# The Linguistic Specification of Speech----

Chin-W. Kim

WPP #5/UCLA
December 1966

# THE LINGUISTIC SPECIFICATION OF SPEECH

Chin-Woo Kim

Working Papers in Phonetics 5

December 1966

University of California, Los Angeles

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

It need not be said that the present study would not have been possible had it not been my good fortune to have sat at the feet of Professors Peter Ladefoged and Robert P. Stockwell at the University of California, Los Angeles. It is from the former that I have acquired an invaluable training in physiological and instrumental phonetics of the Daniel Jones tradition, which is probably unattainable anywhere else in the country. That my view of phonetic theories is essentially based on physiological phonetics is but a logical consequence of his abundant influence on me, as is to be seen in almost every page of this monograph. From the latter, I learned generative phonology and syntax. Knowledge of these subjects made me see the issues from a perspective and unbiased angle. In view of the fact that phonology or linguistic phonetics as it is called by Ladefoged is not to be studied independent of higher level abstractions of grammar, the value of Stockwell's teaching is inestimable.

Next my gratitude goes to the American Council of Learned Societies whose generous award of a predoctoral fellowship made it possible to concentrate on writing the dissertation without having to worry about living, and to the National Institute of Health (Grant 4-443850-24147) which in part supported the final stage of this study.

My acknowledgements are also due to Professor Paul Schacter, and my colleagues Kalon Kelley and John C. McKay who kindly read the manuscript and gave me valuable comments and criticisms. I am solely responsible, however, for any mistakes and faulty arguments in the monograph.

*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.

# I

# THE PHONOLOGICAL COMPONENT

"Whenever we describe a language, at some point we have to talk about the sounds." (the opening sentence of Ladefoged's forthcoming *Linguistic Phonetics*) There is no doubt that sound is a fundamental and inseparable part of human language, although communication can occur with written codes or visual signals. For this reason no complete linguistic description can dispense with the description of sounds, i.e., the "expression" plane of language in glossematic terms. De Saussure analogized this aspect of language as two sides of a sheet of paper:

> Language can also be compared with a sheet of paper: thought is the front and the sound the back; one cannot cut the front without cutting the back at the same time. (De Saussure 1915. English translation by Baskin 1959, p. 113)

The generative grammar of a transformational model (Chomsky 1957; Lees 1960; Katz and Postal 1964; Chomsky 1965) recognizes this fact of language, and establishes the phonological(P) component as one of the three fundamental components of a grammar, the other two being the syntactic component and the semantic component.

The syntactic component generates, via rules, a syntactically well-formed string of formatives, whose meaning is interpreted by rules of the semantic component. The phonological component connects the string into an utterance, that is, into speech sounds. The rules of the P component apply to derived strings, i.e., strings derived after the application of transformational rules. This kind of phonology as a component of a generative grammar is called generative phonology; but since Halle is largely responsible for its development just as Chomsky is largely responsible for the formalization of the syntactic component, and since there may be other types or theories of generative phonology (e.g., Lamb 1964; Ladefoged, forthcoming), let us call this particular type of phonology Hallean phonology.

We will briefly survey the design of Hallean phonology since this monograph proposes another theory with some modification and extension of the former.

The present form of Hallean phonology establishes two levels of phonological abstraction: the level of systematic phonemics and the level of systematic phonetics. The representations at the first level are purely abstract and relational markers for the designation of morphemes as they appear in the dictionary in such a way that the base

form of the same lexical derivatives will have the same phonological representation. For example, each of the pairs of sounds [ay] of *divide* and [i] of *division*, and [d] of *divide* and [ʒ] of *division* will have the same representation /divid/. On the other hand, the representations at the second level characterize phonetic differences in actual utterances so that allophonic differences will have different representations at this level.

