatiki (morfologii)" in *Voprosy grammatičeskogo stroja* (Moscow, 1955), pp. 92–112.

——, "Soglasnye, protivopostavlennye po sposobu i mestu obrazovanija, i ix var'irovanie v sovremennom russkom literaturnom jazyke," *Doklady i soobščenija Instituta jazykoznanija AN SSSR*, No. 8, 3–23 (1955).

——, "Fonologičeskie zametki," *Voprosy jazykoznanija*, *6*, 2, 101–102 (1957).

Richter, E., "Die italienische č- und š-Laute," *Archives néerlandaises de phonétique expérimentale*, *16*, 1–38 (1940).

Romportl, M., "Fonetická studie o ruském ы," *Slavia*, *22*, 529–556 (1953).

Rousselot, P., *Principes de phonétique expérimentale* (Paris, 1897–1908).

Rožanskaja, E. V., "K voprosy o matematičeskom obosnovanii teorii razborčivosti," *Trudy komissii po akustike, 7*, 53–60 (1953).

Russell, J. O., *The Vowel* (Columbus, Ohio, 1928).

Sacia, C. F., and Beck, C. J., "The Power of Fundamental Speech Sounds," *Bell System Technical Journal, 5*, 393–403 (1926).

Sapir, E., *Language* (New York, 1921).

——, *Selected Writings* (ed. D. G. Mandelbaum) (Berkeley and Los Angeles, 1949).

Saussure, F. de, *Cours de linguistique générale* (Paris, 1949).

Schatz, Carol D., "The Role of Context in the Perception of Stops," *Language, 30*, 47–56 (1954).

Sievers, E., *Grundzüge der Phonetik* (Leipzig, 1881).

Smith, Svend, "Chest Register versus Head Register in the Membrane Cushion Model of the Vocal Cords," *Folia Phoniatrica, 9*, 32–36 (1957).

Sovijärvi, A., *Die gehaltenen, geflüsterten und gesungenen Vokale und Nasale der finnischen Sprache* (Helsinki, 1938).

Steinitz, W., *Russische Lautlehre* (Berlin, 1957).

Stetson, R. H., *Bases of Phonology* (Oberlin, O., 1945).

Stevens, K. N., "Autocorrelation Analysis of Speech Sounds," *JASA, 22*, 769–771 (1950).

Stevens, K. N., and House, A. S., "Auditory Testing of a Simplified Description of Vowel Articulation," *JASA, 27*, 882–887 (1955).

——, "Development of a Quantitative Description of Vowel Articulation," *JASA, 27*, 484–493 (1955).

——, "Studies of Formant Transitions Using a Vocal Tract Analog," *JASA, 28*, 578–585 (1956).

Stevens, S. S., and Davis, H. W., *Hearing* (New York, 1938).

Stumpf, C., *Die Sprachlaute* (Berlin, 1926).

Šapiro, A. B., *Russkoe pravopisanie* (Moscow, 1951).

Šaxmatov, A. A., *Očerk sovremennogo russkogo literaturnogo jazyka* (Moscow, 1941).

Ščerba, L., "Quelques mots sur les phonèmes consonnes composés," *Mémoires de la Société Linguistique de Paris, 15*, 237–241 (1908–1909).

——, *Court exposé de la prononciation russe* (Association phonétique internationale, 1911).

——, *Russkie glasnye v kačestvennom i količestvennom otnošenii = Zapiski istoriko-filologičeskogo fakul'teta Imperatorskogo St. Peterburgskogo universiteta, 107*, (1912).

——, *Izbrannye raboty po russkomu jazyku* (Moscow, 1957).

Tarnoczy, T., "Resonance Data Concerning Nasals, Laterals and Trills," *Word, 4*, 71–77 (1948).

——, "Vibration of the Vocal Cords and their Opening Quotient," *Linguistica, 1* (1941–1945), *Acta Instituti Hungarici Universitatis Holmiensis*, Series B, 43–56 (Stockholm, 1952).

——, "Die akustische Struktur der stimmlosen Engelaute," *Acta Linguistica* (Budapest) *4*, 313–349 (1954).

Technical Aspects of Visible Speech = Bell System Technical Monograph B-1415 (1946).

Thienhaus, E., and Barczinsky, L., "Klangspektren und Lautstärken deutscher Sprachlaute," *Achives néerlandaises de phonétique expérimentale, 11*, 47–69 (1935).

Tomaševskij, B., *O stixe* (Moscow, 1929).

