

UCLA

Working Papers in Phonetics

Title

WPP, No. 8: Synthetic Elocution

Permalink

<https://escholarship.org/uc/item/9zc6c6gp>

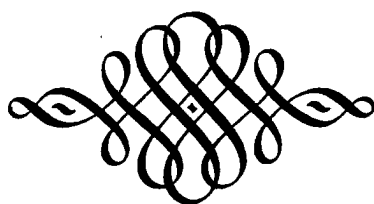
Author

Vanderslice, Ralph

Publication Date

1968-02-01

SYNTHETIC
ELOCUATION



Considerations in Automatic
Orthographic-to-Phonetic
Conversion of English
with special reference to
PROSODIC FEATURES

by Ralph Vanderslice

Working Papers in Phonetics, Eight
UCLA: February, MCMLXVIII

SYNTHETIC ELOCUTION

Considerations in Automatic Orthographic-to-Phonetic Conversion
of English with Special Reference to Prosodic Features

Ralph Vanderslice

Working Papers in Phonetics 8

February 1968

University of California, Los Angeles

TABLE OF CONTENTS

Acknowledgements		iii
List of figures and tables		iv
 Chapter		
I. Elocution, Linguistics, and the Vocal Über-Marionette		1
Orthographic-to-phonetic conversion	1	
Elocution and linguistics	3	
<i>Elocution</i> as tabu word	5	
Theory vs. pedagogy	9	
Thought vs. feeling	14	
Indexical vs. linguistic features	20	
Aesthetic features	32	
Rules or impulses	35	
"Rules" or algorithms	41	
Transliteration or analysis	44	
O-P conversion and machine translation	45	
Orthoepic and prosodic levels	47	
II. Towards a Prosodic Model of American English		51
Tasks in a reading machine	51	
Canonical orthoepic forms	53	
Orthoepic vs. lexical representations	59	
The place of "expressive" prosodic variations	60	
Pitch accent and cadence	63	
IPA stress marks reinterpreted	67	
Weak vs. strong syllables	69	
Accent and emphasis	75	
Cadence and endglide	77	
Pauses and tessitural shifts	83	
A fragment of the prosody	87	
Sample text exemplifying prosodic output	101	
Putative stress distinctions reconsidered	107	
Summary and outlook	114	
 Bibliography		 117

When a man starts to acknowledge his debts, he is beginning his autobiography. -- E. Gordon Craig

ACKNOWLEDGEMENTS

Considerable financial support for this work was provided through a grant from the Office of Naval Research for the study of attitudinal and emotional cues in speech. I am grateful for the freedom I had to pursue the line of research reported here, which, as the work progressed, increasingly appeared to be the paramount prerequisite to a useful study of indexical features, including paralinguistic ones.

I gladly acknowledge my debt to the teachers to whom I owe most of my small knowledge of English intonation. Those in addition to my committee I have been fortunate enough to study with are Professors Charles Hockett, Arthur Bronstein, Robert Stockwell, and Noam Chomsky. I have also benefitted enormously from conversations with a number of visiting faculty at UCLA, especially Professors Kerstin Hadding-Koch and Alexandr Isačenko, and with other visitors to the Phonetics Laboratory, notably Professors J.C. Catford, Gunnar Fant, Lee Hultzén, and John Morton.

I am grateful for the suggestions and criticisms of Professors Kenneth Pike, Seymour Chatman, and Robert Beloof. I am indebted also to Professors Dwight Bolinger and David Abercrombie (whom I count among my teachers although I have yet to meet them) and to Franklin Cooper and Ignatius Mattingly.

The views put forward in this work incorporate large borrowings from the teachings of these scholars; inevitably they entail also certain deviations. But I believe with Sledd (1957) that "to good teachers, the best gratitude is the best criticism of which the student is capable, [which] may at least suggest the extent to which the educational efforts of the linguists have succeeded (or failed)."

Valuable guidance in application of computers for language data manipulation was provided by John Olney, Mark Stein, Professors George Moore and Martin Kay, and Dr. David G. Hays. I have benefitted from the insightful interest of Professor George Allen.

The technical staff of the Laboratory, especially Richard Harshman, Stan Hubler, Willie Martin, and Rick Ratcliffe, have helped generously, as have Dr. Minoru Hirano and John Ohala. Julie Haaker and Janice Yamane gave valuable secretarial assistance. And Jeanne Yamane merits special thanks for her supremely competent work on the manuscript.

I owe a large debt of gratitude to my family, and particularly my wife, Carolyn, for patience, encouragement, and help during the past many months.

Finally, I am glad to have this opportunity to express, however inadequately, my appreciation and gratitude to Peter Ladefoged, who as my mentor and employer has done so much to make my years at UCLA both academically and personally rewarding. His friendship and his counsel have meant more to me than I can say.

LIST OF FIGURES AND TABLES

Figure 1.1	Block diagram of orthographic-to-phonetic conversion	2
Figure 1.2	Diagram of diachronic shifts in domain of concern with O-P conversion (after Hargis)	12
Figure 1.3	Relations among phonetic features, their bases, and information conveyed (after Ladefoged)	26
Figure 1.4	A further-elaborated version	27
Figure 1.5	Block diagram of linguistic and indexical feature	31
Figure 1.6	Block diagram of conceptual O-P converter	48
Table I.	Basic prosodic output	85
Table II.	Partial inventory of related indexical features	86
Table III.	Sample text transcribed, with respect to canonical orthoepic stress patterns and pause-assignment indices, as it might be represented at the input to the prosody	100

Working Papers in Phonetics is an irregular series put out by members of the Phonetics Laboratory, University of California, Los Angeles. The principle object of the series is to inform colleagues in the field of current work, so that we might benefit from comments and criticism before regular publication. The series also serves as a continuing progress report on our sponsored research.

This work was a dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in Speech. The committee in charge comprised Professors Victoria A. Fromkin, Daniel E. Vandraegen, Donald E. Hargis, Eugène J. Brière, and Peter Ladefoged (Chairman).

Such examples of historical precedence . . . are not to be thought of as mere curiosities, as fascinating relics of a dead past, or as amusing glimpses at the vague fumbings of our forefathers for ideas beyond their grasp. No; study of the history of science shows over and over again the cyclic process of its evolution -- ideas and theories coming to a stop because of a lack of technique, and the later reciprocal effect of new techniques upon revival and extension of earlier theory. We cannot escape our past; it continually shapes our ideas and our actions.

-- Colin Cherry (1961)

I

ELOCUTION, LINGUISTICS, AND THE VOCAL ÜBER-MARIONETTE

The function which maps sentences of a written English text onto corresponding sentences of spoken English is of considerable interest. It has practical interest for such applications as a high performance reading machine for the blind, for example, and it has theoretical interest for the art of elocution and the science of linguistics.

The process of translating from texts to spoken renditions of them can be viewed as a black box system: the black box being a device of unknown construction whose character must be inferred from the nature of its output in comparison to its input -- i.e. the changes wrought on what passes through it. In this case the input comprises orthographic texts, the output is phonetic renditions, and the device will be called, for reasons to be developed below, the *elocutionary transfer function* (figure 1.1A -- cf. Lazier 1966).

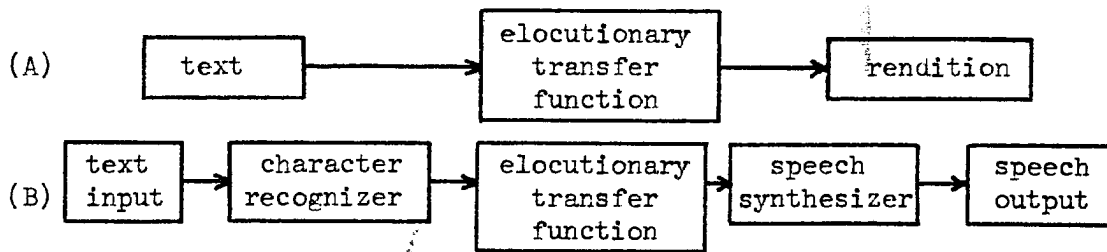


Figure 1.1 Black box model of orthographic-to-phonetic conversion: (A) simplest version; (B) first elaboration of tasks for reading machine.

An automatic high performance reading machine would require, as schematized in figure 1.1B, a means of getting the text into computer-compatible form at the input, a speech synthesizer at the output, and some rather sophisticated processing in between (Cooper 1963) to simulate the elocutionary transfer function. The conceptual design of such a machine furnishes a model of certain crucial aspects of linguistic competence, and that portion of the device already realized -- the speech synthesizer -- offers the best testing ground for aspects of an elocutionary theory.

A speech synthesizer is a sort of "vocal über-marionette" (cf. Craig 1907) which will perform exactly as it is told. It will repeat the same performance precisely, each time it gets the same instructions, or it will vary one parameter (upon instruction) keeping all the rest the same. But the synthesizer, like any computer, requires instructions which are complete and explicit. That is of course its great virtue for experimental phonetics and experimental elocution: it produces exactly and only those vocal features which are generated by the rules to be tested. And the traditional objections to avowedly rule-governed human performances -- the charges of insincerity and exhibitionism routinely laid against the elocutionists of an earlier

era -- are quite clearly irrelevant vis-à-vis a machine. Indeed the ultimate insult used by the currently ascendant "oblique" school of acting and oral interpretation against their elocutionary predecessors -- the charge of being "mechanical" -- describes not only a desideratum but a basic prerequisite for procedures that are to be carried out by computer.

The prosodic model put forward in this study grew out of efforts aimed at solving at least some of the problems inherent in deriving from orthographic text the information needed to correctly assign prosodic features in a synthetic speech rendition of it -- and indeed the problem also of what the prosodic features are which need to be assigned, at least for a restricted range of input texts. The aim of this introductory chapter is to relate the considerations involved in the conceptual design of such a device to the fields of study concerned with prosodic feature assignment: viz. elocution and linguistics. *Elocution* is the traditional name for the art which has claimed orthographic-to-phonetic conversion as its purview for upwards of two centuries (from classical times, in fact). The main invasion of linguistics into this precinct may be dated from the publication of the Trager and Smith (1951) outline -- although earlier forays may of course be discerned, especially in the British tradition of phonetics.

In one sense this confrontation is a reconvergence. The eighteenth century did not share our present-day departmentalizations of knowledge, and elocution was felt to be just as closely related to orthoepy, lexicography and grammar as to the art of acting: the two best known figures in the golden age of the English elocutionary

movement -- both professional actors -- both published dictionaries of the English language (Sheridan 1789 [1780], Walker 1791) to which were prefixed substantial treatments of accent assignment and the like (e.g. Sheridan's "Prosodial Grammar"). The elocutionists of that era are to be placed quite as squarely among the forebears of modern linguistics as of modern "oral interpretation".

In the intervening two centuries, the separate development and evolution of these respective specializations (especially in this country, where linguistics waxed under the aegis of cultural anthropology) have produced a wide divergence suggesting perhaps even total speciation. Nevertheless a fecund union of the two is a prospect which has intrigued a number of observers (Whitehall 1951, 1956; Chatman 1957; Epstein and Hawkes 1959; cf. also Sebeok 1960).

Neither side of the dichotomy is entirely unaware of the existence of the other. Akin (1961) exhorted her fellow interpreters:

We should be setting the pace, not explaining our place.
 . . . We have sat by and allowed other departments to take from us. . . . They are experimenting. They are doing extensive research. They are publishing the results. . . . We borrow the results and welcome them home.

The warning was put more explicitly by Parrish (1957):

. . . we must, building upon the contributions of our predecessors, carry forward their analysis of the normal speech pattern and its relation to meaning, and attempt to establish definite norms of expression. . . . And . . . we might attempt experimental studies to learn if and when and how communication may be made more effective by violating the norms. This area of linguistic study, which has remained pretty much neglected since the early attempts of Joshua Steele and James Rush, is a special province which belongs to elocutionists, and to no one else.

. . . Already our neglect of this our proper province has left the field open for others, and the phoneticians are fast moving in to occupy it.

Among linguists, preoccupation with scientism, and the fact that investigation of prosodic features is usually only a sideline, have made the art of oral interpretation appear rather remote from their concerns. But Gleason (1955), for example, in noting that a written sentence "may be read in a variety of ways" differing in features about the selection of which "in actual speech a native speaker would not be in the least capricious" remarks:

The nature of this difficulty is shown by the very existence of a special art, the oral interpretation of literature, which promotes the ability to select the oral rendition which will be accepted by the native speaker as most suitable for the context. That it is an art, rather than a science, demonstrates that there is no reliable method currently in use by which an author can inform his readers how he wants his work to be pronounced. If there were, expressive reading could be largely mechanical. As it is mechanical reading is far from expressive.

Which is perhaps the most gracefully epigrammatic characterization of the art of elocution extant.

Most of those concerned with the art of oral reading, especially in America, eschew the term *elocution* in any but pejorative contexts. This proscription was challenged by Parrish (1957 -- cf. quotation above), who wanted to rescue and rehabilitate the word:

We have ceased to give proper attention to that vital aspect of spoken language which those who shun the word "elocution" sometimes call "vocal expression."

For this aspect of linguistic study "elocution" is a useful and needed word. Better than any other it denotes that constant and inescapable element of

all live speech -- the movement of the voice with relation to meaning and feeling, the proper management of emphasis and subordination, pause and tempo, word-grouping, intonation and inflection.

But the proscription persists, even though more than a half century has gone by since the National Association of Elocutionists gave way to that of Academic Teachers of Public Speaking (now the Speech Association of America). The tabu on the term *elocution* is symptomatic of a fundamental schism within the art of oral interpretation which has contributed significantly to the gap which separates it from linguistics today.

The nature of this schism has been variously characterized. The best known and least adequate terminology for the opposing points of view is that which discerns a "natural" school, with Sheridan as its leader, and a "mechanical" school surrounding Walker (Fritz 1930, Robb 1941). The trouble with these terms as characterizations of the main opposing viewpoints in acting and elocution is that they are polysemous to the point of being useful only for purposes of polemic. If *natural* for instance is taken to designate the view that oral renditions should be based on and resemble closely the patterns of communication used in "real life" situations, then it is impotent for the purposes of discriminating, let alone characterizing, the opposing schools of elocutionary thought. As Parrish (1957) justly observes, "In the discussion and criticism of elocution a bewildering confusion has arisen from the loose and uncritical employment of the words 'natural,' 'artificial,' and 'mechanical.' The fact is that all the early elocutionists took nature as their guide." Nor can a valid distinction be drawn between the two viewpoints on the basis of belief in the

necessity for a reader to understand the text. It will be obvious that such a necessity, if it exists, has enormous implications for the design of a reading machine. But those who make facile use of the term *mechanical* in reference to one of the two main trends in elocutionary thought exploit its deprecatory connotations of mindless and unemotional perfunctoriness of manner -- thus implying a failure on the part of the opposing camp to appreciate the importance of understanding the meaning of the text. Bacon (1960) remarks that "much modern discussion of the seventeenth and eighteenth-century teachers of the art of speaking and reading -- for they did not normally distinguish between these arts . . . -- has confused our view of these teachers by dividing them artificially into 'natural' and 'mechanical' schools. . . . But *all* stressed the significance of meaning in the act of reading and speaking . . ." Indeed as will be illustrated extensively below, Bacon is entirely correct in saying that "meaning has always been at the center of the art of elocution." He is regrettably correct also in adding that

It is worth urging the fact simply because teachers of interpretation are sometimes led to say nowadays that the recognition of the significance of meaning is a modern contribution to the study.

A sample of the egregious anachronisms committed in this regard is the statement by Geiger (1962) attributing this "discovery" to Curry (1891, 1907); he generously allows that

twenty years ago one could find many a teacher who believed that effective public performance could be expected only from an interpreter who deeply understood literature. Such a teacher was probably either a conscious disciple of, or thinking along the same lines as, Professor Curry, who, in the early years of the century, was already laying the groundwork

for the fundamental importance of textual understanding to effective public reading. . . . probably no one more than Dr. Curry has insisted on understanding as central to the interpretive act itself.

The importance of correcting this historical solecism is that neither diachronically nor synchronically can recognition of the role of meaning be used to differentiate two schools of elocutionary thought.

Thus *natural* in the sense of taking real life communication as a guide, and *mechanical* in the sense of disregarding the importance of understanding the text, are neither one able to separate one side of the elocutionary schism from the other. This has been increasingly recognized in recent years and there have been various attempts to substitute other more precise terms. Vandraegen (1953) proposed *neoclassical* versus *romantic*. This choice is a reasonable one for capturing the fact that while all elocutionists accepted nature as their guide some of them, taking their goal to be "Nature methodized," were more favorable to rules and imitation in their pedagogy, whereas others laid more emphasis on the effect of spontaneity. But these words are scarcely an improvement from the standpoint of polysemy, and have already too many connotations to be pinned down usefully as technical terms. And there is an additional objection to "neoclassical" as an attributive for anyone interested in rules in the present period; it is not merely pejorative (like "mechanical") but anachronistic.

Whiting (1954), dealing specifically with acting, employed the terms *technical* and *psychological* for the competing schools. But these terms too are problematic. For example the chief exponent of the "psychological" (or "natural") school in acting, Stanislavski (1963

[c. 1924]), derogated those actors who "believe that any conscious factor in creativeness is only a nuisance," and held that "there can be no art without virtuosity, without practice, without technique." And those whose concern is with rules of orthographic-to-phonetic conversion as theory may not be at all interested in techniques or other facets of pedagogical technique -- and may not willingly concede all claims to psychological insight.

These caveats point up two confusions which regularly plague discussions of the opposing schools of elocutionary thought and which must be dispelled before the notion of synthetic elocution can be insightfully explored. One is a confusion between theory and pedagogy; the other involves the dichotomy which may be roughly equated with "thought-feeling" and will be elaborated below.

One discovers throughout the writings on elocution an almost complete failure to distinguish between *elocutionary theory* -- which must deal with and make predictions about the relation of phonetic output to orthographic input -- and *theories of elocutionary pedagogy*, which deal with and predict how students can best be taught to read aloud well. Reading aloud "well" may be glossed as "according to the rules" -- so that the ingredients of a theory are at least implicit in any pedagogy which gets down to cases (though this is contrary to the fashion in speech pedagogy nowadays). But a theory must state its claims formally. A theory of elocution would comprise explicit statements of the regularities which govern appropriate pronunciations of texts. In view of the continuous confusion of theory with practical application in this field, it is scarcely surprising that no such

theory has developed. Parrish (1957) points out that

most of the controversy over natural and mechanical methods has been concerned with methods of teaching or learning. The reasons for the current antipathy to teaching by the so-called mechanical method are seldom clearly stated or understood. We prejudge the issue by denominating anything mechanical as bad, and anything natural as good.

The primacy of pedagogy is expressed equally well, from the opposing camp, by Geiger (1962):

I recall how hard it is to become [an] oral interpreter . . . and then I recall how few instructional hours in oral interpretation most of our students take in their courses of study. . . .

Once one considers one's oral interpretation course, or program, as a matter of doing what one can . . . educational priorities are involved and, at that point certainly educational aim becomes determinative.

It is only by considering changing fashions in speech pedagogy that it is possible to account for "the durability and extent of the opprobrium" in the word *elocution* (Parrish 1957). The reaction was directed almost entirely against lapses of taste identified with the elocutionary movement; theoretical issues concerning the relation of rendition to text were only implicitly treated. Something of the nature of the "excesses" indulged in under the title of elocution during the decadent last days of the movement is suggested by Sansom (1960) in a list of malpractices observed more recently in choral speaking:

The choice of unsuitable material. . . .

Music and sound effects used as a background to the speaking. . . .

Flamboyant costumes . . . like glorified marching girls;

The "staging" of poems with backdrops and elaborate lighting, e.g., Tennyson's "The Splendour Falls,"

complete with castle, bugle, and fading sunset;

The arbitrary breaking-up of lines irrespective of the poem's form or meaning;

.

The use of artificial speech tunes;

In fact, all the old elocutionary tricks which public opinion had forced teachers to abandon in the individual speaking of verse.

Parrish (1957) parodies the current view of elocution as "the age of 'hollow bombast and grandiloquence,' of 'simpering adolescents' and 'affected females' in public recitation," and reminds his readers that

The violations of good sense and good taste . . . were due to conditions of the time and to the universal passion for ostentatious display to be found in people of a certain level of culture, and still manifested in some present-day radio and television shows, in popular entertainments, and in school declamation contests.

Hargis (1960) accounts for the shifting trends in pedagogical emphasis in an analysis which discerns three phases:

- [1] With . . . the mid-19th century . . . mechanical theories of elocution . . . the focus of concentration came to be almost exclusively upon the conscious manipulation of the vocal-physical mechanism to the exclusion of such considerations as meaning, mental-emotional response, unconscious vocal-physical response, and listener reaction. . . . [This] led to a rather complete absorption in the purely mechanical, outward aspects of vocal-physical operation *in vacuo*.
- [2] In the gradual movement away from these practices at the end of the century, the totality of the act gained the center of attention with particular emphasis upon analysis for meaning, "thinking-the-thought," and subconscious vocal-physical response. . . . Such an eclectic approach was the *sine qua non* of the revival of oral interpretation as a respectable companion in the field of speech and serves as the base for our contemporaneous theory. . . .
- [3] Currently, there appears to be a conscious gravitation from the pole of the mechanical excess

of elocution to that of the mechanical excess of criticism.

This analysis postulates a sort of pendulum-swing from excessive concern with oralization (system output), through a balanced and eclectic view of the whole process, to an overemphasis on textual analysis (system input), and can be indicated graphically in a block diagram of orthographic-to-phonetic (O-P) conversion as redrawn in figure 1.2.

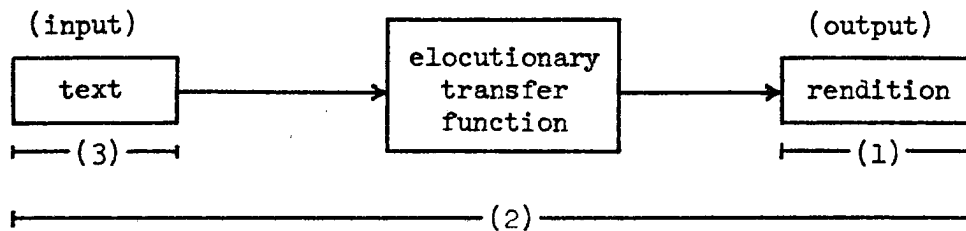


Figure 1.2 Model of O-P conversion showing domain of interest taken in total process by three successive elocutionary phases (Hargis 1960): (1) latter 19th C.; (2) contemporaneous eclectic approach; (3) current overemphasis on literary criticism.

Hargis' third stage, the development recently of a heavily "text-centered" approach to interpretation, has been noted by other observers. Thus Parrish (1957) states that teachers of elocution or oral reading "began to call their discipline 'expression,' then changed that to 'interpretation,' and in the process shifted their interest and their emphasis from the study of vocal expression to the study of literature." Glenn (1960) identifies "two tendencies which are having an adverse effect on the actor's vocal training, [viz.] recent attitudes in the teaching of oral interpretation, . . . and the alarming prevalence of misinterpreted and distorted Stanislavski methods. . . . Both . . . stress . . . the interpreter's understanding of the written word over his ability to communicate. . . ." He points out that while

"the predominantly literary attitude assumes . . . that the comprehension which comes with careful analysis and vocalizing will almost automatically produce a performance which, while not professionally expert in delivery, will adequately interpret the meaning of the selection read . . . this sequence is not necessarily automatic. . . ."

As an example of the tendency he deplores, Glenn cites a statement by Ostroff (1954) that "oral interpretation should be used as an approach to serve literature rather than vice versa." A related comment is that of Campbell (1967):

If the reader is to become a performer, he must perform for himself. . . . Oral interpretation has as its primary goal, not performance before an actual audience, but a quickened and deeper response to, and appreciation of, literature.

And Bacon (1960) contrasts Sheridan's definition of elocution ("the just and graceful management of the voice, countenance, and gesture in speaking") with

. . . a modern view of interpretation as the study of literature through the medium of oral performance. Both definitions reveal interest in the fundamental contributions of the oral performance. . . . But Sheridan's definition uses the text as a point of departure whereas the other uses the text as a point of return. And this, very simply, seems to be the clear line of change from the eighteenth century to our own time in the teaching of oral reading.

Geiger (1962) furnishes perhaps the clearest statement of the inversion of focus implicit in such views: "It was for Curry to emphasize once and for all the importance of understanding the literary text to performance. . . . recent writers, more than Curry, have stressed the importance of performance to understanding the literary text." And although the claim that the latter conception is new is as erroneous

(cf. Legouvé 1877) as the one dating the former from Curry, the implication is clear (despite Geiger's intention not to "speak with contempt of the art of delivery"):

I mean here, of course, that good performance helps an audience to understand the text. But also I mean that the performer's work helps *him* to understand the text -- and not good performance only is required in this case, for often enough the rather mediocre student job can be insightful to the performer himself, if not always particularly so to his audience or classmates.