As is well known, Hallean phonology is to be represented, not with phonemes, but with Jakobsonian distinctive features (DF) (Jakobson, Fant, and Halle 1951 [henceforth, *Preliminaries*]; Jakobson and Halle 1956 [henceforth, *Fundamentals*]). One graphical consequence of this DF representation of a morpheme is a "matrix" (hence the terms: systematic phonemic or classificatory matrix, systematic phonetic or descriptive matrix), in which segments are, to keep the convention of left-to-right writing, designated by columns and features by rows, e.g.,

| | y | æ | n | k | i | . . . |
|---|---|---|---|---|---|---|
| Vocalic | - | + | - | - | + | |
| Consonantal | - | - | + | + | - | |
| Diffuse | + | - | + | - | + | |
| Grave | - | - | - | + | - | |
| Nasal | - | - | + | - | - | |
| Continuant | + | + | - | - | + | |
| Voice | + | + | + | - | + | |
| . | | | | | | |
| . | | | | | | |
| . | | | | | | |

Table 1. An example of a distinctive feature matrix in which columns designate segments and rows features.

Advantages of the use of DF in terms of simplicity in phonological desctiption and as a feasible frame of universal phonetics have appeared in a number of Halle's writings (Halle 1959, 1961, 1964a, 1964b; Chomsky and Halle 1965), and since more will be said about this later (See Chapter IV), I abstain here from a further discussion of DF, and go on.

Hallean phonology is the first of its kind that is explicitly designed and formalized in order to meet the descriptive and explanatory adequacy of a grammar, the highest goal of linguistic description as was set by Chomsky (Chomsky 1957, 1962, 1965). Although the theory will undoubtedly reveal several local deficiencies and inadequacies as more extensive investigations are made (of which this monograph claims to be one), the author accepts Hallean phonology as an essentially correct and

adequate phonological theory as a component of a generative grammar, and presents this monograph in the hope of enriching and improving the theory.

This attempt is made in three directions in this monograph:

(1) extension
(2) modification
(3) application

*Extension* is proposed in the domain or scope of phonological desctiption, that is, in the end point of phonetic specification by rules. This is discussed in Chapters II and III.

*Modification* is proposed in several aspects of the DF theory, or, in more general terms in the framework of universal phonetics. This and some other issues in the theory of phonology will be discussed in Chapters IV, V, and VII.

*Application* of the extended theory will be made on the synthesis of English vowels by rules. It is discussed in Chapters VI, VIII, and IX.

The concluding Chapter X, presents a summary and some residual issues.

# II

# THE SCOPE OF PHONETIC SPECIFICATION

We ask now, what is the precise nature of the phonological representation at the systematic phonetic level? Is this representation essentially a phonemic transcription with added detail of distributionally determined allophones? Or is this representation supposed to be minutely detailed to the end like a photograph? Obviously, the latter measure is impossible and unnecessary. It is impossible because no description or formalization can give a photographic reproduction of the continuous and infinitely varying nature of speech. It is unnecessary because language is a code system involving a relatively high degree of redundancy and most of the minute details are redundant and irrelevant to speech perception and understanding. On the other extreme, it is also evident that a phonemic transcription with allophonic differentiation is not detailed enough to give the fine but systematic phonetic differentiation that is found in speech that we want our phonetic description to capture. Secondly, we want our phonetic theory to show the different phonetic characterization of, say, two different languages which have a series of speech sounds that may be said to be the same from a phonemic point of view but are nevertheless phonetically different in a systematic way.

I give an example or two of each case.

In a dialect of American English, the phonetic difference between *rider* and *writer* is said to be in the length of the first vowel rather than in the stops themselves (cf. Chomsky 1964, p. 96). For those speakers who make the distinction in this way, it is presumably because a general phonetic feature of English that makes vowels shorter, other conditions being equal, before voiceless consonants than before voiced consonants is somehow still kept even when the medial /t/ becomes a voiced flap (cf. also Joos 1942). An allophonic transcription (distributionally determined, not arbitrarily detailed) will not be able to characterize this kind of phonetic differentiation. One might say that the length difference in *rider* and *writer* may be allophonically specified as [ray:dɣ] and [raydɣ] respectively. But there are other factors, e.g., tenseness of vowel, the manner of articulation of the following consonant, etc., that influence the length of vowel systematically, so that the interaction of these factors makes a given vowel length range anywhere from 100 to 400 msec, and it is impossible to transcribe this variation allophonically, except in an *ad hoc* way. (see Chapter VIII for the details)