Tomson, A. I., "Fonetičeskie ètjudy," *Russkij filologičeskij vestnik, 54*, 199–244 (1905).

——, *Obščee jazykovedenie* (Odessa, 1910).

——, "Die Erhärtung und Erweichung der Labiale im Ukrainischen," *Zapiski ist.-fil. vidd. A. N. Ukr. SSR, 13–14*, 253–263 (1927).

——, "Über die weichen Konsonanten," *Zeitschrift für slavische Philologie, 8*, 92–101 (1931).

——, "Bemerkungen über die š-Laute," *Zeitschrift für slavische Philologie, 11* (1934).

Trager, G. L., "The Phonemes of Russian," *Language*, *10*, 334–344 (1934).

Trendelenburg, F., "Elektrische Methoden zur Klanganalyse," *E. Abderhaldens Handbuch der biologischen Arbeitsmethoden*, *5*, 7–1, 787–870 (Berlin-Vienna, 1930).

Trendelenburg, F., and Franz, E., "Sprachuntersuchungen mit Siebketten und Oszillograph," *Wissenschaftliche Veröffentlichungen aus den Siemens Werken*, *15*, 2, 78–94 (1936).

Trubetzkoy, N. S., *Das morphonologische System der russischen Sprache = Travaux du Cercle linguistique de Prague*, *5*, 2 (1934).

——, "Die Aufhebung der phonologischen Gegensätze," *Travaux du Cercle linguistique de Prague*, *6*, 29–45 (1936).

——, *Principes de Phonologie* (Paris, 1949).

Türk, W., "Über die physiologisch-akustischen Kennzeiten von Ausgleichsvorgängen," *Akustische Zeitschrift*, *4*, 129–145 (1940).

Twaddell, W. F., *On Defining the Phoneme = Language Monograph #16* (Baltimore, 1935).

Twistleton, Edward, *The Tongue not Essential to Speech, with Illustrations of the Power of Speech in the African Confessors* (London, 1873).

Ulaszyn, H., *Zarys wymowy rosyjskiej* (Warsaw, 1953).

Ušakov, D. N., *Kratkoe vvedenie v nauku o jazyke* (Moscow, 1929).

——, "Svedenija po pravopisaniju, proiznošeniju i grammatike," *Tolkovyj slovar' russkogo jazyka*, *1*, (Moscow, 1935), col. xxx–lxxvi.

Van den Berg, J., *Physica van de stemvorming, met toepassingen* (Diss. Groningen, 1953).

Vietor, W., *Elemente der Phonetik des Deutschen, Englischen und Französischen* (Leipzig, 1923).

Vinogradov, V. et al, *Grammatika russkogo jazyka* (Moscow, 1952–1954).

Vinokur, G., *Russkoe sceničeskoe proiznošenie* (Moscow, 1948).

Wegel, R. L., "Theory of Vibration of Larynx," *Bell System Technical Journal*, *9*, 207–227 (1930).

Wheatstone, Sir Charles, "Reed Organs, Speaking Machines, etc.," *The Scientific Papers of Sir Charles Wheatstone* (London, 1879), 348–367.

Whitney, W. D., *A Sanskrit Grammar* (Cambridge, Mass., 1941).

Whorf, B. L., "Linguistics as an Exact Science," *The Technology Review*, *43*, 2 (Dec. 1940).

Winckel, F., "Die menschliche Stimme in elektrotechnischen Modellvorstellungen," *Elektrotechnische Zeitschrift*, *73*, 708–711 (Nov. 1952).

——, "Elektroakustische Untersuchungen an der menschlichen Stimme," *Folia Phoniatrica*, *4*, 93–113 (1952).

——, "Analyse und Synthese der Sprachlaute," *Funk und Ton*, *8*, 50–54 and 99–102 (1954).

Zipf, G. K., *Psychobiology of Language* (Boston, 1935).

——, *Human Behavior and the Principle of Least Effort* (Cambridge, Mass., 1949).

Zwirner, E., "Phonologie und Phonetik," *Acta Linguistica I*, 29–48 (1939).

——, "Phonologische und phonometrische Probleme der Quantität," *Proceedings of the Third International Congress of Phonetic Sciences* (Ghent, 1939), pp. 57–66.

Žinkin, N., "Vosprijatie udarenija v slovax russkogo jazyka," *Izvestija Akademii Pedagogičeskix Nauk RSFSR*, *54*, 7–82 (1954).

——, *Mechanizmy reči* (Moscow, 1959).