It should be clear that where pedagogy is (a) taken as the basis of theory, and (b) directed toward explications de texte rather than the elocutionary transfer function itself, there is little hope of contributions to a formal elocutionary theory.

The second of two sources of confusion mentioned above as obfuscating the nature of the schism separating two schools of elocution is, roughly, that of thought and feeling. Of course the dichotomy itself has been almost universally recognized (except for the somewhat self-conscious rejection of it by adherents of the "New Criticism") yet those aspects of theories of elocutionary pedagogy intended to apply to cues of attitude and emotion in the rendition of texts are often treated as if they applied equally well to the communication of logical structures of thought (as in straight factual prose) and vice versa. An example of this is seen in the "distorted Stanislavski methods" (Glenn 1960, above) in which the complex mystique involving "memory of emotion," "living the part," etc. is assumed to apply throughout the range of the actor's problems of communication. Actually Stanislavski (1963 [c. 1924]) distinguished sharply between thought and feeling (he

in fact discerned a "troika" comprising thought, WILL, and feeling) and dealt only with the latter. So far as the reading of straight factual prose is concerned, Stanislavski offered no theory at all; he regarded this skill as part of the technical equipment the actor was presumed to bring to the task of mastering the higher aspects of the art.

The thought-feeling dichotomy maps onto a useful division of the tasks involved in O-P conversion in terms of the class or literary type of text to be pronounced. This was recognized by Legouvé (1877) who contrasted "correct reading" -- the ability to read aloud any factual document simply, directly, and clearly without fatigue -- with "art reading", of which he said: "To be worthy of such a name, Reading must rise to the dignity of being able to reproduce works of undoubted art, must become the interpreter of masterpieces of genius." Hargis (1964) observes that this dichotomy "was based on the stylistically plain, informative, non-emotional nature of the materials for correct reading as opposed to the stylistically more individual, imaginative, and highly emotional components of the literature for art reading." Parrish (1957) refers to the same distinction:

The utterance of highly emotional passages, as in drama and poetry, is much more variable and may not always have a standard pattern. What our students need primarily, however, is direction in the speaking of plain prose. And if there is not always a "right" way to speak an ordinary expository sentence, there are at least many wrong ways.

From this point of view he delivers a penetrating analysis of a paradox in the use of the term "natural" by modern teachers of oral interpretation and acting:

It is odd too that while insisting that speech must

be natural, they often prefer to read, and have their students read, materials that require them to be *un-*natural. Not concerned with effective expression of normal thought in normal situations, they seek out odd and eccentric characters to be impersonated and often prefer to catch them in the throes of some violent and perverted passion.

Subsumed in all these observations is the view that reading plain or factual prose calls upon a skill which is basic to all reading, and to which something additional must be added for handling fictive texts. The paradox noted by Parrish relates to the fact that most of those in the field today prefer to restrict the domain of "oral interpretation of literature" to fictive texts, excluding utilitarian, scientific, and in general all of what is here called "straight factual" prose. Thus Bacon (1960) points out that "interpretation tends (though not absolutely) to restrict its materials to works of imaginative literature, in keeping with its position in the liberal arts-humanities tradition in educational institutions." And Parrish puts it more strongly:

Writers on interpretation instead of teaching first how to communicate the meaning of a simple sentence plunge at once into the most difficult passages of literature . . . seeming to derive their knowledge and their methods from inspiration of the gods. This, supposedly, is "the natural method."

It is scarcely surprising that those who embrace this emphasis most avidly -- the exponents of the current "dramatistic" vogue -- attach little importance to the thought-feeling dichotomy, since both thought and feeling are felt to be inextricably intertwined in the types of text they regard as coming within the purview of their art. Thus Geiger (1962) states that "we cannot accept . . . bifurcation of poetic content into thought and feeling. . . ." and that "such dubious

categories as . . . 'logical thought' and 'emotional content' . . . seem to some of us imprecise, and even misleading in their implications for oral interpretation." This implies that greater precision will result from conflating the categories. It is true of course, as is discussed more fully below, that the results of the operations which assign correct (nonanomalous) prosodic renditions of straight factual prose may be considerably modified by the additional processes required for nonanomalous renditions of fictive texts. And the advantages of treating the former separately are not likely to seem particularly compelling to those who disdain the task of reading factual prose. Where, on the other hand, as in the present study, the prosodic features and rules to be incorporated in the model are delimited by restriction to just those needed for reading straight factual prose, it is helpful not only to observe the distinction but also to define it as precisely as possible.

The restriction of the model on the basis of what is needed for factual prose should not be taken as implying that the residue (the set of additional features needed for correctly reading fictive texts, including drama and poetry) is uninteresting or lies outside the purview of scientific elocutionary and linguistic study. On the contrary, the desideratum is obviously to construct as complete a model of oral reading (and of speech in general) as possible; including, ultimately, paralinguistic and kinesic factors. (An approach to a kinesic übermarionette may be discerned in the late Walt Disney's "audio-animatorics" -- e.g. the Lincoln figure at Disneyland and, formerly, at the New York World's Fair.) The restriction to plain prose is adopted

rather in the spirit expressed by Kruisinga (1943): "There is a time for all things, and the present writer believes that he serves the study of the subject best by not undertaking everything at once, and proposes to leave the examination of [further matters] for another opportunity, or for a successor of a younger generation." But there is, in addition, a substantive reason for focusing attention on a minimal model for O-P (orthographic-to-phonetic) conversion of straight factual prose: viz. the claim here postulated that this provides a motivated basis for distinguishing the grammatically relevant features from what are variously called expressive, paralinguistic, or referential ones. Thus Kurath (1964) notes that

English intonation is partly grammatical, partly referential. Before we can deal effectively with its expressive, attitudinal, and directive uses, which have reference to the speaker's feelings, to his attitude toward the content of his message and his audience, and to the effect he intends to produce upon the person(s) addressed, we must isolate, as best we can, all of its grammatical functions.

At the outset of this study the object was to develop rules for synthesizing vocal cues of attitude and emotion. It quickly became apparent, however, that the scope of this task was too great for the time available. The phonetic correlates of emotions and attitudes are themselves complex and not well understood; the problem of giving explicit directions for recognizing from the context when a particular attitude or emotion is appropriate is even more severe and may require extremely sophisticated pattern-recognition procedures. The first thing is to find out how to make a speech synthesizer sound angry, serene, amused, sarcastic, or whatever, even if at first it is necessary to insert "stage directions" in the text. But even this part of

the task is formidable, and it cannot be separated from that of constructing a model of the grammatically relevant prosodic features. That is, there is not much point in studying how attitudes and emotions are expounded by the so-called "paralinguistic" features (Trager 1958, Pittenger, Hockett, and Danehy 1960, Crystal and Quirk 1964) without knowing something about the ways in which they are also expounded by "marked" assignments of phonologically contrastive features; and clearly impossible to say much about the latter without reference to a model by which the unmarked assignment of such features may be accounted for: thus the task here essayed is inherently a part of the original object. What is claimed here is that a prosodic model for O-P conversion of factual English prose will incorporate just those features which expound *all* (but not, for the reasons just given, *only*) the grammatically relevant prosodic contrasts of English. This of course asserts a purview of grammar which is somewhat broader than that recognized by many linguists (see below); specifically it counts as belonging to grammar certain structural features "that chiefly manifest themselves in text segments comprising more than one sentence" (Olney and Londe 1966).

Undoubtedly the best treatment to date of the vexed question of what is and what is not linguistic in speech is that of Abercrombie (1967), who draws a distinction between *language* and *medium*:

This is a distinction which is not popularly made, and is not immediately obvious

. . . the language itself lies in the *patters* which the mediums form, and not in the physical objects or events, as such, of which the mediums consist. When we distinguish language from medium, what we are doing is to distinguish a pattern from its material embodiment, of which, in a sense, it is independent. Language, we could say, is *form*, while the medium is *substance*.

By making this fundamental distinction, Abercrombie is able to deal insightfully with properties of the medium which exist by virtue of the fact that "the pattern-forming potentialities of a medium [e.g. speech or writing] are much greater than any one language requires to use".

Two sorts of extralinguistic properties are discriminated: *indexical* and *aesthetic*. Indexical features comprise

- (a) those that indicate membership of a group;
- (b) those that characterize the individual;
- (c) those that reveal changing states of the speaker.

Abercrombie's subcategories and examples of such indexical features may be arranged in outline form as follows (headings in italics are his terms; those in roman are adopted here for brevity):

- (a) Sociosyncratic indices
 - 1. *Regional indices*
 - 2. *Status indices*
(e.g. the RP [Received Pronunciation] accent in England)
- (b) *Idiosyncratic indices*
(of sex, age; speech defects, etc.)
- (c) Nonce indices
 - 1. *Physical states*
(fatigue, catarrh, inebriation, talking with mouth full)
 - 2. *Affective indices*
(amusement, anger, contempt, grief, sympathy, suspicion)

It is at once apparent from this array of indexical feature types why the thought-feeling dichotomy is an inadequate basis for analyzing aspects of an elocutionary transfer function and their corollary literary types: *thought* may adequately identify *linguistic features*, but *feeling* falls far short of encompassing the *indexical features*, since it applies only to one subtype of the nonce indices; viz. the *affective* ones. But it is the whole range of indexical features which a

well trained actor or oral interpreter (and equally well the ultimate ideal reading machine) must be prepared to simulate. *Simulate* is here employed as a neutral term to mean "produce convincing and/or conventionally accepted cues of" some index, without intending any commitment in the old argument (cf. Plato's *Ion*) as to whether in portraying affective indices the actor must, or ought to, "feel" them. The need for the speaker himself to feel the emotions he wished his audience to feel was part of the intellectual baggage inherited by the early elocutionists from prior theories of rhetoric. Dodsley (1735) for example stating that

'Tis not enough that what you say is true,
To make *us* feel it, *you* must feel it too.

But contemporaneous expressions of the contrary viewpoint are found, as that of Holmes (1766 [c. 1755]):

Pronunciation, or *Moving Delivery*, which is the very Soul of all Rhetoric, consists in a due Management of the *Voice* and *Countenance*, as well as proper *Gesture* of the *Body* and *Hands*, according to the Nature of the *Passion* or *Thing* spoken of.

.
*Vary your Tone just as your Subjects go,
Cant not, nor pitch your Voice to high or low
Strain not, nor speak your Words too fast or slow.*

.
*Whatever different Points your Speech demand
In Joy, Grief, Hope, or Fear; with Art command
Your Body's Gesture, Countenance and Hand.*

.
Adorn with TROPES and FIGURES your *Oration*.
By VOICE and ACTION grace *Pronunciation*.

The statement that a speaker should "command" his gestures and facial expressions "with art" strongly implies that the vocal instruction to "vary your tone just as your subjects go" also refers to

deliberate, conscious design by the speaker. This will be seen to constitute one of the basic issues in the elocutionary schism of which a proper characterization is sought. Whiting (1954) notes that controversy has raged for decades over whether the actor should feel his part:

This question was originally posed in 1770 by the French philosopher and critic, Denis Diderot, who insisted that the actor should remain completely insensible to the emotion he portrays. The controversy reached a climax in 1888 when the English critic William Archer published a book on the subject entitled *Masks Or Faces?*

He observes that the French actor Coquelin (1841-1909) was "famed for his logic and mental control," and "established what Stanislavski calls the representational school of acting, which advocates creative imagination, study, and emotion during rehearsals but insists upon complete objectivity during performance. . . . This unemotional approach was challenged by Henry Irving" Whiting perspicaciously remarks that

Logic tells us that in the final analysis it does not matter whether the actor feels or not, so long as he presents a pattern of voice and action that will make his audience feel.

He thus directs attention to the real issue, i.e. the empirical question "Is the actor more likely to assume the correct outward pattern -- one that will convince and move his audience -- when feeling or not feeling?" It is useful to separate this issue from that of how the reader or actor should go about preparing for performance (which is ultimately a pedagogical question) even though a desideratum for a reading machine might be to simulate idealized sight reading, where there would be no separate rehearsal period. The speed of modern

digital computers is such that, given an algorithm adequate to map the text onto the rendition, and considering the restrictions on the syllabic rate of output speech acceptable to human listeners, real-time on-line O-P conversion should be possible with only negligible delays for processing the opening sentences (fortunately most of the context searching required in connected text is, as will be seen below, regressive). Humans have to rehearse because they haven't the information-handling capacity to read complex literary texts well at sight: they can't look and parse far enough ahead of nor remember far enough back from what they are pronouncing at any given moment to avoid sight-reading solecisms of the sort that are routinely heard from radio and TV announcers -- e.g. the reading on an 8:30 station break of "Next: Gemini Five Report. At 11:00," where adjective *next* was mistaken for its adverb homograph -- and in under-rehearsed readings of books, especially for children, where unexpected idioms cause misreadings such as

it w'd be the BIG YELLOW CAT that
picked the LITTLE cow to be HIS cow.
[instead of "It WOULD be . . ."],

and

TIME was when they HAD all they NEEDED
[instead of "time WAS . . ."].

It is perhaps not surprising, then, that some who recognize the human element in the work of the reader-actor regard it as other than an unalloyed advantage. In particular, Craig (1907) asserted that

Acting is not an art. . . . Art is the exact antithesis of Pandemonium, and Pandemonium is created by the tumbling together of many accidents; Art arrives only by design. Therefore in order to make any work of art it is clear we may only work in those materials with which

we can calculate. Man is not one of these materials.

.
The actor must go, and in his place comes the inanimate figure -- the über-marionette we may call him. . . .

Whatever the value or importance of "feeling" to the human interpreter in representing affective indices, it should be obvious that it cannot assist him in the simulation of other indexical features. Thus to take an obvious example an actor may be cast in a role requiring him to manifest a regional dialect-accent such as the Irish brogue, but an Irish "feeling" is unlikely to contribute to a solution of the phonetic problems posed.

Much of the technical training of an actor or oral interpreter obviously relates (or ought to relate) to mastering so far as possible the indexical features of his own speech; especially the sociosyncratic and idiosyncratic ones which in the monodialectal layman are substantially invariant. It is a commonplace that for an adult to learn the subtle distinctions in the pronunciation of another dialect of his native language equals or exceeds the difficulty of learning a foreign language. This may set a limit on what can be done by actors in this respect -- although clearly there are large individual differences and a knack for dialects may be part of what is comprehended by "talent". Furthermore, factors involving the audience, such as the exaggeration expected in dialectal stereotypes, may set limits too on the verisimilitude required or perhaps even desirable. A trading relationship between stereotype and verisimilitude in fact obtains in the portrayal of indexical features generally, and changing popular taste (audience expectation) may render obsolete a particular way of balancing these ingredients which previously found favor. This explains the curious

fact that the innovations of popular actors in every period -- e.g. Garrick (1717-79) -- have been praised for their "naturalness," irrespective of the theory or method underlying them. Stanislavski castigated his predecessors (and unreconstructed contemporaries) for their "rubber-stamp" simulations of affective indices, attributing their lack of verisimilitude to the imitation of other actors rather than life. The same theme runs through the writings in the anti-elocutionary revolt, accounting for the uniform use of the term *mechanical* in the sense of "stereotyped" -- and not in the sense of "performed without conscious ordering or deliberate structuring" (a sense which applies better to the romantico-psychological "natural" or "impulsive" theory; see below). Sansom (1960; above), stated that the revolt against elocutionary exhibitionism was caused by changes in public opinion and taste. The acceptable aesthetic balance between stereotype and verisimilitude had altered, or at least the fashion in stereotypes had changed.

Simulating the sociosyncratic indices of a group other than ones native speech community is difficult enough, but a more severe restriction applies to the simulation of idiosyncratic ones. Ladefoged (1959, 1967c) points out that although the latter may include some which are (like the linguistic and sociosyncratic features) part of an individual's learned speech behaviour, they "may also be due to anatomical and physiological consideration, such as the particular shape of the vocal cavities." Figure 1.3 reproduces Ladefoged's diagrams showing the relationships between the origins of the features observable in speech and the information they carry.

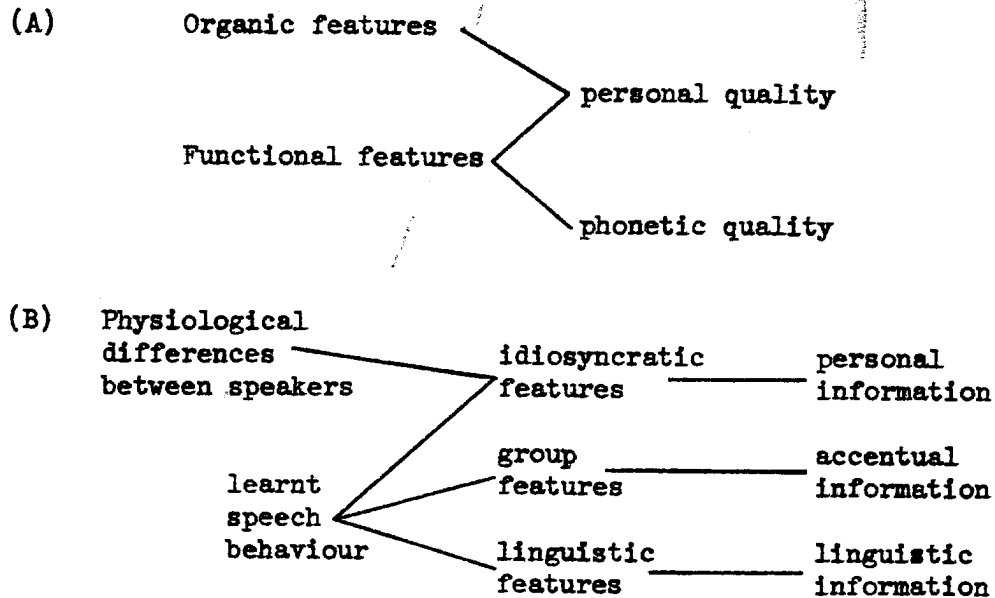


Figure 1.3 Simple (A) and elaborate (B) versions of relations between origins of phonetic features and information they convey (Ladefoged 1959, 1967c).

These diagrams do not include the third category in Abercrombie's analysis of indexical features (those that reveal changing states of the speaker). To show the nonce features properly requires the unlearned or asocial origins to be divided into, roughly, physiological and psychological ones. In view of Pittenger's (1957) caveat concerning this "traditional but somewhat problematical and perhaps misleading division," it may be better to use portmanteau terms such as *anatomical-physiological* versus *psychophysiological* so that the effects of phlegm blocking the nasal passages fall into the first category (along with the structural shape of the vocal cavities) but those of adrenalin in the bloodstream fall into the second, along with the emotions less obviously based in biochemistry. The inclusion of nonce indices suggests also an analysis of the personal information conveyed by phonetic differences into appropriate subcategories. These

elaborations are shown in figure 1.4.

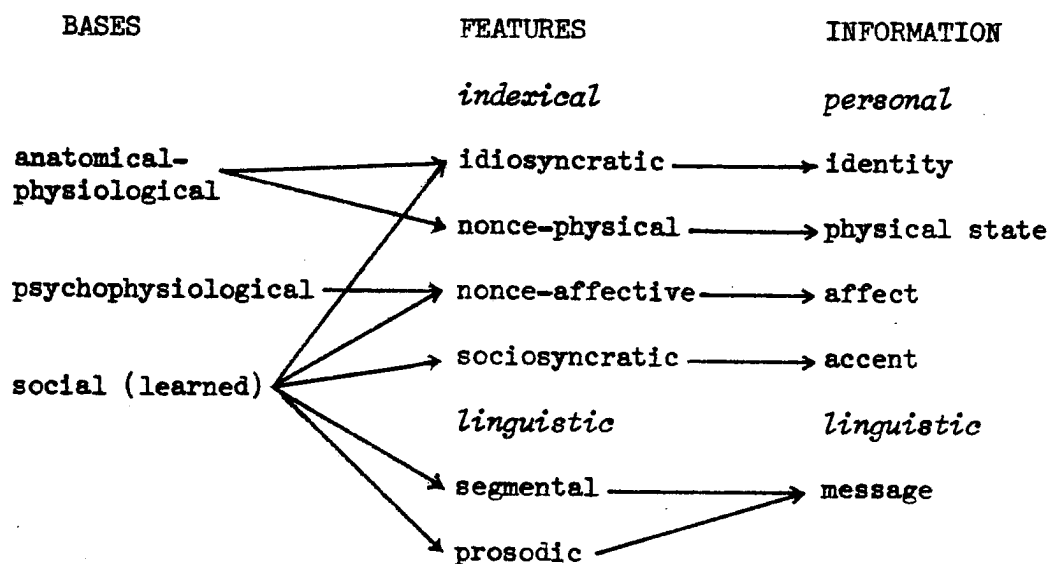


Figure 1.4 Relations among features, their bases, and the information conveyed: a further elaboration.

There is, of course, no assurance that "information" inferred from indexical features is valid. For example, although proponents of the so-called "voiceprint" method of speaker identification have postulated the existence of idiosyncratic features unique to each individual, experiments have shown that there is considerable intra-subject variation and inter-subject overlap even without attempts at voice disguise (Ladefoged and Vanderslice 1967). And the range over which learned indexical adaptation and compensation can operate, whether for group acceptance or for

vocal disguise, is quite large -- as shown by the arts of acting and elocution, the sine qua non of which is the ability to simulate indexical cues. Obviously, in the case of the idiosyncratic indices which arise largely from anatomical-physiological causes, there is a limit on the extent to which any human speaker can modify the vocal cues of his age and sex -- or other subtler indices of individuality -- and assume those of another person (real or fictional). Nevertheless, from ancient times this has been emphasized as part of the actor's, orator's, and rhapsode's arts. For example, Le Faucheur (c. 1702 [1657]) says that

In a *Prosopopeia*, nature her self shews us; First that the *Orator* ought to *change* his *Voyce*, to the end it may appear as if it were not *he* a *speaking*, but *some other Person* brought in by the *By*: And secondly, that he must *vary* it according to the *Diversity*, *Character* and *Business* of the *Persons* that he introduces, and feigns a *speaking* in this *Disguise*.

.

Upon a *Dialogism*, or *Conference*, where *two Persons* are brought in as 'twere *Dialoguing* one another, one of 'em moving the *Question* and t'other making the *Answer*, you must change your *voyce* by turns, as if *two Men* were really a talking together.

.

Upon these *Conferences* and *Rencounters*, we must always observe to *pronounce* the *Answer* with a *different Tone* from the last *Cadence* of the foregoing *Question* and *Objection*.

This advice was perpetuated by native English writers on elocution, e.g. Mason (1748). It is quite clear that the whole range of indexical features, not just affective ones, fell within the purview of the art of elocution as its pioneers defined it.

Whatever the difficulty for the human reader, with his in-built indices of idiosyncrasy and socio-geographical provenience, in simulating or suggesting the features appropriate to characters who may differ

from him in age, sex, dialect, physique -- and even, for some literature, in species -- these problems do not apply to the speech synthesizer, which in theory can simulate any sound that can be made by a human -- and even in practice can produce a large number of markedly inhuman ones. The best examples of synthetic mimicry are quite difficult to tell from the recorded speech of the human models (e.g. Fant and Mártony 1962, Strong 1967). On the other hand all examples of synthesis by rule to date betray their electronic origins by what may be regarded as indexical features of the terminal analog synthesizer group, with various idiosyncratic indices of the particular synthesizer thrown in.

By distinguishing between linguistic features and indexical features (following Abercrombie and Ladefoged) it is possible to treat with more precision than has heretofore been usual both the old dichotomy of thought versus feeling and the newer one of suprasegmental versus paralinguistic prosodic phenomena. Feelings may of course be manifested as the content of the message encoded in segmental linguistic features as well as by affective indices, but it is well known that in case of conflict or lack of congruence between the two it is the affective cues which are more likely to be attended to by the listener (Pittenger 1957). Extreme cases of incongruity are of course likely to produce a comic effect. As Le Faucheur (1702 [1657]) says:

... it would be ridiculous to speak *Common and Ordinary Things*, that happen every day, with a *Tragical concern* or a *Tone* of Admiration; and as absurd on the other hand, to speak of *Great Affairs* and matters of extraordinary moment with a *low unconcerned* and *familiar Voyce*.

The importance of congruence between matter (linguistic content) and manner (indexical, especially affective, cues) is stressed by almost every writer on elocution. Mason (1748) in commenting on a fault of pronunciation in which "persons speak too quick" says that

This Manner of reading . . . is very improper in reading Books of Devotion and Instruction, and especially the sacred Scriptures, where the Solemnity of the Subject or the Weight of the Sense demands a particular Regard. But it is most of all inexcusable to read Forms of Prayer in this Manner as Acts of Devotion.

One of the few writers diverging even mildly from this opinion is Cockin (1775), who claims "that the warmth and Energy of our delivery in reading ought to be inferior to that used in speaking upon subjects in which we are immediately interested." But in general, elocutionary writers have stood for complete congruence of affective indices and linguistic content in communications concerned wholly or in part with affects.