Both English and French have a series of voiceless stops and corresponding voiced stops. (The much argued question whether the DF in this case is Voicing or Tenseness is irrelevant here.) Since within each language the voicing difference gives a sufficient phonemic opposition, I presume that the stop sounds of both languages will be marked exactly the same as to their feature specification, i.e., [-Voc, +Cons,

-Cont, -Nasal, -Strd, αVoice] (α is a variable implying either - or +). Yet, the substitution of one for the other, i.e., full voicing of English /b/ series and non-aspiration of English /p/ series (initially), or devoicing of French /b/ series and aspiration of French /p/ series, may render utterance of both languages unintelligible.

We want our phonetic representation to be not merely a phonemic description of the oppositions within a language, but also to be able to show how the same feature of one language differs phonetically from that feature in other languages (e.g., the case of English and French stops). It should also be able to capture systematic phonetic differences among dialects and idiolects (e.g., the case of *rider* vs. *writer*). Then it is clear that the apparatus in which the phonetic output is represented in broad allophonic transcriptions is not adequate for our purpose (cf. Kim 1965; Ladefoged, forthcoming).

Ideally, we would want a frame that will give a phonetic representation that is detailed enough to capture the differentiation mentioned above but not so detailed as to include a redundant specification. The question is then how detailed and specific should the P rules be that are to convert the abstract representation into real sound. In this chapter, we will explore this question of the scope of phonetic specification.

With regard to this question, Householder once flatly stated that

> the terminal alphabet of the phonological grammar should be, in the main, phonemic, including only such allophones as are distinguished by the native speaker, but not the fine phonetic distinctions required for exact international communication. (Householder 1965, p. 29)

This is a representative of the view that the mechanism of the phonological component is satisfactory if it is capable of performing an essentially phonemic characterization of a language. But we have already seen that we want our theory of speech specification to be more capable than this. Besides, Householder's own criterion for phonetic specification (i.e., of "only those allophones distinguished by the native speaker") immediately breaks down, since, earlier in the same place, discussing Chomsky's 'descriptive adequacy,' he calls "the linguistic intuition of the native speaker" as "too shifty and variable to be of any critical value." (Householder 1965, p. 15. See also his fn. 2)

At the other extreme, Halle argues that in the case when a phoneme, which resulted from the merger of two phonemes, behaves like its historical antecedents in the phonological system of the language, then we should postulate that, e.g., "/ɑ/ and /æ/ remained distinct entities even though every /ɑ/ was actualized phonetically as [æ]" and therefore

the distinction is not present in any utterance (Halle 1964b, p. 351).[1] Still, the following statement clearly indicates that the specification goes beyond the phonemic level:

> . . . the rules supply values to *nonphonemic* features, change the values of certain features, and assign a *phonetic interpretation* to the individual rows of the matrix. (Halle 1964a, p. 333) [emphasis mine]

In the "phonetic interpretation," Halle states, the features no longer have to be binary but they may have a numerical representation:

> The phonological component will include rules replacing some of the pluses and minuses in the matrices by integers representing the different degrees of intensity which the feature in question manifests in the utterance. (ibid.)