The reader or actor, obviously, in the absence of explicit stage directions, has to infer the existence and nature of the affect from the orthographic text in order to be able to assign appropriate, or congruent, affective indices. These two tasks have usually been separately discerned; the first being treated as part of understanding the text, the latter as emotional response. Thus the "golden rule" of Quintilian (c.AD 90) is given by Bishop Burnet in his *Discourse on the Pastoral Care* as quoted by Mason (1748); states

that to make a man speak well, and pronounce with a right Emphasis, he ought thoroughly to understand all that he says, be fully persuaded of it, and bring himself to have those Affections which he desires to infuse into others.

But the task of understanding itself has not usually been subcatego-

rized such that information relevant to assignment of indexical features is considered separately from that needed directly in the assignment of linguistic ones. This latter dichotomy of course underlies the assumption throughout the present work that the task of assigning prosodic features can be usefully divided into two parts: one the *unmarked* assignment of grammatically relevant units (which is required for high-quality O-P conversion of even straight factual prose); the other the *marked* assignment of these features, and/or "paralinguistic" ones, which taken together constitute the indexical features (and would be required to render non-anomalously most fictive works of literary art). Thus the elocutionary transfer function may now be represented as in figure 1.5, where linguistic features are assigned

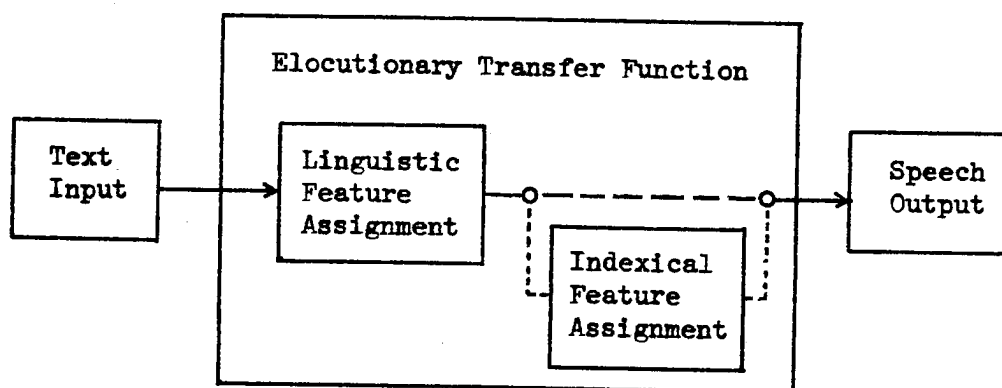


Figure 1.5 Elaboration of O-P conversion scheme to show alternate paths for fictive texts (dotted lines) and straight factual prose (dashed line).

for all O-P conversion but the route through indexical feature assignment is optional, depending on the nature of the text. Obviously every reader conveys indexical cues even when reading plain prose. The distinction postulated here is based on whether the text can be read aloud appropriately without the simulation of any marked or time-

varying indexical features. It is assumed that, within certain limits, features which can serve as affective indices if they come and go will, if they remain invariant, be interpreted as group or idiosyncratic (or possibly nonce-physical) indices forming a neutral background for the communication -- much as a shift, within limits, in the ranges of formant frequencies for all the vowels in an utterance is interpreted as an alteration of indexical features rather than affecting the identification of the words (Ladefoged 1959, 1967c).

The above discussion of the relation of indexical features to linguistic ones has skirted carefully around the other class of extralinguistic features discussed by Abercrombie (1967) as among the properties a medium may carry: viz., aesthetic features. It is apparent that aesthetic features of the aural medium, speech, fall within the purview of elocution. Mason (1748) emphasized that

. . . a decent and graceful Pronunciation (provided that there be no Defect in the Organs of Speech) may be obtained by Rule, Imitation, and Practice; and, when attained, will give a Beauty to your Speech, a Force to your Thoughts, and a Pleasure to the Hearers, not to be expressed

But it is difficult to show the relation of aesthetic features to the others in an explicit scheme such as that in figure 1.5. Some voices are more beautiful than others, of course, but these differences matter much less in acting and elocution than in singing. Practically every statement about beauty and aesthetic merit in the elocutionary literature implies that they consist in doing well those tasks which contribute to communication of the linguistic message and whatever indexical cues may be appropriate. Indeed, to give attention to aesthetic features per se -- ones which do not relate directly to

communication -- is to risk the charge of exhibitionism. This is a sin for which there is no forgiveness in eyes of anti-elocutionists. Babcock (1916) says that anything distracting the audience from the *what* to the *how* destroys unity and harmony; her proscription covers whatever "appeals to the visual rather than the imaginative".

The charge of exhibitionism is nowadays flung freely at that figure of fun, the elocutionist (who is not present to defend his -- or her -- views); usually with airy vagueness in lieu of corroborating details. Thus Kaucher (1945) conjures up a putative picture of the past, in which she sees

. . . the elocutionist, that eager, smiling creature trailing doubtful clouds of glory and a silken train, moving with correct and carefully learned grace to the platform. There she fluted her wild carol in what many must have hoped would be her death, with gestures mathematically taped off from her torso and a voice that rose and fell with the tide; if not with the meaning.

As is usually the case with such polemical tirades, this exaggerated account conveys considerably more about the attitude of the author than about the supposed object of her disdain. But it illustrates the virulence of the opprobrium attaching to any suggestion of deliberate cultivation of aesthetic features -- i.e. "exhibitionism". It is only a short step from this to proscription of deliberate structuring of any features in the rendition -- aesthetic, indexical, or linguistic -- which is a criterial attribute of the theory of acting and oral reading currently in vogue.

But references to aesthetic factors in the earlier literature do not support the linking of elocution with exhibitionism. Most of them indeed are generalized exhortations, as that of Holmes (1766; above) to

Vary your Tone just as your Subjects go,
 Cant not nor pitch your Voice too high or low,

.

Adorn with Tropes and Figures your Oration.

By Voice and Action grace Pronunciation.

Mason (1748) drops occasional remarks such as "*Emphasis* is raising the Voice, *Cadence* is falling it; and when rightly managed is very musical." One of the few sustained treatments of aesthetic features is that by Austin (1806):

In just articulation the words are not to be hurried over, nor precipitated syllable over syllable; nor as it were melted together into a mass of confusion: they should be neither abridged, nor prolonged; nor swallowed, nor forced, and, if I may so express myself, shot from the mouth; they should not be trailed, nor drawled, nor let to slip out carelessly, so as to drop unfinished. They are to be delivered out from the lips, as beautiful coins newly issued from the mint, deeply and accurately impressed, perfectly finished, neatly struck by the proper organs, distinct, sharp, in due succession, and of due weight.

This passage deals primarily with articulation of segmentals, although a prosodic variable is involved in the discussion of rate. The distinction between these types of linguistic feature -- prosodic versus segmental -- is quite important for synthetic elocution. Most of the problems in conventional speech synthesis, especially by rule, lie on the segmental side: the prosodic features for the most part map unidimensionally onto control parameters whose values are not especially critical -- the targets are large, and easily hit. On the other hand for the elocutionary transfer function as distinct from speech synthesis, where the problem is to extract these features from orthographic texts, the relative difficulties are reversed. That is, the orthography carries much fuller specification of the segmental features than of the prosodic ones. These differences

are appropriately reflected by separating the tasks of assigning two types of linguistic features in the next elaboration (below) of the reading machine.

The nature of the schism in elocutionary theory discussed above can now be characterized in terms of views taken of the relationship connecting rendition and text -- i.e. the elocutionary transfer function -- without becoming embroiled in issues of pedagogy; and with the *thought-feeling* dichotomy rewritten and clarified in terms of linguistic versus indexical features. There are two competing conceptions of how the phonetic, and especially prosodic, features of a proper rendition of a text are, or should be, assigned by an actor or interpreter; these conceptions may appropriately be referred to as the *direct*, versus the *oblique*, approach.

The early elocutionists attempted to provide formulas which could be applied to texts directly and deliberately to help map each text onto an appropriate oral rendition. Their rules were not offered as a substitute for understanding the text -- indeed, as will be seen below, they were not intended even to be taken literally. But they did provide guide lines for the O-P conversion process. What makes their rules seem often quaint and stilted is that they expected to uncover a universal "alphabet of emotions" -- indices onto which there existed a biunique mapping of affects. Burgh (1762) in particular assumed that "Nature has given to every emotion of the mind its proper outward expression, in such manner; that what suited one, cannot, by any means, be accommodated to another." Today no one seriously

expects this condition to prevail. Trager and Smith (1951) emphasize the fact that even true universals, like laughter, may be employed differently and with different meanings in various cultures. And the range of discriminable facial expressions, gestures, and paralinguistic cues is scarcely great enough to permit a one-to-one mapping even within a culture. The attempt to discern such a mapping is what makes Burgh's catalog seem ludicrously mechanical in the sense of unnatural (cf. Hargis 1957).

The currently ascendant theory was enunciated by Whately (1828) and holds that an oral performance depending in any way on direct application of rules for relating phonetic features to the text is bound in principle to be inferior (its harmony and unity destroyed). This oblique or impulsive school insists on an approach in which the meaning of the text is grasped (in some undefined sense) and an appropriate pronunciation derived by largely inexplicit and usually unconscious, semi-mystical means. Thus Whately says:

Impress but the mind fully with the sentiments, etc. to be uttered; withdraw the attention from the sound, and fix it on the sense; and nature, or habit, will spontaneously suggest the proper delivery.

and

A reader is sure to pay *too much* attention to his voice, not only if he pays *any at all*, but if he does not strenuously *labour to withdraw* his attention from it altogether.

Parrish notes that nowadays the term *mechanical* is generally applied to any reading "that is *deliberately designed* instead of being left to the speaker's thinking and feeling at the moment of utterance." His own view is that correct readings are not discovered

c merely by studying our own nature, by following our impulses, or encouraging individual self-expression, or . . . "revelation of the soul." Such a method says in effect that once you have the meaning of a sentence there is nothing to learn about elocution; there is nothing to study; all you need to do is let yourself go. It implies also that the teacher has nothing to teach.

Adherents of the oblique approach often attribute its origin to Stanislavski (1936, 1949, 1963), but this is misleading. His method was certainly oblique and mystical, but it applied only to emotions, i.e. to the assignment of affective indices, not to linguistic prosodic features, for which he offered no theory whatever -- only a number of hints about the importance of technique. The current "dramatistic," text-oriented slant in oral interpretation on the other hand, rejects this dichotomy and advocates a pedagogy the implicit O-P theory of which is "think-feel (experience) the thought-feeling, and the phonetic features will take care of themselves." Indeed the claim is even made that this theory represents an improvement in the way it relates text to rendition. Geiger (1962) for instance argues that

accepting the literary text as essentially a dramatic form of discourse, [and stressing] its experiential relations with the life it represents and with the oral interpreter who expresses it . . . permits us to trace in a single continuous movement, from text through public reader and on to his auditor, organic relations of literature and its understanding to oral interpretation. . . . this dramatic or experiential emphasis in analysis and terminology, more clearly and efficiently than earlier theory, relates what the text is to what the interpreter does.

The clarity and efficiency of this terminology for characterizing the relation between text and rendition may be illustrated by Geiger's passage contrasting two types of instructor. The old fashioned one, for whom Geiger expresses tepid tolerance, thinks about his student as

follows:

"If only he will stop gargling his words, if only he will talk a little more like an articulate, intelligent young adult, I will have taught well. And if in this course he stops slouching up to the rostrum and twitching in front of the class, so much the better."

But the modern instructor, with whom Geiger obviously identifies, thinks

"If only the poem, in its rise and fall in my student's voice, will claim him, if even for a moment, he will see a new world, or an old world freshly, and we will both be happy." This instructor wants his student to be possessed by, and so to possess, the literary text. He wants the student during his interpretation to have a little seizure, a seizure of the sort and quality appropriate to the artful drama of the text.

And Geiger adds that "the liberal arts college oral interpretation teacher at least, remembering again the limitations under which he must operate, will, if he must choose, simply hope that the slouch will disappear with the text-directed dramatic seizure."

Obviously such a view of the orthographic-to-phonetic conversion process has little to offer toward the practical design of a reading machine. It is no use instructing a computer to "think the thought" or to "open to itself the situations in literature," or to cultivate a "vivid realization of meaning at the moment of utterance." (cf. Sloan 1966). Nor does this school of oral reading offer any useful insights for a theory of language. The rules governing the output of oral reading are just a subset of the rules by which speech is ordinarily generated, with the addition of a few rather trivial special rules to cover some purely or primarily literary phenomena -- such as medial and final ascriptions (to use Bolinger's term for quotation-attribution clauses such as *he said*). But most people somehow pick up a set of

"pseudo-rules" for reading aloud -- rules which say that whenever confronted by a printed text one should shift linguistic gears and start talking a singsong, isosyllabic accent of English in which functors are emphasized, anaphoric relations ignored, and rheme treated exactly like theme.

This style of reading aloud is characterized by Pike in his undeservedly neglected step by step procedure for marking limited intonation with its related features of stress, pause, and rhythm (1945b; "reading style" being his more or less pejorative term for such unskilled O-P conversion and "conversational style" representing the desideratum):

. . . a conversational style is characterized by few centers of special attention and by many repressed lexical stresses, while a reading style is characterized by many retained lexical stresses

A conversational style has relatively few intonation curves, a reading style has many intonation curves.

Sheridan (1787 [1762]) remarks that

the chief reason, of this general abuse of emphasis, seems to be, that children are taught to read sentences, which they do not understand; and as it is impossible to lay the emphasis right, without perfectly comprehending the meaning of what one reads, they get a habit either of reading in a monotone, or if they attempt to distinguish one word from the rest, as the emphasis falls at random, the sense is usually perverted, or changed into nonsense.

A program for helping students unlearn these pseudo-rules -- by, say, "thinking the thought" -- is of course a useful art. But there is nothing very profound about it. It merely clears the way for native competence in the language to assert itself; it offers nothing to help a foreigner learn to read aloud well, or a member of the culturally disadvantaged, or anyone who is not already a native speaker of

the standard language -- including, and in particular, a computer. Indeed it is evident that in terms of the oblique theory of elocution there is little to say at all about the act of oral reading ("giving the meaning"), and it is not surprising to find that theoretical discussions in the field have, as Hargis (1960) notes, gravitated heavily toward the art of literary criticism ("getting the meaning").

The present study uses *elocution* as a neutral term for the process of assigning appropriate pronunciations (in the classical sense of *pronuntiatio*: subsuming prosodic features as well as segmentals). That elocution is rule-governed behavior follows trivially by definition from its universally admitted classification as a form of communication. The only questions in this connection which merit discussion are: (1) to what extent can the transfer function relating orthographic input to phonetic output, as for a reading machine, be formalized; (2) what implications does this conversion process have for linguistic theory; and (3) what bearing does it have on elocutionary (a) theory and (b) pedagogy? Subpoints (3a) and (b) have, as discussed above, often been conflated, with the result that the issue of whether regularities exist, and can be abstracted and formalized, has not been treated separately from that of whether the presentation of such rules is the best way to teach pupils to read aloud well. But quite clearly *a theory of elocution can only comprise statements explicitly relating orthographic texts to their phonetic renditions, viz. rules.* The pedagogic efficacy of the rules is irrelevant to the theory.

From this point of view it is clear that the early elocutionists were well justified in seeking to define rules for orthographic-to-

phonetic conversion, and the obloquy currently directed against such undertakings is simply a mistake. But it must not be supposed, just because earlier elocutionists are often traduced with being too mechanical, that the rules they offered constituted, or were intended to be, really explicit mechanical procedures or algorithms. A good example is the very old rule for "observing the stops" -- i.e. assigning pauses from punctuation marks:

A Comma stops the Voice while we may privately tell
one, a Semi-colon *two*; a Colon *three*: and a Period *four*.

This is offered by Mason (1748); Haberman (1954), evidently regarding Mason as its inventor, remarks "Although he knows that the best advice is to 'make the Ideas seem to come from the Heart,' he cannot avoid the prescriptive rules which became a commonplace in the elocutionary movement" -- and quotes the above as an example. But in fact the rule was lifted verbatim from the grammar of Brightland (1711) and ultimately derives from the seventeenth century if not earlier. Simon Daines (1640; see also Fries 1925) has a rule differing only trivially:

The *Comma-colon* [i.e. semicolon] . . . [requires] the
time of pause about double that of the *comma*
generally, which yet is very small.

The *Colon* . . . exacts half the pause of a *Period*; and
halfe as much againe as a *Comma Colon*.

The *Period* . . . requireth double the time of pause
. . . [of] a *Colon* . . .

He anticipates Mason also in assigning twice as much pause between paragraphs as between sentences in the same paragraph.

This rule was repeated, with minor variations, throughout the elocutionary movement; nor was it confined to England. Woehl (1927) in "Goethe's Rules for Actors" [1803], cites evidence that Goethe

demanded a pause for each of these marks, and denoted their varying lengths graphically in this fashion:

—, —; — —! — — —? — — — — —.

The rule looks explicit enough if one doesn't quibble about establishing a time-constant for the private counting. But Mason offers the rule in the context of a remark that "you will in a good measure in reading be directed by the Points: but not perfectly; for there are but few Books that are exactly pointed", and adds

After all, there is so much License admitted, and so much Irregularity introduced, into the modern Method of Punctuation, that it is become a very imperfect Rule to direct a just Pronunciation. The Pauses therefore, as well as the Variations of the Voice, must be chiefly regulated by a careful Attention to the Sense and Importance of the Subject.

These passages are of course not quoted by Haberman (1954), who, in ignoring these important qualifications (and failing to recognize the traditional character) of the rule, exhibits a characteristic syndrome of modern writers schooled in the oblique, impulsive Whately tradition: any evidence tending to countervail or ameliorate the stereotyped image of the elocutionists (as wishing to substitute mechanical rules for an understanding of the text) seems to strike them completely on the blind side. Two other examples, among many which might be adduced, are a remark by Woehl (1927) in his introduction to Goethe's rules for actors:

it will be noted that they are, in general, mechanical and arbitrary

-- which is followed by a collocation of "rules" of a character typified by

28. The declaimer is free to select his own stops, pauses, and so forth; but he must guard against destroying the true meaning

-- and a completely inexplicable reference by Hale (1941), in discussing the work of James Rush (1827), to "his labored listing of inflexible patterns"; whereas any rules in the welter of notions offered by Rush in fact are so flexible, ad hoc, and indefinite as scarcely to qualify for the name.

Apropos of the punctuation-to-pause "algorithm" and the caveats with which its advocacy has typically been tempered, Goold Brown (1873) remarks that

children are often told to pause at a comma while they might count *one*; at a semicolon, *one, two*; at a colon, *one, two, three*; at a period, *one, two, three, four*. This may be of some use, as teaching them to observe their stops, that they may catch the sense; but the standard itself is variable, and so are the times which good sense gives to the points.

No rule in the elocutionary literature has been more widely reiterated, and none is closer to the explicit and mechanical nature of an algorithm, than this one. It seems likely that speech actually synthesized by this recipe would sound somewhat anomalous; however, the means now exist for testing elocutionary claims empirically, and such experimentation would be preferable to armchair speculation and namecalling as a basis for rejecting (or accepting or modifying) them. But the remarks quoted above show that the grammarians and elocutionists who repeated the rule never intended it to be taken literally nor applied mechanically. They were not writing algorithms for computers but pedagogical helps for human pupils. The status of their "rules" was comparable to that of "imitation labels" in phonetics (Pike 1943, Abercrombie 1967) -- i.e. they served as impressionistic descriptions of phenomena learned by imitation (overt or otherwise), which could then

be discussed and elicited in terms of their conventional labels, but for the production of which the labels (or rules) provided no sufficient recipe. Thus while it is possible to submit the rules of earlier elocutionists to the acid test of synthesizing speech according to a literal interpretation of them (and this should be done) the results may well prove only what anyone who has an elementary acquaintance with computer programming already knows: that humans are not very good at being consistently precise and rigorous and explicit (i.e. mechanical), especially when strong motivation is absent, as when communicating with other humans.

A number of quotations have been given above which stressed the importance of understanding the text. Similarly, Beattie (1788) observed:

One of the greatest niceties in the art of reading is the right application of emphasis. And of this they only are capable, who perfectly understand what they read, and attend to the full import of every clause, and of every word. If we read without understanding, or without attention, we continually misapply the emphasis; and the hearer, if he is not very acute, must often mistake the sense.

Despite reiteration, throughout the span of elocutionary thought, of such warnings, the task of obtaining explicit commands for controlling a synthesizer from orthographic texts might appear from a linguistic point of view to be a low-level one involving merely transliteration from graphemes into phonemes, i.e. from one form of derived phrase marker or surface structure to another. Such would be the case if English orthography uniquely transcribed each significant phonological unit. But this is not true even for segmental units in English, far less the prosodic ones. The theoretical interest of orthographic-to-

phonetic conversion lies partly in the fact that, given a defective or incomplete representation, a human or mechanical reader can generate a correct phonetic rendition only if the textual ambiguities are resolved through a sophisticated syntactic and semantic analysis amounting to an understanding of the text.

Morton (1964) has shown experimentally that when reading aloud subjects do in fact analyze the written message to a fairly deep level and then "resynthesize" it as speech. Thus oral reading provides one of the few cases of a high-level language processing activity in which both the input and the output are accessible. It is therefore of particular interest, because a fundamental theorem of cybernetics holds that if the input and output of a system are known, its transfer characteristics can be deduced.

In this respect oral reading may be compared with translation of texts from one language to another. Both tasks require "understanding" of the input text and are thus formidable ones to simulate mechanically. But the constraints on the respective outputs -- the extent to which their form is specified by the input -- is greater for the reader than the translator. The reading machine may therefore be expected to provide more insights than machine translation into the workings of the black box of "competence" -- i.e. the native speaker's tacit knowledge of his language (Chomsky 1965).

The requirement for machine translation that the computer must "understand" the text was not recognized until rather recently. It seems to have come as a surprise to the early enthusiasts for machine translation (MT) that more was involved in translating from one tongue

to another than having a computer look up each input word in a bilingual dictionary and print out the target-language equivalent. In reaction to the disappointment which inevitably followed the initial sanguine optimism, various experts began taking gloomy views of the potential of computers to do the task at all (Taube 1961). Most of the examples which purport to prove the impracticability of MT -- such as the well known, if rather puerile, case of *the box is in the pen* (Bar-Hillel 1959) -- happen to involve full homonymy: i.e. the *pens* of *playpen* and *fountain pen* are homophones as well as homographs. There is therefore no immediately apparent problem of orthographic-to-phonetic conversion within English, whereas it would be a most unlikely coincidence if these two concepts could be correctly represented by the same gloss in any given target language.

But equally well, certain homographic forms of English such as *read* [rɪd] (present) versus *read* [rɛd] (past), which have to be discriminated for a correct oral rendition in English, might be translated identically into target languages having a different tense structure, such as Hawaiian. The differences between the two types of difficulty is merely fortuitous. No workable translation program can be based on the expectation that English homographs will be homographic or homonymous also in the target language, nor can any high-performance reading machine treat English homographs as automatically homophonous. Either type of translator must in principle be able to disambiguate both lexical and constructional homographs -- viz. to "understand" the text. Whether mechanical procedures can be developed for doing this is an open question. But a solution of the MT problem will imply solution

of the orthographic-to-phonetic conversion problem and probably vice versa.

As suggested earlier, the problem of assigning linguistic features may be dichotomized into the more or less discrete tasks of (a) assigning segmental features or phonemes to words, and (b) assigning prosodic features to sentences. It is convenient to include "word stress" in the first category. The task of assigning appropriate segmental pronunciations and "innate" lexical stresses (Pike 1945a) to the words of the input text may be subsumed under the traditional term *orthoepy*. The term *prosody* will be used in a corresponding fashion to designate the assignment of prosodic features -- "sentence stress" or accent, intonation, and various indexical phenomena whose treatment will be postponed for the present as not essential for straight factual prose. Furthermore, the tasks of syntactic and semantic analysis of the input text can be represented as separate steps preceding the orthoepy and prosody. These elaborations are shown in figure 1.6.

Most work in orthographic-to-phonetic conversion of English thus far has dealt solely with the orthoepic level (Bhimani and Dolby 1966, Venezky 1966, Lee 1967, Gaitenby 1967). Thus Venezky (1966) in discussing "Automatic spelling-to-sound conversion" states:

Attention will be placed upon spelling-to-sound relationships in isolated words, without considering inter-word alterations and intonation contours. While the "phonetically pleasing" reading machine must handle these problems, the basic problems are met in dealing with isolated words and must be attended to first.

Pronunciations may be assigned to words by algorithms, dictionary lookup, or a judicious combination. Lee (1967) has illustrated some of the difficulties confronting the use of rules or algorithms: e.g. the

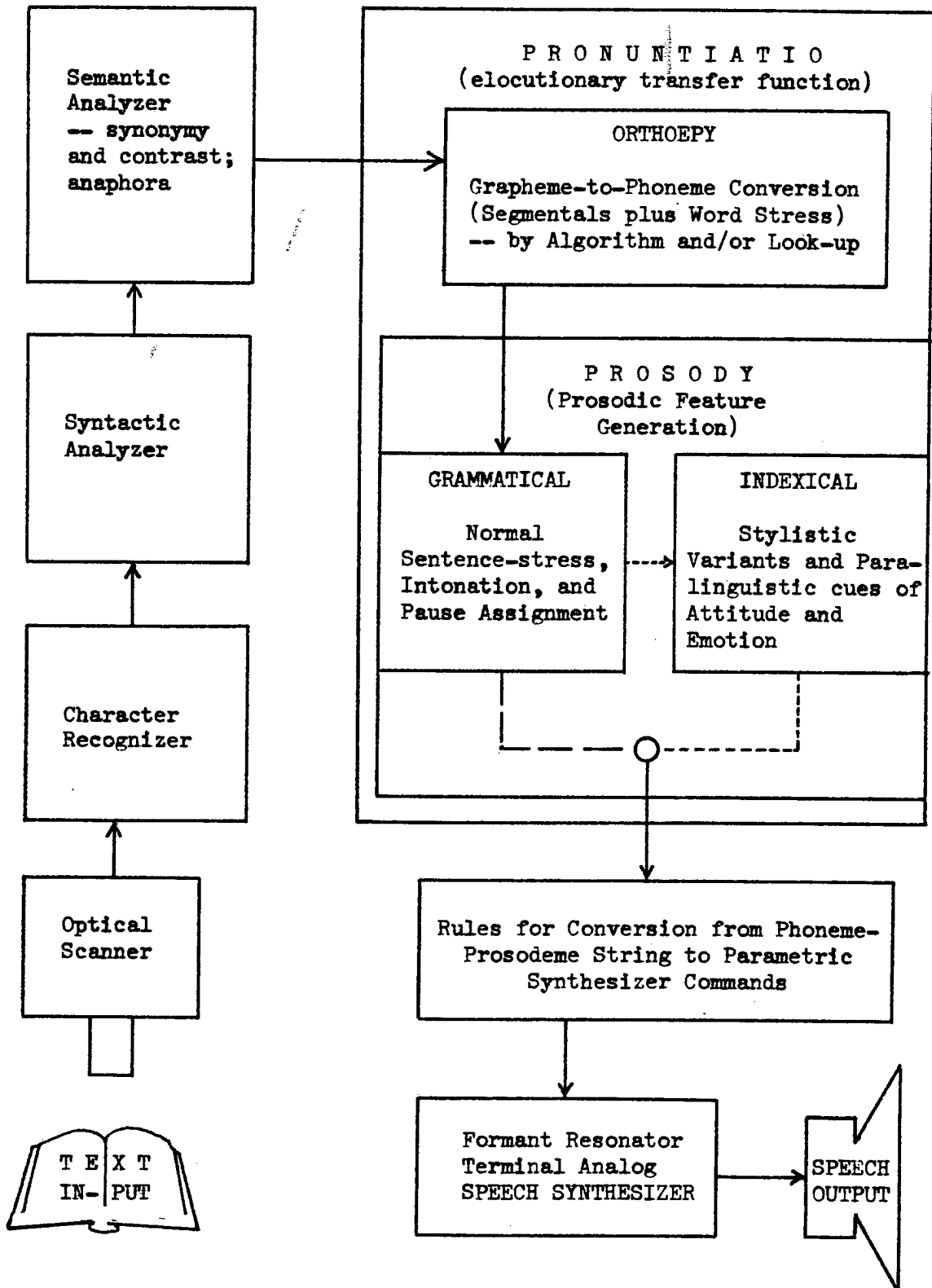


Figure 1.6 Block diagram of O-P converter

final *e* of *epitome* and many other words (*apocope*, *syncope*, *macrame*) represents a separate syllable and is not a silent marker as in words like *dome*, *hope*, *frame*. But there are problems with dictionary lookup too. It presupposes access to a very large dictionary unless the range of input texts is to be severely restricted, or unlisted words are to be spelled out in the output, or other compromise made.

With either rules or lookup there is a rather large residue of problems due to homography which can only be resolved by syntactic and semantic analysis of the text. Examples of pairs which may seem trivial but whose disambiguation in oral reading in fact calls upon very sophisticated skills of parsing and pattern recognition are *lead* [lɛd] metal versus *lead* [lɪd] wire and *bow* [bau] of a ship versus *bow* [bou] of an archer, etc.

The question of whether satisfactory orthoepic algorithms can be devised to eliminate the need for dictionary lookup is not directly relevant to this study, which primarily concerns the prosodic transfer function. The answer for a practical application will depend largely on the willingness of users to accept mispronunciations of irregular words (Venezky 1966, Lee 1967, Gaitenby 1967). And in the reading machine here envisaged, access to a large dictionary is independently required for the syntactic and semantic analyses needed for prosodic feature assignment.

As a result of the recent resurgence of interest in orthoepy in the form of phonological rules for word stress and pronunciation, which has been occasioned by the work of the M.I.T. group, especially Chomsky and Halle (forthcoming), it may seem tempting to suppose that

homographs and other exceptions can be handled by listing them, or appealing to "underlying forms" etc., and that thus the problems disappear. But it should be noted that the Chomsky-Halle (forthcoming) orthoepy does not start from spelled forms and therefore would not eliminate lexical lookup. Derivative forms can (usually) be satisfactorily predicted from their base forms, but base forms would have to be looked up; they do not in general coincide exactly with the orthography. (There is no reason why they should, of course, since Chomsky and Halle are not concerned with practical problems of O-P conversion.)

For present purposes it is quite irrelevant whether the orthoepy employs rules or lexical lookup or both, so long as the output which plugs into the prosody carries suitable representations of the orthoepic information ultimately needed for a correct phonetic rendition of the text. The prosodic model incorporated in the next chapter treats:

- (a) the units which have to be specified at the input to the prosody;
- (b) the output units, convertible by low level rules to sounds; and (c)

the paradigm relations between the two sets -- i.e. the way in which input units are rewritten into output units under various contextual conditions, and particularly by the context-free rules which yield "citation-form" pronunciations. Many rules of prosody will come into the discussion by way of showing the motivation for some of the features. The chapter closes with a presentation of a fragment of the prosody, illustrating its operation on a sample text.

While phonologists agree fairly well about the inventory of segmental phonemes of . . . English, they do not agree about the prosodic features: facts, theory and terminology are still very much in question.

-- Ignatius G. Mattingly (1966)

II

TOWARDS A PROSODIC MODEL OF AMERICAN ENGLISH

As has been suggested in the previous chapter, one way of partitioning the task of prosodic feature generation in an elocutionary model of English such that attention is centered on the unmarked assignment of grammatically relevant prosodic units to the exclusion of indexical features is to address only the task of orthographic-to-phonetic (O-P) conversion of straight factual prose as opposed to fictive texts. This can be conveniently subdivided further into a consideration of the output units to be assigned, the information required at the input to enable correct specification of these units, and the rules or transfer function which will generate the former in terms of the latter.

A conceptual version of a reading machine, as discussed above, could have a character recognizer at the input and a speech synthesizer at the output. Between these, in a more or less serial order, would have to be programs for syntactic and semantic analysis, word

pronunciation, and prosodic feature assignment.

For present purposes it is convenient to start from the point where some of these tasks have been done -- such as character recognition, parsing, semantic analysis, and grapheme-to-phoneme conversion -- so that the input comprises a concatenation of words with their pronunciations (including word stress) assigned, and with discrete signals for punctuation marks, italics, new paragraphs, etc; and, along with this text, categorizations indicating the syntactic and semantic relations. It is true that none of these steps can be done really well at present by automatic procedures. However, the point here is not how to build a practical, workable reading machine, but rather how -- given ideally efficacious procedures for the tasks just listed -- to get from there to the point where conventional speech synthesis by rule takes over (i.e. to a string of phonemic symbols with prosodemic markers incorporated) in such a way that the speech output is prosodically as well as orthoepically correct (at least for the restricted class of input texts here considered, i.e. straight factual prose). The aim of this chapter is (1) to consider what prosodic patterns are required for this minimal model (with occasional glances at variations or subcategories which a fuller theory, i.e. one capable of dealing with expressive features, would need to explain) and (2) to show how the input information can be structured so as to maximize the load borne by the rules of the transfer function, or *prosody* as that term is used here, in securing the context-appropriate variations of prosodic patterns, thus minimizing the need for such variations to be already specified in the input. Ideally the input to

the prosody should be the sequence of text words rewritten (represented) in what will be referred to as their *canonical* orthoepic form (transcription of segmentals and word stress as in a pronouncing dictionary), with punctuation, paragraphing, and italic (etc.) type distinctions also encoded.

The notion that words (and certain phrases) have an abstract "canonical" orthoepic form -- a simplest representation of their vowels, consonants, syllables, and stress patterns from which all the phonic variations produced by context-sensitive rules of the prosody can be predicted -- may require some clarification, since it is not explicitly recognized in most discussions of phonetic theory. Such a conception of stress is evident in the terms "innate stress" and "repressed lexical stresses" employed by Pike (1945a, 1945b), and in the definition of stress by Bolinger (1958, 1965) as the "potential for pitch accent." Kurath (1964) states it quite explicitly:

Words and phrases as such have inherent stress patterns, as *fáther*, *foréet*, *bláckbird* and *black bírd*, *gét up*, *not ríght*. The precise manifestation of these inherent patterns is determined by the prosodic structure of the sentence -- the higher structural unit of speech.

The concept of canonical orthoepic forms is also related to -- but not coterminous with -- that of the "underlying representations" postulated by Chomsky and Halle (forthcoming). This relationship will be discussed below; briefly it may be said that such words as *¹tele,graph* and *te¹legraphy*, which include the same underlying forms in the Chomsky-Halle theory, have differing canonical forms (corresponding to their transcriptions in pronouncing dictionaries). The canonical forms may be thought of as a level of abstraction between

each underlying base form and its concrete phonic manifestations which show not only the derivational changes, but also phonetic variations depending on location of sentence accent, direction of intonation, etc. that are ignored in the canonical representations. Another view would be that the canonical form, which can be equated with the ideal entry in a pronouncing dictionary, may correspond most closely to the way the pronunciation of a word is stored in the minds of native speaker-hearers. This, while perhaps ultimately an empirical question, can only be speculative in the present state of knowledge, and even if true would not thereby challenge the still more abstract characterization of the speaker-hearer's internalized system sought by Chomsky and Halle. But it would be an interesting and important fact about "the system of rules represented in the speaker-hearer's mind, normally acquired in early childhood, and used in the production and interpretation of speech signals."

A good deal of confusion arises from failure to distinguish, in discussions of phonetic transcriptions, between two quite different functions which they serve: (1) the specification of events at the level of phonic substance, as in transcribing particular *tokens* of words or utterances, and (2) the recording of the abstract canonical orthoepic form of word (or utterance) *types*, as in a pronouncing dictionary (where certain idiomatic phrases and clichés have to be entered with canonical forms just like words). This is not merely a matter of descriptive versus prescriptive uses, for the pronunciations in respectable dictionaries are drawn from observations of the usage of educated native speakers and are thus "descriptive". Thus Jones

(1956) states that his entries are "to be taken as expressions of opinion by the author" and mean that "the word . . . is, in the opinion of the author, generally pronounced [as transcribed] in ordinary conversation by the people referred to . . ." (i.e. speakers of RP).

Abercrombie (1967) observes that

Strictly speaking, phonetic transcription records not an utterance but an analysis of an utterance, and the analysis may be of two basically different kinds: it may be in general phonetic terms [impressionistic], or it may be in phonological terms [systematic].

Elsewhere (1964) he subdivides systematic transcriptions by two independent distinctions which allow more precision than the old *broad-narrow* dichotomy: *simple* versus *comparative* symbols; *phonemic* versus *allophonic* classifications. The two separate descriptive functions of systematic phonetic notation postulated here -- the canonical and the phonic -- are not coterminous with either of these interpretations of the "broad-narrow" dichotomy. With respect to segmentals it is perhaps more usual for canonical transcriptions to be broad in the sense of simple and phonemic (in the pre-Chomsky sense of phonemic), but it is by no means clear that phonic transcriptions show any corresponding tendency to be narrow in the sense of comparative and allophonic: this depends on the purpose and audience for which they are intended. In the case of prosodic features, however, this choice is not even available: the stress marks sanctioned by the IPA, for example, afford machinery suitable only for *broad* transcription with *simple* symbols (by any intuitive standard, though they are scarcely "romantic") of a *systematic* analysis.

The conventional analysis of stress (viewed as phonic substance)

in terms of three levels or degrees will be examined and reinterpreted below; it turns out to be, systematically, quite adequate to characterize the *input* to the prosody, and it is therefore ideal for use in canonical transcriptions, as in a dictionary. It is not a satisfactory way of representing the phonic output of real utterances -- though these are of course what the phonetician-lexicographer must use as his data for inferring the canonical forms. The phonic realizations of a word as it occurs in running conversation regularly manifest obligatory prosodic influences of its context via rules of the prosody which it is the aim of this chapter to exemplify. In addition they manifest other prosodic modifications functioning as indexical features of the speaker's group membership (relevant to whether his pronunciation should be counted at all), individual idiosyncrasies, and, often, of physical and mental states, including attitudes and emotions (cf. chapter I). Quite clearly the phonetician-lexicographer, catching such tokens on the fly (or even via tape recording) has to abstract away from these various perturbations in the phonic data; he has to distill out the quintessential pattern underlying the diverse phonic renderings and represent it in, usually, a single entry. It is his knowledge of the prosody which enables him to discount the effects of its context-sensitive rules so as to enter alternate pronunciations only when real divided usage exists as to the canonical form of a word's stress pattern (or segmental representation). One way around the difficulty is to ask, e.g., "how do you pronounce the word spelled f-e-t-e," so that the answer will be a "citation pronunciation," which as will be seen, manifests the canonical form most perspicuously. The abstract nature

d of canonical orthoepic representation in relation to phonic substance,
although of course tacitly taken into account by every lexicographer,
has not usually been explicitly noted. Accordingly it is not surpris-
ry to find some contamination of the class of strictly lexicographic
divided usage by the inclusion of prosodic variations. Thus Jones
s (1956) frequently adds notes to pronunciations such as that at
l 'un'leavened indicating it may be 'un,leavened "when attributive" al-
ry though he elsewhere (1964) shows this to be a quite regular phenomenon.

is The distinction between the canonical and phonic functions of
er systematic transcriptions would be more obvious, of course, if the use
of notation types were correlated with it in a rational way; e.g. if
IPA stress marks were reserved (as they will be in this work) for
oe transcription of canonical forms; and other indicators, such as accent
marks, used for spoken forms.

Unfortunately, no such distinction is normally followed in the use
of these symbols. In broad (phonemic) IPA transcriptions of connected
texts, whether prescriptive (as for teaching English as a foreign
language) or phonic, "stress" is quite often left unmarked, to be pre-
e dicted by the user. Since the features subsumed in the stress marks
si- are by no means regularly redundant, this practice imposes a rather
complex task of disambiguation on the reader -- comparable to, though
different from, that imposed by the orthographic text itself -- a task
requiring considerable tacit knowledge of the language, or competence.
On the other hand it is usual in connected texts transcribed in even
the narrowest of comparative allophonic notations to find prosodic fea-
e tures only vaguely suggested or hinted at by the use of broad stress

notation; thus they are represented as passed through a rough filter -- or, in another metaphor, at a low level of delicacy. The particular prosodic form of the utterance transcribed -- whether the pitch went upward or downward on accented syllables, for example -- is not uniquely recoverable from the transcription. It is quite true that a fair reconstruction of grammatically relevant features is possible by a native speaker of the dialect transcribed; on the present interpretation this merely means that the reader of the transcription knows the rules of the prosody and treats the stress marks as input to them. (The need for syntactic and semantic information is probably not altogether obviated.)

What is needed is a systematic transcription (i.e. analysis, with appropriate notation) which represents the structurally patterned prosodic units in the output as adequately as IPA stress marks represent, in the canonical orthoepic "stored forms" of words and idioms at the input, the minimal information needed for correct assignment of these prosodic elements. Many attempts have been made to provide a satisfactory system for transcribing prosodic features such as accent, emphasis, terminal pitch contours, and pauses -- too many, in fact, to allow systematic review of them here. For various reasons none is entirely satisfactory as the prosodic model for O-P conversion of English, but a number of them have furnished important insights which are utilized in the highly eclectic model proposed in this work.

One of the most influential works in this respect is that of Chomsky and Halle (forthcoming) even though the authors omit pitch and intonation from consideration and assume (while keeping "an open mind

on this matter") that such gaps have no serious bearing on the questions they wish to deal with. The analogy was pointed out above between the attempt in the present study to relate prosodic output features to abstract canonical orthoepic forms and that of Chomsky and Halle in devising rules of phonological interpretation which produce some of the variations found in lexical items and their derivative forms in surface structures from the information contained in postulated abstract and invariant underlying lexical representations of the base forms.

Both the lexical representations postulated by Chomsky and Halle and the canonical forms of words assumed in this study are abstract in the clear sense that they relate to the phonic signal only indirectly, through the medium of rules that apply in ways determined by the intrinsic representation of a word and the context in which it appears in a surface structure. But it should be noted that the *output* of the Chomsky-Halle lexical stress cycle and phonological rules for vowel alternation corresponds approximately to the *input* to the prosody as viewed here -- i.e. this part of their phonological rules provides in general a good description of the canonical orthoepic forms of the various derivatives of basic lexical items. In this sense it can be viewed as an orthoepy; it is certainly elegant and explanatory in comparison with any lookup scheme. It is of course unlikely (nor is it intended) to be of practical use in O-P conversion, since it does not provide a way of mapping the spelled forms onto the underlying lexical representations.

Thus despite some similarities the phonological rules of Chomsky

and Halle and the prosodic rules of interest here deal with different phenomena and at different levels of abstraction. The present study concerns itself with the prosody and assumes canonical forms to be already assigned at the input. Chomsky and Halle on the other hand regard most prosodic variations as outside the domain of their concern. It is true that they attempt to subsume within the stress cycle certain aspects of the prosodic manifestations of phrases and whole sentences. But this part of their enterprise is considerably less successful than the insightful "orthoepic" rules by which they predict canonical stress and vowel variations at the level of the word.

Even at the word level Chomsky and Halle's rules provide more structure than will be found to be needed as input to the prosody (rules for accent and intonation assignment). Indeed Chomsky and Halle themselves propose to limit the number of surface stress distinctions by low level rules whose function is to delete structure assigned by the stress cycle in excess of that needed for the phonetic output. But it is because they view the output of their stress rules as separated from the phonetic level only by low level rules that they assign so much structure in the first place. And that is a view which they can only maintain by steadfastly refusing to recognize as belonging to the domain of linguistic description most of the processes involved in prosodic feature assignment -- viz. all of the processes triggered by semantic relations and suprasentential structural features, or in short most of the prosodic processes having any theoretical interest.

This refusal was explicit and outspoken in Chomsky, Halle and Lukoff (1956), where

we have specifically excluded from consideration all forms of expressive stress, including contrastive stress. In language, expressive elements are deviations from the normal pattern. The possibility for such deviation is, of course, enormous, and almost any stress arrangement can occur under special circumstances. Therefore, if this distinction between normal and expressive stress is not made at the very outset . . . any hope for a systematic account has to be abandoned.

. . . as a consequence . . . we do not provide for the normal stress patterns of such utterances as "This is the brown house, not the white one," where there is extra heavy stress and extra high pitch on "brown" and "white." . . . We feel that these utterances are best regarded as being in a special sense deviations from the normal pattern

The ex cathedra tone of this dismissal (but not the rampant confusion in the application of the term *normal*) is somewhat ameliorated by the accompanying assertions that "the description of such utterances poses many problems which have never been adequately handled" and that "a satisfactory description of them will require the development of methods not currently in use in phonemics." These statements are indeed unarguable.

A certain mellowing may be perceived in Chomsky (1965) where after discussing at some length the notions of "logical" or "psychological" (as opposed to grammatical) subject and predicate, as in the sentence *glass is eLAsTic* versus *GLASS is elastic*, he concludes mildly:

Whatever the force of such observations may be, it seems that they lie beyond the scope of any existing theory of language structure or language use.

. . . the syntactic and semantic structure of natural languages evidently offers many mysteries, both of fact and of principle, and . . . any attempt to delimit the boundaries of these domains must certainly be quite tentative.

Further concessions may confidently be anticipated from Chomsky and Halle (forthcoming), where

[among] the phenomena not accounted for are those involving obligatory contrastive stress (sometimes, stress shift) as determined by syntactic parallelism, as in such sentences as *he wanted to study electrical rather than civil engineering, or instead of encouraging the teacher to make the work interesting, the school administrators actually DIScourage her*, etc. Many other problems can be cited, all indicating that many questions of fact and, perhaps, of principle still remain unresolved, in this area.

The examples cited in these discussions illustrate precisely the phenomena which an adequate prosodic model must account for even if it is restricted (for the present) to handling factual prose. It should be noted that none of them show features which can be called "expressive" in the sense of indexical, and certainly they are not "paralinguistic". A grammar which does not account for them is simply incomplete.

In attempting to sketch a prosodic model to fill this gap it is convenient to regard "stress" not as a phonetically real or pronounceable entity, but rather as an aspect of the canonical orthoepic form of words input to the prosody that is assigned to syllables of a text by the orthoepy. (As discussed in the preceding chapter, it is unimportant whether word stress and pronunciation is handled by algorithm or by lookup or both, so long as appropriate units are correctly assigned.) Once the canonical or "stored" form (stress and segmentals) is assigned to each word in the sequence constituting the input text, all remaining prosodic variations -- sentence accents, emphases, rhythmic variations, intonations -- are taken care of by the prosody. (This is a slight oversimplification: certain phenomena such as nominal compounds and idiomatic phrases may be more conveniently regarded as belonging to the domain of orthoepy; this of course merely sidesteps

some thorny but quite obviously grammatical problems.) On this view "stress" is not a phonic reality at all, nor can it be mapped onto the speech signal by low level rules. Rather, it is regarded for the purpose of this work as part of the abstract orthoepic representation of the canonical or "innate" forms of words. Such representations are pronounceable only in the sense that, subject to certain conditions, there is a virtually vacuous path thru the prosody, comprising context-free rules which treat the word as an independent sense group and assign what is called "citation-form" pronunciation. Thus a subject speaking a word list typically employs phonic output units which are related simply and biuniquely to the canonical orthoepic representations of the words. But even in hypostatic uses the phonic realization of words is subject to prosodic variation by context sensitive rule -- as when one corrects (or comments on) the pronunciation of normally unaccented syllables: "he said hypostaTIZE, not hypostaSIZE" (where the canonical forms are both forestressed).

The present study treats both syllabicity and stress as higher-level units specified in the input to the prosodic rules and having no invariant phonetic correlates in the output. The canonical orthoepic forms of words contain representations of their abstract stress patterns -- including which syllables have potential for pitch accent. Following Bolinger, the term *accent* has been adopted here to designate the output unit, a pitch obtrusion, which expounds canonically full-stressed syllables whenever this potential is realized.

The overriding importance of pitch (voice fundamental frequency) in determining the perception of what is usually called "stress" (or

"sentence stress") was shown even before Bolinger's theory of pitch accent in English (1958) by an unpublished experiment utilizing the Parametric Artificial Talking Device (PAT) developed at the British Ministry of Supply, Signals Research and Development Establishment (Lawrence, 1952). The advantage of synthetic speech is of course that it can be precisely controlled in all respects (cf. Chapter I). In the experiment in question, Lawrence synthesized several versions of the sentence "what did you say before that?" The only difference between the versions was in the F_0 (fundamental frequency) or "pitch" parameter. The rhythm and amplitudes, as well as all the strictly segmental parameters, were repeated exactly for each version. The synthesized sentences may be shown in an iconic transcription as follows:

(1) What did you SAY be fore that

(2) What did you say be^{FOR}E that

(3) What did you say be foreTH A
T

Although neither the durations nor the amplitudes of the three words *say*, *before*, and *that* were allowed to vary, there is a very convincing shift of the sentence stress -- or as it will be called here, following Bolinger (1958, 1965) *accent*. This term is appropriate because the perceived prominence is solely due to changes in the pitch parameter.

Two things are to be noted about the aspects of the F_0 contour which the three versions share: (1) there is an upward pitch obtrusion reaching its maximum during the vocalic portion of the accented syllable; (2) there is a pitch subsidence throughout the remainder of the sentence which takes a roughly exponential form -- i.e. falling rapidly at first and then asymptotically approaching a low value. The latter is of course the often noted "nuclear fall" or glide. This subsidence of pitch from the nuclear accent (i.e. the sole accent in these cases but always the rightmost one) to the end of the sense group will be referred to in this work as *cadence*, following Mason (1748), who observes:

This is directly opposite to *Emphasis*. *Emphasis* is raising the Voice, *Cadence* is falling it . . . that is . . . the Sense being almost expressed and perfectly discerned by the Reader, the remaining Words (which are only necessary to compleat the Period) gently fall of themselves without any emphatical Word among them.