Thus, as is given by Halle, the fact that the English [ʌ] as in *pup* is less grave than English [u] as in *poop* will be embodied in a rule replacing the plus for the feature gravity by a higher integer in the vowel in *poop* than in the vowel in *pup*. This non-binariness of features and the use of the "degree" of features in the phonetic matrix have also been expressed by Chomsky:

> The entry in the i-th row and j-th column [in a phonetic matrix] indicates whether the j-th phone of the generated utterance possesses the i-th feature or the degree to which it possesses this feature. Classificatory distinctive features are by definition

---

[1]It is not altogether clear, however, why this should always be the case. Suppose that "a historical behavior" was that consonants were palatalized (say [t] to [c]) before /æ/ but not before /a/, and that this behavior is still kept even after the merger of /a/ to /æ/. Everything being equal, what we have here is

> two vowel phonemes /æ/ and /a/ (for which there is no synchronic distinction)
> a consonantal phoneme /t/
> a palatalization rule

An alternative solution is to postulate only one vowel phoneme /æ/ but two consonantal phonemes /t/ and /c/ (or in DF terms, Plain and Sharp). It seems that the latter solution offers a simpler grammar (i.e., one less P rule, as there is no need for palatalization rule), and reflects reality more truly (i.e., phonemic distinction maps directly into phonetic distinction). For a discussion on the extent to which a diachrony may be included in a synchronic description, see Stockwell 1964 and 1966.

> binary; phonetic features may or may not be. (Chomsky 1964, p. 86)

Similarly, Postal remarks:

> On the lowest level derived by the rules of phonology, the elements are also bundles of features, but here the features are not binary and *their values are best represented with numbers.* (Postal 1964, p. 279. fn. 40) [emphasis mine]

All three quotations in the above point in the same direction: features become non-binary at the phonetic level and the best way to represent them is to state the different degrees of feature manifestations in terms of integers, i.e., numbers. Considering the continuous and varying nature of speech which will most likely defy the description merely in terms of several binary oppositions, the above statements are not surprising. There is yet no work, however, that shows explicitly what the form of these phonological rules using numbers would be like. And furthermore, since the numerical specification can be either crude (e.g., "Korea has a population of 30 million.") or infinitely detailed (e.g., "the value of $\pi$ is 3.14159...."), the statement that a numerical representation or a degree specification is the final form of phonological rules still does not provide an agreement as to the extent of the detail of the phonetic specification. Halle once said that "in principle the phonological rules should be extended to the point where all distinctive features of all segments are specified." (Halle 1959, p. 44) Taken literally, this means that the phonetic matrix is simply derived from the phonemic matrix by filling in all the blanks which were left unspecified because of redundancy. That this is not enough is obvious, and I am sure, judging from Halle's later writings, that Halle himself no longer holds this view. Otherwise, "replacement of some of the pluses and minuses by integers," etc. would not be necessary.[2]

---

[2]That Halle does not regard a description of phonemic oppositions sufficient is clearly seen by the following statement:

> Languages differ also in the way they handle nonphonemic features or feature combinations. For some of the nonphonemic features there are definite rules; for others the decision is left up to the speaker who can do as he likes. For example, the feature of aspiration is nonphonemic in English; its occurrence is subject to the following conditions:
>
> All segments other than the voiceless stops [k], [p], [t] are unaspirated.
>
> The voiceless stops are never aspirated after [s].
>
> Except after [s], voiceless stops are always aspirated before an accented vowel.
>
> In all other positions, aspirations of voiceless stops is optional.

A complete grammar must obviously contain a statement of such facts, for

Recently, Ladefoged explicitly set the domain of the phonetic specification as follows:

> There are three stages which a theory of phonetics must be capable of handling. First, it must permit the oppositions within each language to be specified; this is what Chomsky calls systematic phonemics. Secondly, it must provide a way of accounting for the particular characteristics of each language; this might be systematic phonetics. Thirdly, it must lead to the specification of actual utterances by individual speakers of each language; this is physical phonetics. Linguistic descriptions which do not meet all three of these requirements are apt to be trivial. In practice the first step involves allocating sounds to contrasting categories, the second to designating relative values of each category, and the third to interpreting these values in terms of measurable units. (Ladefoged, forthcoming)

What is to be noted here is the significance of the level of physical phonetics. Contrary to the traditional view that it is irrelevant to linguistics, Ladefoged suggests that it is part of the proper domain of the phonological statement in terms of rules. The similar view has also been expressed by Kim (1965) and by Fromkin (1965). She, stressing the relevance of physical phonetics in linguistics, asserts that, in order for the ultimate phonological features of a grammar to be tested, they must be provided with empirical content, and that, therefore, the physical data is the end-point of phonetic specification.