It should be noted that Mason and others in his period (Beattie 1788) used *emphasis* for what is here called *accent*. They, like a number of modern writers (Gunter 1966, Lieberman 1967) fail to distinguish between normal "sentence stress" (accent) and the additional degree of prominence (emphasis) that is most neutrally realized as "overhigh" or pitch level /4/ in the Trager-Smith (1951) system, or Pike's pitch level one, and the most clearly grammatical role of which is as the exponent of antithesis. Thus Mattingly (1966) in explaining certain deliberate simplifications in his model says:

There is more than one degree of prominence in natural speech; our model allows only one. Moreover, our model makes no use of intensity as an acoustical correlate of prominence. Fry (1955, 1958) and Bolinger (1958) give good

ground for believing that the importance of intensity as a cue to prominence is much less than used to be thought. Our experience in prosodic synthesis has tended to support this view.

The nuclearly accented syllables in the synthetic sentences cited above manifest the conjunction of two separate phenomena: accent and cadence. Either of these phenomena gives prominence to the syllable on which it occurs (in the case of cadence, that with which it begins), as Bolinger (1957, 1965) notes in discussing his "accents A and B". His way of separating the phenomena seems somewhat problematical and less economical than that adopted here, but it is preferable to the usual practice of conflating them.

It is the conflation of accent and cadence which, apparently, gives rise to the notion that the nuclear "stress" (accent) is characteristically stronger than any others in the same sense group. Such an assumption has far-reaching and unfortunate consequences. Those adopting it have to hold, for example, that words and phrases generally, (and not merely in the case of rhythmical alternation or recession -- see below) assume different "stress" patterns in nuclear or tonic position than in prenuclear position. Thus Chomsky and Halle (forthcoming), after discussing the "stresses" in *tēlegrāph*³ (where the digits relate inversely to "degrees" of stress) and its derivative forms *telēgrāphy*¹ and *tēlegrāphic*^{3 1}, observe in a footnote that in the sentence *we estāblished*² *tēlegrāphic*^{3 2} *communication*^{3 1} the word has "still another representation because of the stress modifications which take place in that context." (It is not quite clear why their rules do not give *telēgrāphic*^{4 2} in the context cited.)

Similarly, Kurath (1964) accepts the view that *father* and *sea-worthy*, for example, "have full-stress . . . in (*tèll*) *fàther* . . . and (*nòt*) *séawòrthy*," but only half-stress in "*fàther* (*knòws*) and . . . *sèawòrthy* (*vessels*), without any change in meaning or in inherent stress patterns" (his circumflex being weaker than his grave or "half-stress") although he distinguishes between the cases where "(1) the stresses are proportionately reduced, thus preserving the [inherent -- i.e. canonical] pattern" and those where "(2) the stress pattern of the phrase is rhythmically reversed."

The phenomenon of rhythmical recession and suppression of accents is well attested and important for a model of O-P conversion. But the notion that two or more accents in the same sense group are related by monotonically increasing strength, so that the last (rightmost) or nuclear one is necessarily the strongest, is one for which there has never been the slightest evidence. It is difficult to see why this continues to be maintained. Even if it were factually correct it would be just the sort of predictable variation in the signal which a linguist attempts to abstract away from in setting up theoretical units. But the impression that in prenuclear position a word's stress pattern "is toned down, but not effaced" seems to be solely due to the fact that the supporting cue of cadence -- pitch subsidence -- does not (cannot) occur with prenuclear accents. In a sentence or sense group where no cadence takes place -- as in a sustained or a rising (yes-no question) intonation -- the putative difference disappears.

By rigorously distinguishing between the high-level units assigned by the orthoepy which constitute the input to the prosody, and the

phonetic (or prosodemic) units forming its output, it is possible to illuminate the question of what is encoded in the traditional lexicographical (canonical) stress transcriptions where three "degrees" of stress are discerned. Since the IPA system of stress marking is widely used in pronouncing dictionaries in both America and Britain, and has recently been adopted (or, actually, adapted) for the pronunciations in the foremost American unabridged dictionary (NID₃), this system will be used here for marking canonical stress patterns at the input to the prosody. Such stress patterns are taken to be abstract representations of certain units of information needed for the correct assignment of accent, emphasis, and other prosodic output units -- including cadence.

The IPA system marks a vertical bar before a "stressed" syllable and nothing before an unstressed one; if the stress mark is high [¹] it indicates "primary stress", if lowered [₁] it shows "secondary stress": as

¹Red₁ coat
¹light₁ house₁ keeper
¹abso¹lutely
₁con₁den¹sation

IPA stress marks may be regarded as recording two choices: (1) whether for each syllable there is a vertical bar or not: (2) whether, if present, the vertical bar is high or low. It thus uses two bits (binary digits' worth) of information to encode what is usually interpreted as a ternary distinction (leaving aside the matter of information contained in the serial location of the stress marks in the string and assuming that the locus for each "potential" stress mark, i.e. the "beginning" of each syllable, is fixed and known -- a quite

counterfactual assumption).

But two bits can also encode two independent binary distinctions, and this is the basis for the reinterpretation of traditional lexical stress marks that is suggested by the model here postulated. One of the binary distinctions has already been suggested: presence or absence, in a syllable's abstract representation, of Bolinger's "potential for pitch accent". This is the difference between ¹Red₁coat and ¹red ¹coat: the morpheme *coat* in the first item does not have potential for accent in the canonical representation of the word, whereas in the phrase *red coat*, *coat* has canonical "full stress" -- indicating its potential for accent -- even though this potential may not be realized in a particular context: as

she didn't buy a RED coat, she bought a GREEN one.

where the de-accentuation of *coat* is due to context sensitive rules of the prosody.

But there is more wrapped up in the IPA transcriptions of stress -- even just "word-stress" -- than can be accounted for by Bolinger's important observation, which yields only a binary division between canonically accentable and unaccentable syllables. The other binary distinction involved in stress patterns is ordinarily encoded by the presence of either a high- or low-set stress mark as opposed to none. However, in discussing stress patterns it is desirable to have some means of positively transcribing a canonically weak syllable (as in the alternative plurals of *process*: ¹prəʃe₁sɪz versus ¹prə₁ʃesɛz) and for this purpose it is convenient to adopt the low-set minus sign so used in NID₃:

'pro_ces,ses ~ 'pro,ces_ses

The distinction which may be transcribed in this way is that which is at least putatively to be found in the well known pairs:

'effi_gy 'refu,gee

'post_men 'milk,men

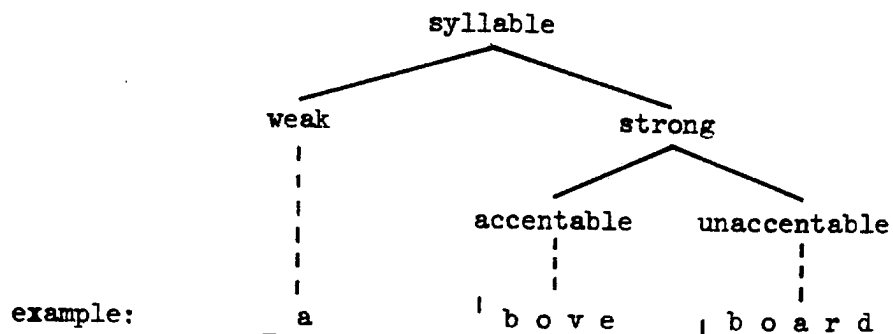
'win_dow 'el,bow

There is no entirely satisfactory nomenclature for this distinction, which in traditional Trager-Smith terms is that between tertiary and weak stress. Since all remaining distinctions of "stress level" in their system are interpreted here as matters of accentuation, this one could be designated as just *stressed* versus *unstressed* were it not for the fact that these terms are too well established in other senses to admit of this use. *Short* and *long* might be used; this pair captures the most significant cue in the phonetic output and has good historical precedent (cf. Obadiah Walker 1659). But it might seem perverse to use these terms for the distinction in question inasmuch as the durational correlates can only be noted by abstracting away from the welter of other duration effects in the data of real speech (especially phonematic and word-boundary effects -- see Abercrombie 1964, Bolinger 1965b; also below).

The distinction in question underlies the phonetic differences between the weak and strong forms of various function words such as *him, and, the* (Jones 1964 [1918]), and it is therefore convenient to refer in general to syllables with and without stress markings in their canonical representations as "strong" and "weak" syllables respectively. A "secondary" or half-stress mark is then interpreted

as identifying a canonically strong but unaccentable syllable; a "primary" or full-stress mark identifies a strong accentable one.

This may be shown in a decision tree:



The word *aboveboard* has two strong syllables, *bove* and *board*, of which only the first is canonically accentable: but of course its potential for accent might be unrealized in a particular context because of rhythmic variation: as "it was an OPEN and aboveboard MEETING". This shows clearly the relations of the two binary distinctions which in the present model are assumed to be encoded in traditional stress marks:

1. For each syllable: whether it is weak or strong;
2. For each strong syllable: whether it is accentable.

Or these distinctions can be tabulated as binary features characterizing syllables in the canonical representations of words:

	strong	
	+	-
+	'Σ	(non-occurring)
-	Σ	_Σ
accentable		

The distinction between weak and strong unaccented syllables is not accounted for by the deliberately simplified model used for synthesis by rule of prosodic features by Mattingly (1966), which "does not take any direct account of lexical stress," and in which therefore "in a non-prominent word, a stressed syllable is not treated any differently from an unstressed syllable composed of the same segmental phonemes." Mattingly rightly observes that

However, a great deal of information about stress is implicit in the selection and sequence of segmental phonemes and the sequence of morphemes. Thus, the nucleus of the stressed syllable will frequently have greater *inherent* intensity and/or duration than neighboring syllable nuclei; the fact that the phoneme /ə/ or the morpheme /r/ are normally unstressed is often sufficient to determine the stress of the word; and the suffix /k/ generally indicates that the stress must fall on the preceding syllable.

Nevertheless, there is a clear rhythmic difference in certain pairs which seems to be due entirely to the relative durations in the phonic output of weak as opposed to strong-unaccentable syllables. Since the accentuation of monosyllabic function words is quite variable, such contrasts are best illustrated by polysyllabic lexical items having established canonical stress patterns: as

'con₁tour_{ing} ~ 'cen_{ter}_{ing}
 'sit_{in} ~ 'pid_{gin}
 'in_{dex}_{es} ~ 'in_{di}_{ces}
 'op_{er}_{at}_{or} ~ 'op_{er}_{at}_{ive}

It is precisely this distinction which is noted by Chomsky and Halle for such pairs as

com_{pen}'sa_{tion} ~ con_{den}'sa_{tion}

where it is accounted for in terms of derivational morphology.

The weak-strong syllable distinction is partially accounted for in a different way by Catford (1966). By recognizing a special type of phoneme, the "transition," in conjunction with the analysis of tone groups in terms of "feet" or stress-groups (Halliday 1963a, 1966, Abercrombie 1964) he is able to describe a number of phenomena without reference to "stress" as an independent system of contrasting units. In this regard his very provocative model resembles the one put forward here (where stress is not among the phonic output units). But the result is gained in a different way, and these differences can be discussed better below after further consideration of the prosodic units in the model here proposed, and of ways of transcribing them.

The first units or features to be incorporated in the output inventory of the proposed prosodic model, as suggested above, are [+ ACCENT] and [+ CADENCE]. These features are all that are required for specifying "citation form" pronunciations -- viz.: all and only the canonically full-stressed syllables are accented, and everything after the peak of the last such syllable is in cadence. ACCENT refers to a pitch obtrusion on the vocalic portion of a syllable. It is well known that such prominence-marking obtrusions are not necessarily upward. But downwardly-obtruded accents (cf. the accent C of Bolinger 1958) are quite dispensable in reading factual prose, and in general -- for American English at least -- are attitudinally or emotionally marked. Upward obtrusion is therefore assumed as the unmarked case, and the marked cases of downward accent can be differentiated by another, indexical, feature. Malone's (1926) term DIP provides a convenient label for this feature; which, however, is not essential for

the present model. There is another marked variant of accent which may be handled in a similar way. This is the type which created consternation among adherents of the Trager-Smith school when pointed out by Sledd (1956): e.g. ²wón³derful^{1#} -- where "primary stress" is followed first by higher pitch and then a falling contour. Jones (1966 [1909]) transcribes such accents, and mentions that "English as spoken in Wales is noteworthy for the use of [this feature]." It is very common also in Hawaiian Island English (Vanderslice and Pierson 1967). This form of accent, where the upward obtrusion is delayed relative to the prominent syllable, has been called SCOOP by Hockett (Pittenger, Hockett, and Danehy 1960), a term which will be adopted here for the feature characterizing such accents. That it is a marked, affective index at least for standard English is suggested by the discussion in the work cited (dealing with a psychiatric interview). Scooped or dipped accents are extremely hard to synthesize convincingly in a terminal analog by varying only the pitch parameter. Presumably this is because, as Mattingly (1966; above) noted, the acoustical correlates of accents in real speech include also durational and intensity cues. The unmarked accent can be effectively suggested by appropriate pitch pattern alone, but the marked variants require adjustments of syllable durations (and perhaps amplitudes) simulating the subtler cues which identify the location in real speech of the invariable physiological correlates of accent: viz. a gesture of the respiratory muscles (Ladefoged 1959, 1967c) and a gesture of the larynx (Vanderslice 1967).

The features DIP and SCOOP can co-occur with the same accent, as is shown by Jones' example



. . . remarkable

-- which, interestingly, is one of Sledd's examples also (3rè²márk3able¹#). Both of these features are of the sort which can be relegated to a fuller model of O-P conversion -- one in which indexical as well as grammatical features are assigned by the prosody (see figure 1.5 above).

But there is one variation on the form of accents which is required even for the simpler model sought here: an additional degree of pitch prominence which we may refer to as EMPHASIS. This is needed -- to expound antitheses, for example -- even in factual prose. The failure to distinguish between normal and emphatic accents -- or ACCENT and EMPHASIS -- troubles Gunter (1966) a good deal in his effort to explain accent assignment. For example it sends him looking for an implied contrast in perfectly ordinary, neutral sentences like "Bob has acquired a MOtorcycle." He is thus led to ignore the difference -- in

Context: the man can see the BOY.

Response: the man can *SEE* the boy.

-- between BOY, requiring to be read only with nuclear ACCENT, and *SEE* which quite evidently demands nuclear EMPHASIS.

Similarly, Lieberman (1967), by allowing for only one degree of prominence, is unable to account for minimal pairs occurring in his own data: as

Joe ate his s^oup. ~ Joe ate his s_oup.

In fact he seems to identify his prominence feature strictly with the occurrences of an italicized word in the texts furnished to his subjects -- suggesting that its "archetypal correlate" is not physiological but orthographic.

The convention of showing accented syllables in capitals and unaccented ones in lower case will be adopted here, with the further distinction of italics (as well as capitals) to show EMPHASIS. For purposes where this is inappropriate, as in conjunction with phonetic symbols, or where a more rigorous transcription is needed -- as to show the distinction between weak versus strong-unaccented syllables in the phonic output explicitly -- the traditional marks will serve: acute (´) for ACCENT, and doubled (ˆ) for EMPHASIS; macron (¯) for unaccented STRONG syllables, and breve (˘) for weak ones (i.e.

- STRONG). The latter marks are appropriate, since the primary phonic correlate of the weak-strong distinction among unaccented syllables is (ceteris paribus) their duration.

ACCENT and EMPHASIS are related in such a way that much of what is said about the one applies, mutatis mutandis, to the other. Emphasis is a pitch obtrusion like accent, but substantially greater in extent; i.e. the unmarked case reaches a higher pitch (cf. "overhigh"). Similar marked forms occur: a dipped emphasis (or emphatic dip) is a larger-than-usual downward obtrusion of pitch on the syllable so marked. It should be noted that the term EMPHASIS is generally used here as a shorthand label for EMPHATIC ACCENT -- i.e.

+ ACCENT
+ EMPHASIS

-- and that [- ACCENT] implies [- EMPHASIS]. Thus conditions stated

in terms of e.g. rightmost or nuclear ACCENT are of course met equally well by an occurrence of EMPHASIS.

Bearing this in mind, it is important to note that in the illustrative sentences above from Gunter, the domain of CADENCE is determined by the last ACCENT -- i.e. an exponential subsidence of pitch begins on the syllable nucleus (for unmarked accents) and lasts to the end of the sense group (in falling intonations). Thus, in his examples (see page 75) BOY in the "context" starts higher but quickly falls to a lower pitch than the part of the sentence preceding the accent, and in the "response" the words *the boy* are at a lower pitch than *the man*. Similarly the last three syllables of *MOTORcycle* (canonical form 'mo_tor,cy_cle) are low compared with *acquired*.

Since the nuclear accent is defined as the last one in a sense group it is obvious that all following syllables are unaccented. And if the sense group has a falling intonation they participate in the cadence. Sentences composed of the same words in the same order may differ in their pitch patterns and their meanings even if all have falling intonation, depending on which words are accented. Thus arise the paradigms of accent "shift" which have fascinated writers from Mason (1748) to Gunter (1966) and Lieberman (1967). If there are several words without accent at the end of a sense group for any reason -- i.e. because of being either canonically unaccentable (as function words typically are) or prosodically de-accented (as repeated lexical items usually are) -- then the domain of CADENCE is increased. Sometimes the nuclear accent comes quite early and is followed by a long cadence, either in the same sense group or continued by a

succession of enclitic adjuncts which are accentless. Examples of the latter type abound because medial and final vocatives and ascriptions (cf. Bolinger 1966) are normally enclitic, and so are descriptive clauses subordinately conjoined to ascriptions, so that in the following example everything after the first word is in cadence:

"GoodBYE, Gilbert," she said, turning and walking slowly away.

Indeed a frequent solecism of journalistic prose is the inclusion in ascriptions of new information which ought to be accented but cannot be because of the overriding effect of this rule: thus Schmeck (1965) writes what would have to be read

The IDEA developed from exPERiments in MICE,
Dr. Mathe said in a preliminary report in the
British Medical Journal.

Examples of long cadence within a sense group are more apt to depend on what Hultzen (1964) calls "long context," and we must therefore look at enough of the context to see the operation of the rule. For example if the following were considered as an independent sentence, the neutral accentuation would be something like

She USED to have to FIND things, and reMIND
me of my aPOINTments.

But taken in conjunction with preceding context it is accentuated quite differently:

- (i) HIGGINS. [explaining about Eliza to his mother]
... she's useful. She knows where my things
are, and remembers my appointments and so forth.
- (ii) MRS. HIGGINS. How does your house keeper get on with
her?

- (iii) HIGGINS. Mrs. Pearce? Oh, she's jolly glad to get so much taken off her hands: for beFORE eLIZa came, SHE used to have to find things and remind me of my appointments.

EMPHASIS is assigned to *she* in (iii) because it contrasts with *she* in (i) -- i. e. despite their "morphemic identity" (Gleitman 1961) they stand for contrasting antecedents. The remainder of the clause in (iii) is de-accented as functionally synonymous with the corresponding part of (i). Obviously it is no trivial task to identify these synonymies by mechanical procedures; they depend on a knowledge of the partial congruences

- (a) to find $x \approx$ to know where x is,
 (b) to remind y of $z \approx$ to remember z ,

(where the elements of either pair could easily contrast in another environment) and of the anaphoric relations

$$\text{Mrs. Pearce} = \text{she}_{iii} \neq \text{she}_i = \text{Eliza}$$

-- which merely confirms that as pointed out in the preceding chapter, O-P conversion, like machine translation, demands sophisticated analysis of texts. Such problems are not addressed by the fragment of a prosody presented below, which assume semantic and syntactic analysis as prior tasks (cf. figure 1.6). It is beside the point to say that no form of grammatical description at present known can include such complicated relationships. Human speakers can do it; and eventually proper grammars must be able to do so as well.

Superficially there are certain resemblances between the model here proposed and that of Lieberman (1967). Thus ACCENT might seem to relate to his feature "prominence" [+P_s] and CADENCE to his "unmarked breath group" [-BG]. But as noted above, his [+P_s] is

ambiguous, and whether it equates with accent or cadence, it is inadequate to describe the data. Similarly, his [+BG] is not able to distinguish between rising intonations -- as for yes-no questions -- and fall-rise contours of continuation, for example (cf. Delattre, Poenack, and Olsen 1965). And his statement that [-BG] occurs "within the last 150-200 msec of phonation" is completely counter-factual, as shown by instances like those above where, assuming a normal speech rate, the domain of CADENCE may be an order of magnitude longer than this.

Thus each of Lieberman's features needs to be split in two -- or, to put it the opposite way: even a minimal prosodic model of American English requires more distinctions than his two features provide. The distinction between ACCENT and EMPHASIS has been dealt with. The intonational distinctions that are conflated in his [+BG] are provided for in the present model by separating the features that refer to the fall and to the rise instead of using a single feature to separate merely falling from rising "tunes". That is, in this work CADENCE refers to the presence or absence of a post-nuclear fall of pitch. Whether or not the pitch rises at the end of the sense group is (as a first-order approximation) an independent binary distinction, here assigned to the feature [+ ENDGLIDE].

Thus in the case of the fall-rise intonation, as for continuation, the monotonically falling pitch of CADENCE is arrested and succeeded by ENDGLIDE -- a sharp upward pitch glide in the last few centiseconds of the sense group. In the case of rising intonation, as on typical yes-no questions, there is no cadence or pitch fall after

the nuclear accent, and ENDGLIDE is therefore realized as a high rise instead of a low one. Both features are absent in sustained or level-ending intonations. Thus: these two features discriminate the four standard intonations:

		ENDGLIDE (RISE)	
		+	-
	+	Falling-Rising(,)	Falling(.)
CADENCE (FALL)	-	Rising(?)	Sustained(--)

To transcribe these intonations informally, standard punctuation marks can be used (as in some examples and the matrix above). Thus a period, representing a falling intonation, encodes the features

+ CADENCE
- ENDGLIDE

and a question mark the reverse (so that it would not be used after a normal WH-question). Similarly a comma encodes presence of both features and a dash their absence, as shown in the matrix.

But in more formal transcription it is better to represent the features discretely, and as the form of their representation is irrelevant from a theoretical point of view, there is no objection to doing so in such a way as to make the transcription as perspicuous as possible. Accordingly the traditional letters F (= fall) and R (= rise) will be used for their mnemonic value, representing [+ CADENCE] and [+ ENDGLIDE] respectively, and located (quite redundantly, of course, but perhaps usefully) near the points in the segmental string corresponding to where the phenomena represented have effect. Thus when

cadence is to be transcribed, the F can be placed over the line of segmental symbols (orthographic or phonetic) just after the nuclear accent mark, since this corresponds to where the pitch fall is to begin. The R for (rising) endglide can be put above the line at the end of the sense group. (The term "sense group" refers to the constituents between pauses -- assigned as described in the rules below.) If the nuclear accent is on the last syllable, the letters will be located close together, as are the occurrences in time of the phenomena they represent. Thus

Shě [^]used tō have tō ^{^F}find [̄]things/
 and remind mē of my [^]appointmēnts/

and

GO HOME, now YOU go STRAIGHT HOME,
 gō ^{FR}hōme/ nōw yōu gō ^{FR}strāight hōme/

where the "rising slur on the final element turns such commands . . . into admonitions" (Heffner 1964). Or

Tō bē or nōt ^{^F}tō bē/^{̄R} thāt is thē ^{^F}questiōn/
 To BE or NOT to be, THAT is the QUESTion.

In continuous text at least, the end of each sense group has to be shown; and while the R defines it when present, since it is nevertheless an independently motivated unit in the output, an independent boundary marker such as the virgules above must be used; if only because in O-P conversion the task of delimiting sense groups needs to be done ahead of intonation (cadence and endglide) assignment. The virgule is a boundary mark representing "pausal phenomena" such as the drawing or slower articulation of the last few phonemes preceding

it. It is the only notational device distinguishing an utterance in two sense groups, of which the first has sustained intonation, from the same utterance as a single sense group.

One other type of feature is needed in the prosodic model. It has been repeatedly emphasized, by both those who analyze intonation into tunes and those who prefer levels, that the linguistic pitch contrasts of English are relative to the speaker's "characteristic range of notes, or compass, within which the pitch fluctuation of his voice falls during normal circumstances. This range or compass is the *tessitura*." (Abercrombie 1967, italics mine) Variations in a speaker's tessitura -- or "register" as it has been called by the Trager-Smith school (e.g. Pittenger, Hockett, and Danehy 1960) -- have usually been regarded as a paralinguistic or indexical feature rather than as having grammatical significance. But quite aside from the functioning, perhaps in a gradient way, of tessitural variations as affective indices, there are discrete shifts of tessitura which are required for non-anomalous renditions of straight factual prose. These are of two types: (1) a lowering and compressing of the tessitura (with respect to its neutral values) for parenthesis, etc., and (2) a raising and narrowing of the tessitura for material in quotation marks and certain other constructions.

The raising and lowering of the vocal pitch pattern for quotations and parentheses are well-known phenomena. Mason (1748) included them under his basic nomenclature for pitch obtrusion and subsidence:

Every *Parenthesis* is to be pronounced in *Cadence*; that is, with a low Voice, and quicker than ordinary; that it may not take off the Attention too much from the Sense of the Period it interrupts. But all *Apostrophes* and *Prosopopæias* are to be pronounced in *Emphasis*.

These shifts of register are here identified as the features UPSHIFT and DOWNSHIFT. They are transcribed by pairs of angle and square brackets (respectively) enclosing the sense groups to which the features apply. Since the domain of a shift is always coterminous with one or more sense groups, it is convenient to regard these brackets as taking the place of the virgule when they occur. This is merely a notational convention to avoid cluttering the text with redundant marks that do not aid the user.

Table I lists the basic prosodic features adopted in this model for assignment in the output of the O-P conversion process. These features appear to be sufficient to expound the grammatically relevant prosodic contrasts of American English. After each feature is given its symbol, a short definition, and an indication of the formal means by which the feature can be simulated in a terminal analog. The effects shown in these diagrams are assumed to be obtained in the process of translating the symbolic output string of features (phonemes and prosodemes) into parametric control functions for a hardware synthesizer. The pitch obtrusions for accent and emphasis, for example (and for endglide as well) are obtained by step-function changes in the voltage associated with the F_0 parameter. This approach owes much to the work of Isačenko (1965). But to obtain greater naturalness the step function is modified (as suggested by Öhman and Lindqvist 1965) by passage through a low pass filter before reaching the voltage-to-frequency converter. (See also Öhman 1967 and Bakis 1967 for important suggestions, published too recently to be utilized here, concerning step-function intonation control and smoothing algorithms.)

TABLE I. Basic prosodic output features

<u>Feature</u>	<u>Symbol</u>	<u>Definition</u>	<u>Recipe for synthesis</u> (of unmarked form)
ACCENT	'	pitch obtrusion	
EMPHASIS	"	double pitch obtrusion	
CADENCE	F	post-nuclear pitch subsidence	
ENDGLIDE	R	terminal pitch rise	
PAUSE	/	terminal drawl	(decrease rate of reading parameter values during last few syllables)
UPSHIFT	< >	tessitura up and narrowed	
DOWNSHIFT	[]	tessitura down and narrowed	

TABLE II. Partial inventory of related indexical features

<u>Feature</u>	<u>Transcription</u>	<u>Effect</u>
DIP	$D\underset{\sim}{\int}$	inverts pitch obtrusion
SCOOP	$S\underset{\sim}{\int}$	delays pitch obtrusion
FAST SLOW	-- --	continuum of rate -- vary speed of reading commands
PORTAMENTO GLISSANDO	-- --	continuum of pitch control -- vary filter time constant
SPREAD STRAIT	-- --	continuum of tessitura range. -- vary amplifier slope
HIGH LOW	-- --	continuum of tessitura position -- vary amplifier intercept

PRELIMINARY RULES FOR ASSIGNMENT OF OUTPUT FEATURES

The following are tentative rules for assigning the features listed in Table I.

I. Pause assignment

- A. Locate virgules provisionally (to be revised, principally by deletions, in subsequent steps):
1. At each punctuation mark [, ; : . ! ? () --]
 2. At each major syntactic boundary

Virgules are deleted usually in an ordered sequence inversely related to the priority of the nodes they represent -- i.e. those between most closely related elements are deleted first. For simple sentences, the virgule separating subject and predicate is last to go. Extent of deletions is influenced by stylistic options selected ("fast" deletes more virgules; "slow" retains many) and by considerations such as the number of accented words and the total length of the elements separated by a virgule (the more syllables and/or accents, the less probability of deletion).

II. Accentuation

- A. Provisionally assign accents to every canonically full-stressed (strong accentable) syllable in the input text, and macrons to half-stressed ones (strong but unaccentable).

Function words are considered to have either weak- or half-stress in their canonical representations -- or both, for those with weak and strong forms. Pike (1945b) provides the following division of function and content words:

(A) Function words are as follows:

pronouns (personal, reflexive)
auxiliary verbs
prepositions
connectives (conjunctions, relative pronouns, etc.)

articles
adverbs of degree

(B) Content words are as follows:

nouns
principal verbs
 (i.e., verbs which do not stand in
 auxiliary position before other verbs)
adjectives
adverbs of time, place, and manner
demonstratives
interrogatives
indefinite pronouns

These lists may be accepted with minor caveats about the indefinite pronouns, and about certain adverbs like *today*, *last night*, *yesterday*, *tomorrow*, *now*, *then*, *yet*, etc., and *here*, *there*, etc., which if accented in sentence final position are contrastive even without emphasis (cf. the concept "marked tonicity" of Halliday 1963a, 1967).

These must be considered to have canonical half-stresses in order to account for the fact that

I admíred his cónduct lāst nīght

is the neutral rendition (of either parsing of this sentence)-- i.e. with nuclear accent on the full-stressed penult of 'con,duct, but with ,last ,night as well as the ultima in cadence, being strong but unaccentable syllables. Such syllables and even canonically weak ones can of course be accented under appropriate conditions, e.g. of contrast. But *neutral rendition* is here defined as that assigned to a sentence by the context-free rules which assign citation-form pronunciation of words.

- B. Delete accents (i.e. substitute macrons) from all echoic elements, unless
1. They participate in antithesis, or
 2. They are part of a *thematic summary* (see below).

Echoic elements are words or phrases which repeat either the form or the semantic content of a word or phrase preceding in the same text, within certain restrictions as to syntactic parallelism, proximity, etc. The clearest and strongest cases involve both formal identity and semantic equivalence, but either is sufficient. The difficulty arises in cases where it is less than clear from the written text whether the second element, being formally distinct, is being employed as a synonym for the first. An example is

Overt authority used physical force; anonymous
authority employs psychic manipulation.

Here it is obvious that *overt* ~ *anonymous* contrast and so do *physical* ~ *psychic*; the *authority*'s are formally identical, and *use* ~ *employ* are synonymous (in general tense distinctions will produce contrastive accent assignment only when nothing else can contrast). The problem pair is *force* ~ *manipulation*. These terms name extensively overlapping categories, and could be intended here as synonyms of a single category differentiated by the adjectives, in which case the second, echoic term at least should be de-accented. Or the nouns may be intended by the author to refer to non-intersecting portions of their semantic ranges, such that there is a triple antithesis in the sentence. The author's intention being moot, the best strategy for handling such ambiguities in automatic O-P conversion is to keep the accents. The resulting rendition may be opaque, but it is less likely to be positively misleading than if the accents are deleted and the words in question spoken in cadence. To delete the accents is to say definitely that the terms are synonymous.

Except for the case of pronouns, formal identity results in

de-accentuation, given appropriate parallelism and proximity, whether or not synonymy accompanies it. That is, the identity need not be referential -- and it need not even be morphemic as suggested by Gleitman (1961); mere homophony is enough to prevent contrast and thereby enforce de-accentuation. This is shown by the adoption of abnormal nonce pronunciations for the purpose of contrasting pairs like *aural* ~ *oral* or *horse* ~ *hoarse* by speakers who normally use the same vowel in such words.

A thematic summary is one or more sentences, typically at the end of a paragraph or comprising a separate short paragraph at the conclusion of a series on the same theme, which (a) terminate a sequence of sentences sharing an inventory or "natural language list" of related lexical items, and (b) contain only echoic elements, or at least a very small proportion of new lexical items. (The identification of such relationships among the sentences of a text is assumed here to be the function of the semantic analyzer, and the techniques to be employed lie, perhaps fortunately, outside the scope of this work.)

- C. Delete accents on procatarctic elements (the antecedents of echoic ones) if they occur in the same sentence as their echoic partners.

While less mandatory than (B) this rule is usually followed and should perhaps even be extended to two contiguous sentences. It accounts for

JOHN studies eLECTrical engineering, while
BILL is in meCHANical engineering.

- D. If two words meeting the parallelism and proximity conditions are morphemically identical or homophonous except for differing prefixes and are not

canonically forestressed, accent the prefixes and reduce the accents on the subsequent syllables. If one of the words has zero prefix, its accent is unaffected.

Example:

SPEECH uses CONcrete, SINGing DIScrete intervals.

This rule was given by Jonson (1640) as follows:

.

sociable, tolerable.

When they be compounded, they Keepe the same accent: as
insociable, intolerable.

But in the way of a comparison, it altereth thus: some men are *sociable*, some *insociable*; some *tolerable*, some *intolerable*. For, the accent sits on the syllabe that puts the difference: as

sincerity [sic], *insincerity*.

That morphemic identity is not a condition is shown by:

I said *DE*plane. not *COM*plain,

They should just *IN*tercede. not try to *SU*persede,

(see emphasis rule below).

That semantic contrast in the prefix(es) is not necessary is suggested by

YOU RAVel it and I'LL UNravel it.

-- although this is quite evidently a hypostatic joke. In order to accent a canonically weak syllable, it is necessary of course to make it strong. This is one of the few cases where the prosody imposes any change on the distribution of canonically strong and weak syllables assigned by the orthoepy. Since the strong-weak distinction is normally carried through in the output of the prosody exactly as assigned at the input it has not been treated here as expounded by a separate pair of output features to be assigned by the prosody. Weak syllables become strong (redundantly) when they are accented, as by

the above rule, but no other changes are made in the strong-weak distinction. Note that even under so-called "stress shift" a canonically accentable syllable, though de-accented, nevertheless remains strong: as "we have *dōne* that which we ought to have left *úndōne*".

Obviously there has to be either a strategy for regenerating strong forms from weak ones, with the loss of information this entails, or the strong forms have to be assumed to be indicated in the canonical orthoepic representations. The well known cases of "re-stressed" pronunciation such as *óf* \wedge v and *thé* δ \wedge or δ i suggest that the first option is not too gross an oversimplification of human speech behavior. But to get *kánvīkt* from *con'vīct* (under accent shift) would require a representation of the first vowel other than schwa. This of course was the practice in lexicographic pronunciations until quite recently, when the objections of the phoneticians brought about various concessions to the biuniqueness principle. On the view suggested by the model of O-P conversion postulated in this work, the representation by different symbols of the vowels in the weak syllables of *above*, *women*, *pencil*, *lemon*, and *circus* is not a bad idea after all.

Where the differentiating syllables are not prefixes, or are not readily recognizable as such, a shift of accent is not mandatory; in fact such a pair may sound anomalous in any possible rendition -- as for example the propaganda slogan disseminated by the American Medical Association:

Let's *use* not *abuse* health insurance!

Here the *ab* is not readily identifiable as a prefix by most speakers,

for whom the canonical form is $\text{,} \text{e}^{\text{b}} \text{juz}$, so that the two strong syllables are non-identical though rhyming.

Rightward shifts of accent are much less the norm, and in fact occur mainly when the matching accentable syllables are followed by canonically unaccentable but strong differentiating syllables -- especially if distinct morphemes: as

I $\text{s}\bar{\text{a}}\text{i}\bar{\text{d}}$ $\text{h}\bar{\text{a}}\text{l}\bar{\text{f}}$ $\text{l}\bar{\text{i}}\bar{\text{f}}^{\text{FR}}$ / $\text{n}\bar{\text{o}}\bar{\text{t}}$ $\text{h}\bar{\text{a}}\text{l}\bar{\text{f}}$ $\text{l}\bar{\text{i}}\bar{\text{g}}\bar{\text{h}}\bar{\text{t}}^{\text{FR}}$ /

(where the forms are canonically 'half_{N}). But in pairs such as $\text{'pou}\bar{\text{l}}_{\text{t}}\text{ico} \sim \text{'pou}\bar{\text{l}}_{\text{t}}\text{ry}$ where the differentiating syllables are weak there is normally no accent shift. Even where one of the two is strong, as in $\text{'for}_{\text{mat}} \sim \text{'for}_{\text{mant}}$, a shift of accent is anomalous -- but there are dialects which have $\text{'f}\bar{\text{a}}\text{r}_{\text{mant}}$, and for such speakers a contrast like

$\bar{\text{I}}$ $\text{d}\bar{\text{i}}\bar{\text{d}}\bar{\text{n}}\bar{\text{'t}}$ $\text{s}\bar{\text{a}}\text{y}$ $\text{f}\bar{\text{o}}\bar{\text{r}}\bar{\text{m}}\bar{\text{a}}\bar{\text{t}}^{\text{FR}}$ / $\bar{\text{I}}$ $\text{s}\bar{\text{a}}\bar{\text{i}}\bar{\text{d}}$ $\text{f}\bar{\text{o}}\bar{\text{r}}\bar{\text{m}}\bar{\text{a}}\bar{\text{n}}\bar{\text{t}}^{\text{F}}$ /

would presumably be quite normal.

Similarly when the strong accentable syllables of two words are identical but there are differentiating syllables to both right and left there is no accent shift; as in "she THINKS she's the picture of eFFICIency, but she's REALly a picture of offFICiousness." (Note that the first *picture* -- the procatartic -- is de-accented by rule C.) But in contrasting *continuously* with *continually* rightward accent shift would seem to be the norm.

E. Rhythm Rule

1. Delete the accent on the middle one of three consecutive accentable syllables in the same sense group.
2. Delete the middle accent similarly if the middle syllable is separated from the others by not more than one strong unaccentable syllable, or n weak syllables, or one strong plus $n-1$

weak syllables, where n is a small positive integer, variable according to stylistic options such as fast-slow.

3. If a sequence of more than three syllables meeting these criteria occurs, apply the rule to sequences of three as follows until the conditions in 1 and 2 no longer apply:
 - a. if two of the accentable syllables are in the same word (or hyphenated pair), include them in the first three chosen.
 - b. otherwise take the closest three syllables.
 - c. if (a) and (b) do not resolve the choice, start with the rightmost three.

NB: a syllable which has been de-accented is thenceforth treated exactly like a strong unaccentable one.

This rule applies in general to successions of unaccented syllables whether in the same or different words, and it is extensively illustrated in the sample text below. In particular, it accounts for rhythmical alternation or "recession" of accent within a word, since it is assumed that all words in which these phenomena occurs have canonical forms with two strong accentable syllables. Thus ¹clari¹net is the postulated canonical form underlying

hě pláys thě clār(ĭ)nĕt^F/
 ĭt wás á clár(ĭ)nĕt sólo^F/

and in general those words which undergo alternation have to be assumed to have "double stress" -- i.e. two full-stressed syllables in their canonical forms. For example, such well known cases as

the úknōwn sóldier ~ the sóldier is ūknōwn
 some hōmemāde jām ~ the jām is hōmemāde
 a sóuped ūp cár ~ the cár is sōuped ūp

presuppose the canonical forms ¹un¹known, ¹home¹made, and ¹souped ¹up (if the latter is regarded as a single lexical item). Likewise the Irish dialect, where Bridges (1921) says that recession of accent is

still heard as in *éxtrême únción*, must be assumed to have the canonical form 'ex'treme, whereas in standard English the first syllable is not only unaccentable but weak. (Of course Bridges' assertion that, in RP, "recession of accent is not now heard" is incorrect, and shows merely that he noticed it only when it occurred on forms for which he as an RP speaker did not also have canonical double stress.) Such double stressed words would of course be predicted to have two accents in citation pronunciation, or in a context where surrounded by a number of unstressed syllables.

It appears that words with more than one strong syllable can be divided into three main classes, based on whether they have, in their canonical representations, *double*, *rising*, or *falling* stress. Double-stressed words have two accentable syllables, and in addition to those already cited, include (for many dialects of English)

'syste'matic, 'repre'sent, 'peri'odically, 'bene'ficial,
'micro'scopic, 'un'moved, etc.

The other classes have only one canonically accentable syllable. Words with falling stress have an unaccentable syllable following the accentable one. This class is well known and conspicuous because of the added effect, in citation pronunciations, of the last strong syllable's being not only unaccented but also in cadence. It includes, e.g.

'verti,go, 'can,not, 'solip,sism, 'obfus,cate,
'fructi,fy, 'ope,rator, 'macra,me, etc.

It also includes a large number of compound nouns such as *blackbird*, *nitwit*, *isinglass*, etc.

Rising stress refers to the class of words with a strong unaccentable syllable preceding the accentable one. Some examples are:

1compen'sation, 1con,den'sation, (1)con,figu'ration,
 1obser'vation, 1orien'tation, 1saty'riasis,
 1pene'tralia, 1ine'narrable, 1papu'liferous.

The class of rising-stressed words, not having the effect of cadence to point up the distinction, is not so readily separated from that of double-stressed ones, and it is not surprising that a good deal of disagreement is found among the examples used by those who have troubled to discuss these classes at all. Jones (1964 [1918]), to whom the term *double-stressed* is due, gives *unwieldy*, *undoubted*, and *underline* as exceptions -- i.e. having rising-stress -- along with *discourage* and *inordinate*. But the former seem to have canonically double stress (cf. NID₃) despite the fact that they cannot ordinarily be contrasted with *wieldy*, *doubted*, *line* in corresponding senses. Pike (1945a) gives *umbrella*, *publication*, and *educability* as double stressed, though they seem clear cases of rising (cf. NID₃). The test suggested by the present model is whether the words do or do not undergo rhythmical variations in accent placement depending on whether they are closely preceded, or followed, by another accented syllable. It is important to distinguish between these classes of words because in general accent and emphasis are assigned to words, not to syllables. For polysyllabic words, the syllable or syllables to be accented under various conditions of surrounding accents must be assumed to be encoded in the canonical orthoepic form.

F. Verb Rule: delete accent on a verb if it comes

between (a) its subject and (b) an object, complement, or adverbial; and if both (a) and (b) are accentable -- i.e. contain one or more accentable words.

This rule is merely an extension of the rhythm rule, which de-accentuates the verb in such paradigm cases as *John enjoys parties*; the verb rule extends the effect to cases where the number of intervening weak syllables would prevent the rhythm rule from operating. There are various exceptions, of which participation in antithesis with another verb is the most noteworthy.

Personal pronouns constitute a special case of anaphora, and are accordingly unaccented as echoic elements most of the time. Indeed, as function words they can be considered to have, at most, half-stress in their canonical representations. But they are more readily accentable on a facultative basis than most functors.

- G. Pronoun Rule: a personal pronoun is accented whenever (a) it participates in antithesis; or (b) the unmarked antecedent-candidate is not the true antecedent.

This accounts for sentences like "Bill saw John across the room and he ran over to him", in which it is Bill who did the running unless *he* and *him* are accented, since the unmarked antecedent in such instances is in the same case. But if the second clause only were passive, then *by him* would refer to the subject of the first clause: as "John hit Bill and then he was tripped by him", which means that Bill was tripped if *he* is accented and that John was tripped if *by him* is accented. If both or neither are accented, the sentence is a conundrum.

III. Emphasis

Assign EMPHASIS:

1. To all accented syllables in a sentence terminated by an exclamation point.
2. To the accented syllables of elements in antithesis, as follows:
 - a. To the second element, obligatorily except as in (c).
 - b. To the first element, optionally except as in (c).
 - c. If the second element is negated, assign emphasis only to the first.

The first part of this rule can be passed over with the remark that exclamations are not usual in straight factual prose. The second part accounts for the differing placement of emphasis as between

He didn't go to Chicago^{F R}/ he went to New York^F/

and

He went to Paris^F/ not to London^{F R}/

and in several related examples above.

IV. Cadence and Endglide

- A. Assign falling intonation [+ CADENCE, - ENDGLIDE] to a sense group terminated by a period, exclamation point or semicolon in the orthographic text.
- B. Assign fall-rise intonation [+ CADENCE, + ENDGLIDE] to those terminated by a colon or comma.
- C. Assign sustained intonation [- CADENCE, - ENDGLIDE] to sense groups terminated without punctuation, or by an unpaired dash, unless the second of two sense groups in an antithesis; then assign fall-rise.
- D. Assign rising intonation [- CADENCE, + ENDGLIDE]
 - (1) to sense groups terminated by a question mark,
 - (2) to the first of two sense groups in an antithesis.

Part D(1) of this rule is clearly an egregious oversimplification of the complex problem of interrogative intonations. The justification for this is that direct questions are relatively infrequent in texts of the type here essayed, and that WH-questions and alternative questions, which normally end with a falling intonation, can have rising intonation if they are asked as rhetorical echo questions --

a solution perfectly congruent with the role of any such questions occurring in straight factual prose texts.

V. Upshift and Downshift

- A. Assign square brackets (downshift) to
 1. material between pairs of orthographic parens.
 2. material between pairs of orthographic dashes.
 3. nonrestrictive clauses enclosed in commas.
 4. embedded sentence adverbs.
 5. medial and final vocatives.
 6. medial and final ascriptions, and material subordinately conjoined.

NB: (4), (5), and (6) are also [- ACCENT, + CADENCE, + ENDGLIDE]
- B. If an accented function word immediately precedes left square bracket, assign rising intonation; otherwise fall-rise.
- C. Assign angle brackets (upshift)
 1. To any sentence-initial structure not belonging to the subject.
 2. Replacing square brackets as in (A) whenever any element contained in them participates in an antithesis.
- D. If no punctuation, or only a comma, intervenes between left angle bracket and preceding text word, rewrite the intonation contour at that point to *sustain* [- CADENCE, - ENDGLIDE].

The way in which these rules operate can best be illustrated by applying them to a sample text. The text used is the first paragraph of a *Scientific American* article on "Aftereffects in Perception" by W. C. H. Prentice. This text was chosen because (a) it is a good example of the type of prose for which the model is expected to be adequate, and (b) a partially automatic parsing (by the Kuno grammar) and anaphoric analysis (by Olney's post-processing program) were available through the courtesy of John Olney of the System Development Corporation, Santa Monica. It thus furnished some indication of the kinds of information that could be expected to be extracted from such a text and presented in explicit terms at the input to the pronuntiatio, or elocutionary transfer function, of a practical

TABLE III. Sample text transcribed, with respect to canonical orthoepic stress patterns and pause-assignment indices, as it might be represented at the input to the prosody.

- S1. 'Hu_man_ex'per_i_ence⁴ _and² 'hu_man_be'hav_ior⁵ ,are³
_ac'ces_si_ble² _to¹ ,ob_ser'va_tion³ _by¹ 'ev_(e)ry_one.⁷
- S2. _The_psy'cho_lo_gist⁴ 'tries³ _to² 'bring¹ ,them² ,un_der¹
'sys_te'ma_tic 'stu_dy.⁷
- S3. 'What³ 'he² _per'ceives,⁴ _how'ev_er,⁵ 'an_y_one² ,can¹
_per'ceive;⁶ 'for¹ 'his 'task⁵ 'he² _re'quires³ 'no
'mi_cro|scope² _or¹ _e|lec'tron_ic 'gear.⁷
- S4. _A 'gen_u_ine _dis'cov_er_y² _in¹ 'this 'field --⁶ _a 'whol_ly
'new 'it_em² _of¹ _ex'per_i_ence³ _or² _a 'fact¹ _a'bout¹
_be'hav_ior² 'pre_vi_ous_ly 'un'known --⁶ ,is⁴ _ac'cord_ing_ly
⁴ 'rare.⁷
- S5. 'Once _in _a 'while,⁴ 'none_the'less,⁵ 'such _a _dis'cov_er_y²
,is¹ 'made.⁷
- S6. _With,in¹ _the 'past 'few 'years⁵ _some 'sim_ple ,ob_ser'va_ti-
ons⁴ ,have 'un'cov_ered³ 'new² _and¹ _dis'turb_ing² 'facts
³ _a,bout¹ 'hu_man 'sen_sor_y _ex'per_i_ence,⁶ _par'tic_u_lar-
ly³ 'vis_u_al _ex'per_i_ence.⁷
- S7. ,Com_pre'hен_sion² _of¹ 'these 'facts⁵ ,is 'bring_ing _a'bout
³ _a 'rad_i_cal _re'vis_ion⁴ _in¹ _the _pre'vail_ing 'con|cept
³ _of¹ _the 'na_ture² _of¹ 'that _ex'per_i_ence.⁷

NOTE: Superscript digits indicate number of iterations of pause-deletion required to eliminate pause from corresponding word interstices. They reflect primarily the depth or height of the nodes joining constituents (or the number of brackets separating them) but are conditioned also by length and accentability criteria. They show order of pause deletions as faster style-rate options are implemented. Pauses deleted in output below for index ≤ 3 .

reading machine. At present the Kuno grammar results in multiple parsings of many of those sentences it is able to handle. The main motivation for development of suprasentential semantic analysis is to provide procedures to reject spurious parsings and, ideally, to choose the one right one for each sentence. It remains to be shown that such analyses also provide the necessary information to enable correct assignment of prosodic features.