Little attention has been paid to this aspect of linguistic description. Even the most generous linguists, not to mention Glossematicians who threw phonetics out of the window of the linguistics library (cf. Hjelmslev 1943/1961), hardly gave any thought beyond the allophonic or systematic phonetic level. Chomsky's "Current issues in linguistic theory" (1964) contains only one paragraph regarding this matter, half of which I quote:

> Physical phonetics is . . . Bloomfield's 'mechanical record of the gross acoustic features, such as is produced in the phonetics laboratory;' its status is not in question here and no further attention will be given to it." (p. 92)

The similar attitude has prevailed among linguists who somehow assumed that everything beyond allophones is automatically or contextually determined, that the transitional phenomena between phones are physiologically conditioned, and therefore irrelevant to linguistic description, and that, since realizations of phonemes or features are

---

they are of crucial importance to one who would speak the language correctly. (1964a, p. 330)

rather relative than absolute, it didn't matter where a phone is placed in the phonological space as long as it kept its relative distance from other phonemes. Thus, for example, a segment specified as Long would be of any length, theoretically from the time it takes for an electric currect to travel a foot of wire to the period of an ice age, as long as it is longer than a segment specified as Short.

This sort of attitude is the consequence of thinking that matrices of systematic phonetics or phonemic transcriptions (in the taxonomic sense) with the added detail of distributionally conditioned allophones are the final derivations of the rules of the phonological component. I believe that we need to go one step further beyond this level in order for the P rules to produce, from the phonemic strings, speech stretches that are acceptable to the native speakers as agreeable and unjarring. If we follow a Firthian maxim that "part of the meaning of an Englishman is to sound like an Englishman," we must go a step further beyond the level of systematic phonetics and provide it with physical contents. Only in this way, are we able to validate the features or ultimate linguistic units at that level, and only this constitutes a workable and testable phonology. (We will return to the question of linguistic validation in the next chapter.) To quote Fischer-Jørgensen:

> . . . the chief objective of linguistics is the abstract functional system . . . but the units of this abstract system can only be identified if the 'substance' is taken into account . . . once the system is established, the linguist will be interested in its actualization in the speech act . . . . (1961, p. 112)

Ladefoged referred to this level of "actualization" as "physical phonetics" (cf. quotation on p. 8). But this term has an unfortunately misleading connotation in that it has an implication of dealing with "gross acoustical features" (cf. quotation from Chomsky p. 8). But as long as the "actualization" is rule-governed behavior, it is also "systematic," and since the process of actualization is in a way equivalent to speech synthesis,[3] I propose here the term *systematic synthesis*, in lieu of physical phonetics, meaning synthesis-by-rule as opposed to synthesis-by-art.

Summarizing what has been stated so far in this chapter, there are

---

[3]One might argue that the actualization of the system differs from speech synthesis in a non-trivial way in that the input to a synthesizer need not necessarily be functionally relevant linguistic units, whereas the actualization of a system by the speaker is primarily concerned with abstract functional units. But if we concern ourselves only with the process itself, the speaker's operation to produce audible utterances from some sort of phonetic representation may be equated with the operation of a speech synthesizer to produce sounds from some sort of an input. See also footnote 6.

three levels of phonetic description that an adequate phonological theory must be capable of handling:

(1) the level of systematic phonemics - categorization of (morpho)-phonemic oppositions within a language

(2) the level of systematic phonetics - characterization of phonetic differences and similarities among languages

(3) the level of systematic synthesis - specifications of physical phonic substance of utterances as are spoken by an ideal speaker of the language.

We now must ask what are the criteria by which we may classify, not in an *ad hoc* way but with a linguistically significant motivation, a given phonological phenomenon as an object of description at each level. For instance, in traditional taxonomic phonemics, two major criteria for identifying two or more allophones as belonging to one and the same phoneme were complementary distribution and phonetic similarity. What would be such criteria in the case of the above three levels?