A f t e r e f f e c t s i n P e r c e p t i o n

(The text is presented as a column of words on the left. The digits preceding each word are the serial numbers of the sentence and of the word in that sentence. Input is exemplified in Table III.)

1.01	Húman	accent retained on procatartic member of pair (cf. 1.04) because paragraph-initial.
1.02	expérience ^F /B	pause retained because both elements of compound subject contain adjectives.
1.03	and	
1.04	hūman	accent deleted because 1.04 echoic to 1.01.
1.05	beháviór ^F /B	
1.06	are	
1.07	accéssible	
1.08	to	
1.09	ōbservátiōn	canonical form ,obser'vation
1.10	by	
1.11	éveryone. ^F /	

- 2.01 The
- 2.02 psych^Fologist^R/ emphasis assigned because of antithesis with 1.11, pause retained because of emphasis.
- 2.03 tr̄ies de-accented by verb rule
- 2.04 to
- 2.05 br̄ing infinitive forms excepted from verb rule when conjoined with finite verbs.
- 2.06 them anaphoric to 1.01-1.05.
- 2.07 under
- 2.08 s̄ystem̄atic emphasized by contrast with 1.09 canonical form 'sys te'mat ic, but rhythm rule operates.
- 2.09 st̄udy^F./ emphasized heuristically: text is ambiguous whether *observation* (1.09) is a kind of study that is nonsystematic, or an activity contrasting with the entire concept of 2.08-2.09.
- 3.01 <Wh̄at upshift for non-subject initially.
- 3.02 h̄e accented although anaphoric to 2.02 because of antithesis with 3.05.
- 3.03 perc̄eives^F,^R> could optionally be de-accented as procatarctic to 3.07.
- 3.04 [h̄ow̄ever^F,^R] cadence continued by downshift rule (sentence adverb in medial position).
- 3.05 ānyone^F emphasis by antithesis with 3.02.
- 3.06 can
- 3.07 perc̄eive;/ de-accented as echoic to 3.03 semi-colon assigns falling intonation.

- 3.08 $\overset{-F}{\text{for}}$ *for* accented when clause-initial preposition (as opposed to conjunction).
- 3.09 *his*
- 3.10 $\overline{\text{task}}^R$ de-accented as anaphoric to 2.03-2.09 (i.e. "the task of trying to bring them [human experience and human behavior] under systematic study")
- 3.11 *he*
- 3.12 *requīres* de-accented by verb rule
- 3.13 *nó*
- 3.14 $\overset{F}{\text{micr}}\overline{\text{osc}}\overset{R}{\text{ope}}$ canonically falling-stressed
- 3.15 *or*
- 3.16 *elētronic* canonically rising-stressed
- 3.17 $\overset{F}{\text{g}}\overline{\text{ear}}./$ accented heuristically, ignoring possibility that 3.14 implies *gear* procatactically.
- 4.01 *A*
- 4.02 *genuine*
- 4.03 $\overset{F}{\text{disc}}\overline{\text{ove}}\text{ry}$
- 4.04 *in*
- 4.05 *this*
- 4.06 $\overline{\text{field}}^R \text{--} /$ de-accented as anaphoric to 1.01-1.05 and 2.01-2.09; i.e. "The field in which human experience and behavior are studied".
- 4.07 [*a* downshift assigned by dash pair.
- 4.08 *wholly*
- 4.09 *nēw* accent deleted by rhythm rule.
- 4.10 $\overset{F}{\text{it}}\overline{\text{em}}$

- 4.11 of
- 4.12 experience^R/ de-accented as echoic to 1.02.
- 4.13 or
- 4.14 a
- 4.15 fact
- 4.16 about polysyllabic functors have, canonically, only weak versus strong unaccentable distinction.
- 4.17 behavior echoic to 1.05.
- 4.18 previously
- 4.19 unknown^{F R} --] canonically 'un|known, but prefix de-accented by rhythm rule.
- 4.20 is/ sustained intonation assigned to accented functor before left bracket.
- 4.21 [accordingly^R] sentence adverb in medial position.
- 4.22 rare^F./ cf. 5.10 -- ellipsis for "is rarely made"
- 5.01 <Once upshift assigned to non-subject structure in sentence initial position.
- 5.02 in
- 5.03 a
- 5.04 while,^{FR}>
- 5.05 [nonetheless^{F R},] cf. 4.21.
- 5.06 such^F accented as anaphoric flag; whole phrase 5.06-5.08 anaphoric to 4.01-4.06, etc.
- 5.07 a
- 5.08 discovery echoic to 4.03.

- 5.09 ^Fis emphasized by antithesis with 4.22 (if ellipsis there uncovered)
- 5.10 māde./ de-accented as echoic to procatartic *made* implied in 4.22, if ellipsis uncovered; if not, then 5.09-5.10 are accentuated *is māde*, suggesting the somewhat unsatisfactory contrast *is rare ~ is made*.
- 6.01 <Within accented as clause-initial preposition.
- 6.02 the
- 6.03 p̄ast accent optionally deleted by second application of rhythm rule, depending on style option.
- 6.04 f̄ew by rhythm rule.
- 6.05 ^{FR}years >
- 6.06 some
- 6.07 s̄imple
- 6.08 ^F̄bservations^R/
- 6.09 have
- 6.10 ^R̄nc̄overed by verb rule.
- 6.11 n̄ew
- 6.12 and
- 6.13 d̄ist̄urbing accent deleted by rhythm rule (but heuristic should be sought to retain it here).
- 6.14 f̄acts pause deleted by length and accent criteria.
- 6.15 ab̄out cf. 4.16.
- 6.16 h̄uman echoic to 1.01, 1.04.
- 6.17 ^Fs̄ensory cf. 6.20.
- 6.18 exp̄erience^R/ echoic to 4.12, 1.02.

- 6.19 particularly
- 6.20 ^Fvisual emphasized by antithesis with 6.17.
- 6.21 exp̄erience./ cf. 6.18.
- 7.01 ^FComprehension
- 7.02 of
- 7.03 thēse non-accentable anaphora flag, 7.03-7.04 anaphoric to 6.11-6.21.
- 7.04 ^Rfācts/ echoic to 6.14.
- 7.05 is
- 7.06 bringing
- 7.07 ab̄out by rhythm rule.
- 7.08 a
- 7.09 rādical
- 7.10 ^{F R}revision/
- 7.11 in
- 7.12 the
- 7.13 prevāiling
- 7.14 cōncēpt pause deleted by length and accent criteria.
- 7.15 of
- 7.16 the
- 7.17 ^Fnāture pause deleted because no remaining accents.
- 7.18 of
- 7.19 that
- 7.20 exp̄erience./ echoic to 6.21, etc.

These rules are clearly in a very preliminary state as regards their prospects for being programmed and used in a real reading machine. There are a number of questions, such as the form of the syntactical and semantic information assumed at the input, which have been deliberately skirted around. It is hoped that the rules and their exemplification are sufficient to show that, given full syntactic and semantic analysis plus appropriate canonical orthoepic representations of the words and idiomatic phrases of a text at the input to a prosodic-feature-assignment program, it is possible to write explicit rules which will assign features to the output which will expound all grammatically relevant prosodic distinctions. Further it should be noted that the minimal set of output units here postulated constitutes an empirical hypothesis about prosodic distinctions in American English; i.e. it asserts certain consequences at the level of phonetic fact which are entirely testable. It asserts, for example, that "stress" distinctions are either distinctions between strong and weak syllables, or else they are distinctions of accentuation, subject to all the vagaries of context sensitive rules of the prosody. Statements in the literature about correlates of "stress" as it distinguishes noun-verb pairs like 'di,gest versus ,di'gest or 'tor,ment versus ,tor'ment (Ladefoged 1960, 1967c) have to be interpreted, in terms of the present model, as treating the correlates of ACCENT. That is, since the words in these pairs all have two strong syllables, and identical segmentals, it is here claimed that they can be distinguished only by the assignment of ACCENT to one syllable or the other. And in fact in Ladefoged's experiments the test word was always

under nuclear accent, so that the measured correlates would include the effects of CADENCE, for the statement test sentences.

The difference is of course purely one of nomenclature as far as Ladefoged is concerned, since he is obviously discussing the nature of the physiological correlates of features in the phonic output, and simply uses the term *stress* for the feature which is here called *accent*. But it would be a substantive difference if anyone supposed that there were inherent or invariant differences between these pairs irrespective of their context. The present model treats accent as a fugacious feature, assignable to certain strong syllables (even to weak ones under special circumstances) but equally well deletable from them. It follows that in a context where either member of a noun-verb pair is de-accented, there will be, if both have two strong syllables, nothing to distinguish them phonetically: as

I *THOUGHT* he'd say pervert again,

I *HOPE* it was incline he said,

and so forth. If it turns out that people do regularly distinguish these pairs in such contexts (under appropriate conditions of testing such that the subjects are unaware of the contrasts involved) then this would point to an oversimplification in the model here put forward.

This issue is clearly related very closely to the belief that *blackbird* is distinguished phonetically from *black bird* even when the latter's accent is shifted, or when both are de-accented. The possibility of their remaining distinct under these circumstances seems quite remote. Alexandr Isacenko (personal communication) asserts the

same point for German, citing the following as a homophonous pair:

Ich KENNE die Krankenpfleger.

"I KNOW the male nurse."

Ich KENNE die kranken Pfleger.

"I KNOW the sick caretaker."

A related issue is whether differing "degrees" of stress on attributives before accented ("primary stressed") nouns expound syntactically distinct structures. The claim that they do has important consequences. There is not in this case any doubt of a real phonetic difference. It can be described in terms of the present model as realization versus nonrealization of the potential for accent on the attributive: as *the OLD MAID* versus *the old MAID*. But it is necessary to ascertain whether the difference in pronunciation of these two phrases bears an invariant relation to the syntax such that the first means an elderly female servant and the second a spinster (Trager and Smith 1951), in which case it deserves the name stress distinction; or whether the patterns are in free variation, or one is a stylistic variant, or the choice is conditioned by some nongrammatical factor such as "rhythm"; in which case they do not distinguish syntactic structures.

The claim that this is a grammatical stress distinction has recently been taken up by Chomsky and Halle (forthcoming). But what they purport to prove with the putative stress distinctions is a claim not only about the grammar of English but more especially about "universal grammar" -- viz. that phonological sound patterns are best explicated in terms of the principle of a transformational cycle of

ordered, recursive rules.

Thus if it were true that

(i) BILL'S british HISToRY teacher is PRETTY

were regularly distinguished from

(ii) BILL'S BRITish HISToRY teacher is PRETTY

in such a way that (i) meant that Bill has a pretty teacher who teaches British History, and (ii) that Bill has a pretty history teacher who is British; then this would be evidence in favor of the view that the stress cycle operates on the phrase and sentence level and in this case reduces the "degree of stress" on *British* twice in the sentence where *British* and *history* are joined at a lower node than *history* and *teacher*, and only once in the other sentence where they are joined at a higher node.

There is no doubt of "the reality of the syntactic patterns illustrated". What is at issue is the existence of a function which maps the separate syntactic structures onto distinct prosodic manifestations.

The alternative view is that taken by Jones (1918), Pike (1945a), Bolinger (1965b) and many others from at least the eighteenth century onward (Walker 1791, Jespersen 1909, van Draat 1910); a view which accounts for the absence of accent on *British* in (i) as due to the phenomenon of rhythmical variation, which is at least partially explicated in the rhythm rule above. On this view the two syntactic structures above collapse homophonously at the phonological level. Then the resulting single ambiguous string is input to the prosodic-feature assignment rules which, depending on some factor like "rate"

(stylistic variation) or random choice (free variation), assign one of the accentuations shown -- or one of the several other possible phonetic outputs such as

- (iii) BILL'S BRITish history teacher is PRETTY,
 but his {*FILiPINa*} history teacher is ABSolutely
 WORLD
 RAVishing.

That the syntactic ambiguity persists here is shown by the differential effect on its interpretation of the options in the second part of the sentence, which have been devised so that *Filipina* can only be governed by *teacher* whereas *world* must refer to *history*. Obviously the presence in (iii) of accent on *British* and its absence on *history* is the result of the antithesis between two categories of history (or of nationality). The point of interest is that the assignment of accent under "contrastive shift" correlates with meaning (whether contrast is considered to belong to the realm of syntax or semantics) whereas the assignment of accent by the rhythm rule, in (i) versus (ii), does not. For example, the rhythm rule operates in (iii) to produce *ABSolutely RAVishing* rather than *ABSOLUTELY RAVishing* -- cf. Bolinger's (1965) example *ábsolute pówér corrúpts absolútely*.

Catford (1966) postulates a special type of phoneme, called an open transition, or just transition, which accounts for some of the weak-strong syllable pairs, as discussed above, but which also presupposes distinctions between what are here considered to be homophonous weak pairs. He illustrates the three way contrast (close consonant transition, open transition, vowel) with examples which may be

transcribed, in terms of the features assumed here, as follows -- with slight emendations for R-keeping American dialects; viz. substitution of *kapa* for *copper* (since these words usually have the same first vowel in such dialects):

- A (i) Táke thě c^Fóp pārt/
 (ii) Táke thě c^Fóp ăpārt/
 (iii) Táke the k^Fapā pārt/
 B (i) Dád mē^Fděč ĭt
 (ii) Dád ămē^Fděč ĭt
 (iii) Dádďy mē^Fděč ĭt

where (ii) in each set illustrates the "transition". But it seems readily apparent that A(ii, iii) are homophonous in American English, and that B(ii, iii) are distinguished rather by the quality of the vowels than their durations (i.e. $i \neq \dot{i}$). If there is any difference in the syllable quantities in these two sentences it is probably to be looked for in the monosyllable *Dad* versus the penult of *Daddy*, and attributed to the location of word boundary and what Bolinger (1965b) calls "the peculiar status of monosyllables" (but cf. also Abercrombie 1964b).

On the other hand Catford's model does not distinguish systematically between weak-strong pairs. He asserts that

The *foot*, of course, is a stress-group, involving a fixed pattern of falling stress. Once this has been stated, stress need not be mentioned at all.

He then equates the Trager-Smith "weak stress" with his "transition," and T-S "tertiary stress" with other non-initial syllables of any foot. But as has been shown above, T-S tertiary (and also postnuclear

"secondary stresses") are equivalent to strong, unaccented syllables in the present model. And the weak-strong distinction can by no means always be equated with that between Catford's transitions and regular vowels. For example, in the following (transcribed a la Catford)

Do you/ know what there/ was/ yesterday?

there is obviously a weak syllable -- an expletive. But in

Do you/ know what was/ there/ yesterday?

where it is a locative adverb, it is clearly strong -- and, in this case, accented. But it is possible to have a reply to either question in which *there* is unaccented and in the same position, and distinguished only by its weak versus strong forms (transcribed as usual):

(i) NÓ/ whát wás thĕrĕ yĕstĕrdĕy/

(ii) NÓ/ whát wás thĕrĕ yĕstĕrdĕy/

This distinction cannot apparently be accounted for by Catford's analysis (i.e. its durational aspect as opposed to vowel quality).

This is perhaps enough to show that the prosodic model postulated here offers a reasonable balance between the assignment of excessive structure -- distinguishing what are in fact homophones -- on the one hand, and the conflation of grammatically relevant categories on the other. It appears to furnish an important part of the apparatus needed for relating orthographic texts to their prosodically, as well as orthoepically, correct renditions, and thereby to suggest a role for synthetic elocution in linguistic research.

Summary and outlook

The instrumentation of modern experimental phonetics -- especially the speech synthesizer, as a sort of vocal übermarionette -- supplies the lack of controlled means for rule testing which stumped the early elocutionists, and points the way toward a revival and extension of their concerns. Particularly it permits clear-cut discrimination of elocutionary *theory* -- which must explicitly relate orthographic (especially literary) texts to appropriate renditions of them -- from elocutionary *pedagogy*, which deals with how to teach human pupils to read aloud well. The pedagogical efficacy of the rules is irrelevant to the theory which comprises them.

The prosodic model put forward in this work assumes an input in which text words are represented in their *canonical orthoepic forms* (segmentals and "stress" patterns transcribed as for some ideal pronouncing dictionary) along with representations of their syntactic and semantic relations. Traditional IPA stress marks are reinterpreted as encoding two binary orthoepic distinctions: (1) *weak* versus *strong* syllables; and, among strong syllables, (2) *accentable* versus *unaccentable* ones.

The rules of the prosodic-feature-assignment component (or *prosody*), of which a fragment was presented above, are then claimed to predict all of the phonic variations due to the supra-sentential context. The postulated output features include two degrees of pitch-obtrusion, ACCENT and EMPHASIS. These can be upward or downward, as Bolinger has shown, but upward obtrusion is the unmarked case and all that is needed for renditions of straight factual prose, which turns

out to be a good diagnostic test of grammatical (as opposed to indexical) uses of prosodic features. Four intonational tunes are defined by the features CADENCE (nuclear fall) and ENDGLIDE (terminal rise). Additional features assign pauses and tessitural shifts (for handling e.g. parenthetical and quoted matter).

As discussed above, one obvious application of a working program for synthetic elocution would be in a high performance reading machine for the blind (cf. Cooper 1963). But research in this area has potentially a far wider application in man-machine communication. In even the most sophisticated time-sharing systems today the computer-user interface is usually just a typewriter console or the like. It is generally conceded that in many situations spoken interchanges would be preferable. As for man talking to machine, the present work has little bearing on the formidable task of automatic speech recognition (the so called "voice-operated typewriter" problem). But with respect to the machine's talking back, the state of the art of speech synthesis by rule already permits any message that can be typed out as text to be equally well encoded into intelligible speech. What is presently lacking is a prosodic component capable of producing in the machine a "just management of the voice". This task falls clearly within the purview of the computational elocutionist.

Perhaps one may adumbrate future generations of computers whose behavior will verge toward the human sufficiently for the provision of indexical features -- e.g. affective cues -- to become a useful or necessary adjunct to the prosodic component. But for some time at least it will suffice to give a talking computer the ability to dif-

ferentiate among non-homophonous parsings of a text segment, and to express the all-important (at least for English) relations of *contrast* and *synonymy*. These relations are not static characteristics of lexical items as such but result from complex interplay in the operation of two processes basic to all intelligence whether natural or artificial -- viz. *identification* and *discrimination*. For this reason the development of a model of the prosodic component of English is an enterprise with significant implications not only for communications engineering and the speech arts but also for linguistics and psychology.

BIBLIOGRAPHY

- Abe, I. (1955) Intonational patterns of English and Japanese, *Word* 2. 386-98.
- Abercrombie, D. (1949) rev. of K.L. Pike, *The Intonation of American English*, in *Review of English Studies* 25. 369-70.
- (1964) *English Phonetic Texts*, London: Faber.
- (1964b) Syllable quantity and enclitics in English, in *In Honour of Daniel Jones*, London: Longmans.
- (1965) *Studies in Phonetics and Linguistics*, London: Oxford.
- (1967) *Elements of General Phonetics*, Chicago: Aldine.
- Akin, Johnnye (1961) There's a strong wind blowing, *Western Sph. J.* 25. 36-42.
- Armstrong, L.E. and Ward, Ida C. (1931) *A Handbook of English Intonation*, 2nd ed., Cambridge: Heffner.
- Astle, T. (1784) *The Origins and Progress of Writing*, London.
- Austin, G. (1806) *Chironomia; or, a Treatise on Rhetorical Delivery*, London.
- Babcock, Maud May (1916) Interpretive presentation versus impersonative presentation, *Q. J. of Sph.* 2. 18-25.
- Bacon, W.A. (1960) The dangerous shores: from elocution to interpretation, *Q. J. of Sph.* 46. 148-52.
- Bakis, R. (1967) Prosodic feature synthesis for word-segmented messages, IBM Research Paper RZ 234.
- Bar-Hillel, Y. (1959) A demonstration of the non-feasibility of fully automatic high quality translation, Appendix IV to *Report on the State of Machine Translation in the United States and Great Britain*, Jerusalem; also in *Advances in Computers*, I, ed. F.L. Alt, New York, 1960.
- Beattie, J. (1788) *The Theory of Language*, London; 1st ed. 1783.
- Bell, A.M. (1888) *Hand-book of World-English*, New York and London.
- Bennett, E.M. ed. (1963) *Human Factors in Technology*, New York: McGraw-Hill.

- Berg, Jw. van den (1962) Modern research in experimental phonetics, *Folia Phoniatica* 14. 81-149.
- Bhimani, B.V. and Dolby, J.L. (1966) Acoustic phonetic transcription of written English, *Annual Report: Automatic Indexing and Abstracting*, Palo Alto: Lockheed Missiles and Space Co.
- Bolinger, D.L. (1947a) Comments on Pike's *Intonation of American English*, in *Studies in Linguistics* 5. 69-78.
- (1947b) American English intonation, *Amer. Sph.* 22. 134-6.
- (1954) Intonation as stress-carrier, *Litera* 1. 35-40.
- (1957) On certain functions of accents A and B, *Litera* 4. 80-9; reprinted in Bolinger (1965).
- (1958) A theory of pitch accent in English, *Word* 14. 109-49; reprinted in Bolinger (1965).
- (1961a) *Generality, Gradience, and the All-or-None*, 's-Gravenhage: Mouton.
- (1961b) Three analogies, *Hispania* 44. 134-36.
- (1964) Around the edge of language: intonation, *Harvard Educ. Rev.* 34. 282-96.
- (1965) *Forms of English: Accent, Morpheme, Order*, ed. I. Abe and T. Kanekiyo, Cambridge: Harvard Univ. Press.
- (1965b) Pitch accent and sentence rhythm, in *Forms of English* (Bolinger 1965).
- (1966) rev. of G. Faure, *Recherches sur le caractères et le rôle des éléments musicaux dans la prononciation anglaise*, in *Language* 42. 670-90.
- Bolinger, D.L. and Gerstman, L.J. (1957) Disjuncture as a cue to constructs, *Word* 13. 246-55.
- Bridges, R. (1921) *Milton's Prosody*, Oxford: Univ. Press.
- Brightland (1711) *A Grammar of the English Tongue*, London.
- Bronstein, A.J. (1960) *The Pronunciation of American English: An Introduction to Phonetics*, New York: Appleton-Century-Crofts.
- Brown, G. (1873) *The Institutes of English Grammar*, New York: Wood.
- Burgh, J. (1761) *The Art of Speaking*, London.

- Butler, C. (1634) *The English Grammar*, Oxford: Turner; reprinted, ed. A. Eichler, Wein: Halle, 1910.
- Bysse, E. (1710) *The Art of English Poetry*, London: Dolphin.
- Campbell, P.N. (1967) *The Speaking and the Speakers of Literature*, Belmont (Calif): Dickenson.
- Catford, J.C. (1966) English phonology and the teaching of pronunciation, *College Eng.* 28. 605-13.
- Chao, Y.R. (1932) *A Preliminary Study of English Intonation (with American Variants) and its Chinese Equivalents*, Peiping.
- Chatman, S. (1957) Linguistics, poetics, and interpretation: the phonemic dimension, *Q. J. of Sph.* 43. 249-56.
- (1965) *A Theory of Meter*, The Hague: Mouton.
- Cherry, C. (1961) *On Human Communication*, New York: Science Editions; 1st ed. 1957.
- Chomsky, N. (1965) *Aspects of the Theory of Syntax*, Cambridge: M.I.T. Press.
- Chomsky, N. and Halle, M. (forthcoming) *The Sound Pattern of English*.
- Chomsky, N., Halle, M. and Lukoff, F. (1956) On accent and juncture in English, in *For Roman Jakobson*, The Hague: Mouton.
- Classe, A. (1939) *The Rhythm of English Prose*, Oxford.
- Cockin, W. (1775) *The Art of Delivering Written Language; or, an Essay on Reading*, London.
- Coleman, H.O. (1914) Intonation and emphasis, *Miscellanea Phonetica* 1. 6-26.
- Cooper, F.S. (1963) Toward a high performance reading machine for the blind, in *Human Factors in Technology*, ed. E.M. Bennett, J. Degan, J. Spiegel, New York: McGraw-Hill.
- Craig, E.G. (1907) *The Actor and the Über-Marionette*, Florence.
- Crocker, L. (1929) The refrain in oratorical prose, *Q. J. of Sph.* 15. 25-9.
- Croll, M.W. (1929) *The Rhythm of English Verse*, Princeton: Univ. Press.
- Crystal, D. and Quirk, R. (1964) *Systems of Prosodic and Paralinguistic Features in English*, The Hague: Mouton.