At the first level, perceptual discrimination by the native speakers would be a major criterion. That is, if a difference in a pair of sounds in two otherwise identical strings of phones is held to be responsible for their being identified by the native speakers as designating two different phenomena in reality, say [p] and [b] in [pul] and [bul], then the two sounds would qualify as two distinct phonemes in the language.

At the second level, the question is on what basis do we declare two sounds in two different languages as the same or different? Earlier we discussed a difference between English and French stops on the assumption that they belong to the same universal or interlingual phonological category "stop." This assumption is correct only if we take the plosive nature of the sound as the sole qualification to make it a stop. If aspiration has to be marked, then we would have to say that English stops and French stops are not the same but different kinds of sounds. Similarly, if we declare that a click and a stop are two different kinds of sounds, they are so only if we assume that the difference in the airstream mechanism during the articulation of speech sounds renders them as different, although the two sounds are the same in all other respects. Returning to the question posed at the beginning of this paragraph, what would be a general criterion of the phonetic sameness and difference in the cross-linguistic case? Since this question is closely tied with the theory of universal phonetics, we will deal with this question in Chapter IV, where the concept of Jakobsonian distinctive features as a frame of universal phonetics is reviewed and an alternate frame is proposed.

At the third level, the question is: by what criteria do we decide which phonetic phenomena are relevant to the description, i.e., what allophones are subject to rule-description and what are not?

Intuitively, we may say that phonetically very similar allophones do not need separate specification, but those that are phonetically not similar are subject to a separate description. Thus, for example, it

seems reasonable to say that [ʔ] as an allophone of /t/ as in *mountain*, *captain*, etc. should be separately specified by a rule of the form:

(1) /t/ → [ʔ] / $\acute{V}$__n #

since [ʔ] is phonetically not similar to [t] at least from an articulatory point of view, but that a laterally released [tˡ] need not be, because it is phonetically very similar to a centrally released [t].

The problem, however, still remains, since we do not know where to draw an exact, non-arbitrary boundary line between phonetic similarity and non-similarity. It is a familiar, much argued problem in taxonomic linguistics, as phonetic similarity is one of the major criteria in phonemic analysis (cf. Pike 1947; Austin 1957). To be more specific about the difficulty in establishing phonetic similarity as a criterion, we ask ourselves a few questions: Is or is not an unreleased [t°] as in [kut°] 'solid' in Korean phonetically similar to a released [t] as in [kut] *coûte* 'costs' (Present 3rd Person Singular) in French? Is or is not a dental [t] as in [t̪or] 'descend' in Temne similar to an alveolar [t] as in [tɔr] *tore* in English? Is or is not an aspirated [pʰ] as in [pʰul] 'grass' in Korean phonetically similar to an unaspirated [p] as in [pul] *poule* 'fowl' in French? In the case of vowels which are inherently of more continuous nature than consonants, the question of a boundary line of (non)similarity is all but undecidable and arbitrary if decided.

Thus, we seek an answer from a different point of view.

In 1961, Wang and Fillmore published an article titled "Intrinsic cues and consonant perception." In that paper, the authors distinguished two kinds of allophones: "intrinsic" and "extrinsic."

> In most phonetic discussion, it is useful to distinguish those secondary cues which reflect the speech habits of a particular community from those which reflect the structure of the speech mechanism in general. The former is called *extrinsic* and the latter, *intrinsic*. (p. 130)

Ladefoged (1965; and forthcoming) adopted and elaborated these terms, giving a slightly new definition. He defines intrinsic allophones as those allophonic variations "which are due to the partial overlapping of the articulations of adjacent phonemes"[4] (Ladefoged, forthcoming),

---

[4]This definition, however, seems to be not well-defined. That is, there seem to be cases where the definition needs a more precise explication. For example, the voicing-through of voiceless consonants, or the assimilation of nasals to the homorganic position of the following consonants, found in many languages, can be regarded as intrinsic features by the above definition, as the voicing-through and the nasal assimilation

and extrinsic allophones as those which are due to "the effect of other higher level units such as junctures, stress, or vowel harmony marks" (ibid.) but not by means of conjoining rules as in the case of intrinsic allophones.