- Crystal, T.H. (1965) Model of larynx activity during phonation, M.I.T. *Quar. Prog. Rep.* 78. 212-19.
- Curry, S.S. (1891) *The Province of Expression*, Boston.
- (1907) *Foundations of Expression*, Boston.
- Daines, S. (1640) *Orthoepia Anglicana: or, the first principal part of the English grammar*, London, 1640. Reprinted (facsimile) by Scholar Press Ltd., Menston, 1967.
- Darwin, C. (1873) *Expression of the Emotions in Man and Animals*, New York: Appleton.
- Davitz, J.R. (1964) *The Communication of Emotional Meaning*, New York: McGraw-Hill.
- Delattre, P. (1961) La leçon d'intonation de Simone de Beauvoir, étude d'intonation déclarative comparée, *French Review* 35. 59-67.
- (1962) Some factors of vowel duration and their cross-linguistic validity, *J. Acoust. Soc. Amer.* 34. 1141-2.
- (1965) *Comparing the Phonetic Features of English, German, Spanish, and French*, Heidelberg: Verlag; Philadelphia: Chilton.
- Delattre, P., Olsen, C., and Poenack, E. (1962) A comparative study of declarative intonation in American English and Spanish, *Hispania* 45. 134-61.
- Delattre, P., Poenack, E., and Olsen, C. (1965) Some characteristics of German intonation for the expression of continuation and finality, *Phonetica* 13. 134-61.
- Denes, P. (1959) A preliminary investigation of certain aspects of intonation, *Lg. and Sph.* 2. 106-22.
- Denes, P. and Milton-Williams, J. (1962) Further studies in intonation, *Lg. and Sph.* 5. 1-14.
- Dodsley, R. (1735) *The Art of Preaching*, London.
- Draat, P.F. van (1910) Rhythm in English prose, *Anglistische Forschungen* 29; Heidelberg.
- Dreyfus, H. (1965) Alchemy and artificial intelligence, Report #P-3244, RAND Corp., Santa Monica.

- Dvořáková, Eva (1964) On the English and Czech situational adverbs in functional sentence perspective, in *Brno Studies in English* 4, Praha; reproduced in P.L. Garvin, ed., *Papers on Complex Condensation, Functional Sentence Perspective and Word Order*, Canoga Park: Bunker-Ramo, n.d.
- Epstein, E.L. and Hawkes, T. (1959) *Linguistics and English Prosody* (*Studies in Ling. Occasional Papers* 7).
- Fant, C.G.M. (1962) Descriptive analysis of the acoustic aspects of speech, *Logos* 5.1. 3-17.
- Fant, G. and Martony, J. (1962) Speech synthesis, *Quarterly Progress and Status Report* 2. 18-24, Transmission Lab., R.I.T., Stockholm.
- Fintoft, K. (1965) Some remarks on word accents, *Phonetica* 13. 201-26.
- Fónagy, I. (1958) Elektrophysiologische Beiträge zur Akzentfrage, *Phonetica* 2. 59-85.
- Foster, J. (1763) *An Essay on the Different Nature of Accent and Quantity*, Eton: Pote.
- Friederich, W. (1958) *English Pronunciation*, trans. R.A. Martin, London: Longmans.
- Fries, C.C. (1925) Shakespearian punctuation, in *Studies in Shakespeare, Milton and Donne*, Univ. of Michigan publ. in Lang. and Lit. I, New York: Macmillan.
- Fritz, C.A. (1930) From Sheridan to Rush: the beginnings of English elocution, *Q. J. of Sph.* 16. 75-88.
- Fry, D.B. (1955) Duration and intensity as physical correlates of linguistic stress, *J. Acoust. Soc. Amer.* 27. 765 ff.
- (1956) Perception and recognition of speech, in *For Roman Jakobson*, The Hague: Mouton.
- (1958) Experiments in the perception of stress, *Lg. and Sph.* 1. 126.
- (1964a) The correction of errors in the reception of speech, *Phonetica* 11. 164-75.
- (1964b) The function of the syllable, *Zeitschrift für Phonetik Sprachwissenschaft und Kommunikationsforschung* 17. 215-21.

- Fuchs, E. (1935) *Zur Intonation des Englischen Fragesatzes; eine Experimentelle Untersuchung*, Bonn: Ludwig.
- Gaitenby, Jane H. (1967) Rules for word stress analysis: for conversion of print to synthetic speech (read at 74th meeting of Acoust. Soc. of Amer.) *J. Acoust. Soc. Amer.* 42. 1182(A).
- Geiger, D. (1962) Oral interpretation and the teaching of literature, *Sph. Teacher* 11. 202-7.
- Gleason, H.A. (1955) *An Introduction to Descriptive Linguistics*, New York: Holt.
- Gleitman, Lila R. (1961) Pronominals and stress in English conjunctions, *Lg. Learning* 11. 157-70.
- Glenn, S. (1960) Echo to the sense, *Western Sph. J.* 24. 88-93.
- Gove, P.B. (1964) "Noun often attributive" and "adjective", *Amer. Sph.* 34. 163-75.
- Gunter, R. (1966) On the placement of accent in dialogue: a feature of context grammar, *J. of Ling.* 2. 159-79.
- Haberman, F.W. (1954) English sources of American elocution, in *History of Speech Education in America*, ed. K. Wallace, New York: Appleton-Century-Crofts.
- Hadding-Koch, Kerstin (1961) *Acoustico-phonetic Studies in the Intonation of Southern Swedish*, Lund: Gleerups.
- Hadding-Koch, Kerstin and Studdert-Kennedy, M. (1964) An experimental study of some intonation contours, *Phonetica* 11. 175-85.
- Hale, L.L. (1941) Principles of James Rush as applied to interpretation, *Southern Sph. J.* (then *Bulletin*) 7.2. 43-5.
- Halle, M. and Keyser, S.J. (1966) Chaucer and the study of prosody, *College English* 28. 187-219.
- Halliday, M.A.K. (1963a) The tones of English, *Archivum Linguisticum* 15. 1-28.
- (1963b) Intonation in English grammar, *Trans. of the Philological Soc.* 143-69.
- (1966) Intonation systems in English (1963), in *Patterns of Language*, A. McIntosh and M.A.K. Halliday, Bloomington and London: Indiana Univ. Press.
- (1967) Some aspects of the thematic organization of the English clause, U.S.A.F. Project memorandum, Santa Monica: RAND.

- Hargis, D.E. (1957) James Burgh and *The Art of Speaking*, *Sph. Monographs* 24. 275-84.
- (1960) Interpretation as oral communication, *Central States Sph. J.* 11. 168-73.
- (1964) A French elocutionist: 1877, *Southern Sph. J.* 30. 24-35.
- Harris, Z. (1952) Discourse analysis, *Language* 28. 1-30.
- Haugen, E. (1956) The syllable in linguistic description, in *For Roman Jakobson*, The Hague: Mouton.
- Hill, A.A. (1961) Suprasegmentals, prosodies, prosodemes: comparison and discussion, *Language* 37. 457-68.
- Holmes, J. (1766) *The Art of Rhetoric Made Easy: or, the Elements of Oratory*, London; dedication dated 1755.
- Householder, F.W. Jr. (1957) Accent, juncture, intonation, and my grandfather's reader, *Word* 13. 234-45.
- Howell, W.S. (1959) Sources of the Elocutionist Movement in English: 1700-1748, *Q. J. of Sph.* 45. 1-18.
- Hultzen, L.S. (1939) Seventeenth century intonation, *Amer. Sph.* 14. 39-43.
- (1955) Stress and intonation, *General Linguistics* 1. 35-42.
- (1956) The poet Burns again, *Amer. Sph.* 31. 195-201.
- (1957) Communication in intonation: General American, in *Study of Sounds*, Tokyo: Phonetic Soc. of Japan.
- (1959) Information points in intonation, *Phonetica* 4. 107-20.
- (1964) Grammatical intonation, in *In Honor of Daniel Jones*, ed. D. Abercrombie et al., London: Longmans.
- Isacenko, A.V. (1965) Deutsche Satzintonationen als Konfigurationen zweierdiscreter Tonebenen (mit Demonstrationen von Tonbändern), *5th Phon. Sci. Cong. Proc.* 371-3.
- (1967) rev. of V.A. Kozhevnikov, L.A. Chistovich et al., *Speech: Articulation and Perception*, in *Foundations of Language* 3. 105-7.
- Jassem, W. (1952) *Intonation of Conversational English*, Wroclaw: Soc. of Sci. and Lett. (Series A, 45).

- Jespersen, O. (1907) John Hart's pronunciation of English (1569, 1570), *Anglistische Forschungen* 22. 123.
- (1909) *A Modern English Grammar on Historical Principles*, Heidelberg.
- (1933) Notes on metre, *Linguistica* 249-74; First printed in Danish, 1900.
- Johnson, Gertrude E. (1940) *Studies in the Art of Oral Interpretation*, New York, London: Appleton-Century.
- Jones, D. (1909) *Intonation Curves*, Leipzig and Berlin: Teubner.
- (1931) The 'word' as a phonetic entity, *Le Maître Phonétique* 3.46. 60-65.
- (1956) *An English Pronouncing Dictionary*, 11th ed., London: Dent; 1st ed. 1917.
- (1964) *An Outline of English Phonetics*, Cambridge: Heffer; 1st ed. 1918.
- (1966) *The Pronunciation of English*, Cambridge: Univ. Press; 1st ed. 1909.
- Jonson, B. (1640) *The English Grammar...for the Use of All Strangers...*, London; reprinted 1928, London: Lanston Monotype Corp. Ltd.
- Kaucher, Dorothy (1945) Try it again, *Q. J. of Sph.* 31. 47-55.
- Kenyon, J.S. and Knott, T.A. (1953) *A Pronouncing Dictionary of American English*, Springfield: Merriam; 1st ed. 1944.
- Keyser, S.J. (1967) The Linguistic basis of English prosody (read at Ling. Soc. of Amer. meeting, December, 1966).
- Kim, C-W. (1966) *The Linguistic Specification of Speech (Working Papers in Phonetics 5)*, Los Angeles: UCLA.
- Klinghardt, H. (1923) *Sprachmelodie und Sprechтакт*, Marburg: Elwert.
- Kozhevnikov, V.A. and Chistovich, L.A. et al. (1965) *Speech: Articulation and Perception*, Moscow-Leningrad. Translation JPRS 30. 543. Department of Commerce.
- Kreidler, C.W. (1963) The alternative question as a learning device, *Language Learning* 13. 65-75.
- Kruisinga, E. (1943) *The Phonetic Structure of English Words (Bibliotheca Anglicana 2)*, Bern: Francke.

- Kurath, H. (1964) *A Phonology and Prosody of Modern English*, Ann Arbor: Univ. of Michigan Press.
- Ladefoged, P. (1958) Syllables and stress, *Miscellanea Phonetica* 3. 1-14.
- (1959) The perception of vowel sounds, Ph.D. diss., Univ. of Edinburgh.
- (1960) The regulation of subglottal pressure, *Folia Phoniatica* 12. 169-75.
- (1962) Subglottal activity during speech, *4th Int. Congr. Phon. Sci.*, The Hague: Mouton.
- (1967a) Research possibilities in phonetics, *Annual Bulletin* 1. 31-4, Research Inst. of Logopedics and Phoniatrics, Univ. of Tokyo.
- (1967b) *Linguistic Phonetics (Working Papers in Phonetics 6)*, Los Angeles: UCLA.
- (1967c) *Three Areas of Experimental Phonetics*, London: Oxford Univ. Press.
- Ladefoged, P. and Vanderslice, R. (1967) The voiceprint mystique, *Working Papers in Phonetics* 7. 126-42.
- Lane, H. (1965) The motor theory of speech perception: a critical review, *Psychol. Review* 72. 275-309.
- Lawrence, W. (1952) The synthesis of speech from signals which have a low information rate, in *Communication Theory*, ed. W. Jackson, London: Butterworths.
- Lazier, G. (1966) Burke, behavior, and oral interpretation, *Southern Sph. J.* 31. 10-14.
- Lee, F.F. (1967) Automatic grapheme-to-phoneme translation of English (read at 73rd meeting of the Acoust. Soc. of Amer.) *J. Acoust. Soc. Amer.* 41. 1594(A).
- Lee, W.R. (1956) English intonation: a new approach, *Lingua* 4. 345-71.
- Le Faucheur, M. (c. 1702) *An essay upon the action of an orator; as to his pronunciation and gesture*, trans. anon., London: Cox, n.d. (see Howell 1959); 1st publ. in French, Paris, 1657. Orton, F. (1848) is a metrical version of the above translation, erroneously attributed to S. Foucher.
- Legouvé, E. (1877) *L'Art de la Lecture*, Paris; trans. E. Roth, *The Art of Reading*, Philadelphia, 1879.

- Loesch, Katharine T. (1965) Literary ambiguity and oral performance, *Q. J. of Sph.* 51. 258-67.
- Lehiste, Ilse (1965) Juncture, *Proc. 5th Int. Congr. Phon. Sci.*, Basel and New York: Karger; p. 172-200.
- Lehiste, Ilse and Peterson, G.E. (1959) Vowel amplitude and phonemic stress in American English, *J. Acoust. Soc. Amer.* 31. 428-35.
- (1961) Some basic considerations in the analysis of intonation, *J. Acoust. Soc. Amer.* 33. 419-25.
- Leiberman, P. (1967) *Intonation, Perception, and Language* (Research Monograph 38), Cambridge: M.I.T. Press.
- Levin, S.R. (1962) Suprasegmentals and the performance of poetry, *Q. J. of Sph.* 48. 366-72.
- Maack, A. (1958) Rules of the German Syllable Melody, *Phonetica* 2. 3-4, 199-219.
- Magdics, Klara (1963) Research on intonation during the past ten years, *Acta Linguistica* 13. 133-65.
- Malone, K. (1926) Pitch patterns in English, *Studies in Philol.* 23. 371-9.
- Mason, J. (1748) *An Essay on Elocution*, 2nd ed., London; 1st ed. same year.
- (1749) *An Essay on the Power of Numbers*, London.
- Mattingly, Alethea Smith (1954) The Mechanical School of oral reading in England 1761-1821, Ph.D. diss., Northwestern Univ.
- Mattingly, I.G. (1966) Synthesis by rule of prosodic features, *Lg. and Sph.* 9. 1-13.
- McIntosh, A. and Halliday, M.A.K. (1966) *Patterns of Language*, Bloomington: Indiana Univ. Press; London: Longmans.
- Meyer-Eppler, W. (1957) Realization of prosodic features in whispered speech, *J. Acoust. Soc. Amer.* 29. 104.
- Morton, J. (1964) A model for continuous language behavior, *Lg. and Sph.* 7. 40-70.
- Muyskens, J.H. (1931) An analysis of accent in English from kymograph records, *VOX* 17. 55-65.

- Newman, J.B. (1951) The phonetic aspect of Joshua Steele's system of prosody, *Sph. Monographs* 18. 279-87.
- (1953) The role of Joshua Steele in the development of speech education in America, *Sph. Monographs* 20. 65-73.
- Newman, S.S. (1946) On the stress system of English, *Word* 2. 171-87.
- O'Connor, J.D. (1955) The intonation of tag questions in English, *Eng. Stud.* 36. 98.
- (1957) Recent work in English phonetics, *Phonetica* 1. 96-117.
- O'Connor, J.D. and Arnold, G.F. (1961) *Intonation of Colloquial English, a practical handbook*, London: Longmans.
- Öhman, S.E.G. (1967) Word and sentence intonation: a quantitative model, *Quarterly Progress and Status Report* 2-3. 20-54.
- Öhman, S. and Lindqvist, J. (1965) Analysis-by-synthesis of prosodic pitch contours, *Quarterly Progress and Status Report* 4/1965, Sph. Transmission Lab., R.I.T., Stockholm.
- Olney, J. and Londe, D. (1966) A research plan for investigating English discourse structure, with particular attention to anaphoric relationships (report TM(L)-3256), Santa Monica: System Development Corp.
- Ortleb, Ruth (1937) An objective study of emphasis in oral reading of emotional and unemotional material, *Speech Monographs* 4. 56-68.
- Orton, F. (1848) *An Essay upon the action of an orator, or, his pronunciation and gesture; being the metrical version of an old English prose translation, from the French of Simon Foucher, of Dijon, who died at Paris, 1696*, Altrincham: Balshaw (original really by Michel Le Faucheur 1657, q.v.).
- Ostroff, A. (1954) New criticism in oral interpretation, *Western Sph. J.* 18. 39.
- Palmer, H.E. (1924) *English Intonation with Systematic Exercises*, Cambridge: Heffner.
- Parrish, W.M. (1951) The concept of "naturalness", *Q. J. of Sph.* 37. 448-54.
- (1957) Elocution -- a definition and a challenge, *Q. J. of Sph.* 43. 1-11.
- Peterson, G.E. (1955) An oral communication model, *Language* 31. 414-27.
- (1964) Requirements for a phonetic theory, *Phonetica* 11. 228-36.

- Pike, K.L. (1943) *Phonetics*, Ann Arbor: Univ. of Michigan Press.
- (1945a) *The Intonation of American English*, Ann Arbor: Univ. of Michigan Press.
- Pike, K.L. (1945b, with Traver, Aileen and French, Virginia) Step-by-step procedure for marking limited intonation with its related features of pause, stress, and rhythm, appendix to C.C. Fries, *Teaching and Learning English as a Foreign Language*, Ann Arbor: Univ. of Michigan Press.
- Pittenger, R.E. (1957) Linguistic analysis of tone of voice in communication of affect, *Psychiatric Research Reports* 8. 41-54.
- Pittenger, R.E., Hockett, C.F., and Danehy, J.J. (1960) *The First Five Minutes*, Ithaca: Martineau.
- Porter, E. (1824) *Analysis of Vocal Inflections, as used in Reading and Speaking*, Andover: Flag and Gould.
- Post, R.M. (1967) Oral interpretation and the Stanislavski Method, *Southern Sph. J.* 32. 180-7.
- Quintilian, M.F. (c. A.D. 90) *Institutio Oratoria*, trans. H.E. Butler, London: Heinemann; New York: Putnam.
- Reifler, E. (1955) The mechanical determination of meaning, in *Machine Translation of Languages*, ed. W.N. Locke, and A.D. Booth, Cambridge: MIT.
- Robb, Mary Margaret (1941) *Oral Interpretation of Literature in American Colleges and Universities*, New York: Wilson.
- Rush, J. (1827) *The Philosophy of the Human Voice*, Philadelphia (3rd and 4th eds. [1845, 1855] consulted).
- Sansom, C. (1960) Choral speaking, *Q. J. of Sph.* 46. 306-7.
- Schmeck, H. Jr. (1965) *The Semi-Artificial Man*, New York: Walker.
- Schubiger, Maria (1935) *The Role of Intonation in Spoken English*, St. Gall: Fehr.
- (1958) *English Intonation, Its Form and Function*, Tübingen: Niemeyer.
- (1965) English intonation and German modal particles -- a comparative study, *Phonetica* 12. 65-84.
- Scott, N.C. (1939) An experiment on stress perception. *Le Maître Phonétique* 67. 44-5.

- Sebeok, T.A. ed. (1960) *Style in Language*, New York and London: Wiley; Technology Press of M.I.T.
- Seigloch, A. (1939) *Die phonetischen Mittel der Deklamation bei W. J. Holloway in der Wiedergabe der Leichenrede des Mark Anton*, Berlin: Gruyter.
- Sharp, A.E. (1958) Falling-rising intonation patterns in English, *Phonetica* 2. 127-52.
- Sheridan, T. (1787) *Lectures on the Art of Reading*, 3rd ed., London: Dodsley.
- (1789) *A Complete Dictionary of the English Language...to which is prefixed a Prosodial Grammar*, 2nd ed., London: Dilly; 1st ed. (*A General Dictionary [etc.] to which is prefixed a Rhetorical Grammar*) 1780.
- Sledd, J. (1955) rev. of G.L. Trager and H.L. Smith Jr., *An Outline of English Structure*, *Language* 31. 312-35.
- (1956) Superfixes and intonation patterns, *Litera* 3. 35-41.
- (1957) rev. of H.L. Smith, Jr., *Linguistic Science and the Teaching of English*, *Litera* 4. 93-7.
- Sloan, T.O. ed. (1966) *The Oral Study of Literature*, New York: Random House.
- Stanislavski, K.S. (1936) *An Actor Prepares*, trans. Elizabeth R. Hapgood, New York: Theatre Arts Books.
- (1949) *Building a Character*, trans. Elizabeth R. Hapgood, New York: Theatre Arts Books.
- (1963) *An Actor's Handbook*, ed. and trans. Elizabeth R. Hapgood, New York: Theatre Arts Books; material cited first published in Russian c. 1924.
- Steele, J. (1775) *An Essay towards Establishing the Melody and Measure of Speech, to be Expressed and Perpetuated by Peculiar Symbols*, London; reprinted 1775 (v. Newman 1951) and 2nd ed. 1779 under title *Prosodia Rationalis; or, an Essay ... (etc.)*.
- Stene, Aasta (1954) Hiatus in English: problems of concatenation and juncture, *Anglistica* 3. 102.
- Stockwell, R.P. (1960a) rev. of Maria Schubiger, *English Intonation: Its Form and Function*, *Language* 36. 544-8.

- (1960b) The place of intonation in a generative grammar of English, *Language* 36. 360-7.
- Stockwell, R.P., Bowen, D.J. and Silva-Fuenzalida, I. (1956) Spanish juncture and intonation, *Language* 32. 641-65.
- Stevens, P. (1958) Edinburgh's artificial talking machine, *Univ. of Edinburgh Gazette* 20. (reprint) 4-8.
- Strong, W.J. (1967) Machine-aided formant determination for speech synthesis, *J. Acoust. Soc. Amer.* 41. 1434-42.
- Tassin, A. (1930) *The Oral Study of Literature*, New York: Crofts; 1st ed. 1923.
- Taube, M. (1961) *Computers and Common Sense*, New York: Columbia Univ. Press.
- Tempest, N.R. (1930) *The Rhythm of English Prose*, Cambridge.
- Tiffin, J. and Steer, M.D. (1937) An experimental analysis of emphasis, *Speech Monographs* 4. 69-74.
- Trager, G.L. (1950) Prosodic patterns and oral stress, *Science* 111. 582.
- (1951) Stress indicators, *Science* 113. 365.
- (1958) Paralanguage: a first approximation, *Studies in Ling.* 13. 1-12.
- Trager, G.L. and Smith, H.L. Jr. (1951) *An Outline of English Structure* (*Studies in Ling. Occasional Papers* 3); reprinted 1957, Washington, D.C.
- Uldall, Elizabeth T. (1960) Attitudinal meanings conveyed by intonation contours, *Lg. and Sph.* 3. 223-34.
- (1962) Ambiguity: question or statement? or "Are you asking me or telling me?" *Proc. 4th Int. Congr. Phon. Sci.* The Hague: Mouton.
- Uldall, Elizabeth and Anthony, J. (1962) The synthesis of a long piece of connected speech on PAT, *Proc. Sph. Communications Seminar* 2, Stockholm: R.I.T.
- Vanderslice, R. (1967) Larynx vs. lungs: cricothyrometer data refuting some recent claims concerning intonation and "archetypality", *Working Papers in Phonetics* 7. 69-79.

- Vanderslice, R. and Pierson, Laura Shun (1967) Prosodic features of Hawaiian English, *Q. J. of Sph.* 53. 156-66.
- Vandraegen, D.E. (1953) Thomas Sheridan and the Natural School, *Q. J. of Sph.* 20. 58-64.
- Vanvik, A. (1961) *On Stress in Present-Day English (Received Pronunciation)*, Bergen and Oslo: Norwegian Univs. Press.
- Venezky, R.L. (1966) Automatic spelling-to-sound conversion, in *Computation in Linguistics*, ed. P.L. Garvin and P. Spolsky, Bloomington: Indiana Univ. Press.
- Voelker, C.H. (1938) Loudness intonation in English, *Le Maître Phonétique* 62. 21-2.
- Walker, J. (1791) *A Critical Pronouncing Dictionary and Expositor of the English Language*, London: Robinson.
- Walker, O. (1659) *Some Instructions Concerning the Art of Oratory*, London.
- Wang, W.S.-Y. (1962) Stress in English, *Language Learning* 12. 69-77.
- Webster's Third New International Dictionary of the English Language* (1961) ed. P.B. Gove, Springfield (Mass.): G. and C. Merriam.
- Weinreich, U. (1960) Mid-century linguistics: attainments and frustrations (rev. of C.F. Hockett, *Course in Modern Linguistics*), *Romance Philol.* 13. 320-41.
- Whately, R. (1846) *Elements of Rhetoric*, 7th ed., London; 1st ed. 1828.
- Whitehall, H. (1956) From linguistics to criticism, *Kenyon Rev.* 18. 411-21; 1st printed *Kenyon Rev.* 13. 710-14 (1951).
- Whiting, F.M. (1954) *An Introduction to the Theatre*, New York: Harper.
- Wode, H. (1966) Englische Satzintonation, *Phonetica* 15. 129-218.
- Woehl, A. (1927) Goethe's rules for actors, *Q. J. of Sph.* 13. 243-64.