To give a few examples: The difference between the advanced [k̟] in *key* and the retracted [k̠] in *car* would be due to the influence of the neighboring sounds, the following vowel in this case, and therefore, they are intrinsic allophones. But the difference between the "clear" [l] in *leaf* and the "dark" [ɫ] in *feel* cannot be explained in the same way. There is no phonetic feature of neighboring sounds that can predict the difference. Hence, they are extrinsic allophones. Similarly, the difference in the amount of voicing in two [r]'s in *drew* and *true* is totally due to the contextual influence of the preceding consonant (intrinsic allophones), but the difference between two [r]'s in *reed* and *deer* are not explicable in the same manner. There is no inherent articulatory reason or inborn physiological restriction for such a difference (extrinsic allophones).

We find this type of allophonic distinction very convenient and well motivated, and I propose this to be a criterion by which a decision is to be made as to whether or not an allophone should be rule-specified at the level of systematic synthesis.

This proposal is different from Ladefoged's view in one nontrivial respect. Ladefoged views that both intrinsic and extrinsic allophones are language-dependent and that therefore the rules of the systematic synthesis must include descriptions of both kinds of allophones:

> Both the ideal positions in a table of values (extrinsic allophones) and the conjoining rules for specifying intrinsic allophones are language dependent. There are many linguistic universals; but, for example, the effect of neighboring vowels on the articulation of velar stops is not one of them. This may be seen by comparing English and French. In both languages the initial stops vary in much the same way in pairs such as English "key - car" and French "qui - car;" but there is a much greater difference between the final stops in French "pique - pâque" than there is between those in English "peak - pock." The conjoining rules for English and French have to be different. (Ladefoged, forthcoming)

But as was stated in the preceding paragraph, I exclude intrinsic

---

are certainly due to the partial overlapping of the articulations of adjacent phonemes, although few would argue that those phenomena are due to inherent physiological constraints. (I owe this observation to Paul Schachter and Kalon Kelley.) A better definition may be in terms of universal versus non-universal allophonic variations. See below.

allophones from the description by rules of systematic synthesis. The reasons are threefold:

(1) I maintain the view that intrinsic allophonic features are universal, not language-specific. Considering the physiological and anatomical structure of the human vocal tract which is remarkably the same for all people, regardless of their language, this view is not at all unlikely. A conjoining rule like that of Lindblom's (1964) and Öhman's (1966) is formulated in terms of the target positions and the degree by which the targets are missed through the influence of the adjacent items. In this form of a conjoining rules, I am inclined to believe that, if two respective phonemes and the neighboring items in two or more different languages have the same ideal targets, and furthermore, if all other factors are the same, e.g., the rate of speech, the degree of stress, etc., then the degree of missing the target would be equal in both languages by virtue of the fact that the structure and the potential behavior of the human vocal mechanism is universally the same. I believe this is not an unreasonable assumption. It is difficult to imagine that in two languages that have the same target positions for, say, /a/ and /k/, the syllable /aka/ would yield the different coarticulation or transitional phenomena in two languages beyond individual variations. Ladefoged's example of French *pique - pâque* as a pair having a different degree of vocalic influence on /k/ as compared to the English pair *peak - pock* is not too convincing. Granted that the two languages have the same target position for /k/, it is worth noting that English /i/ (F1 - 275 cps, F2 - 2150 cps --- Lehiste 1964, p. 25) and French /i/ (F1 - 250 cps, F2 - 2600 cps --- Malmberg 1963, p. 49) have probably different target positions.

The different degrees of coarticulation or transition in different languages may be due to one or both of the following:

i. Corresponding phonemes in different languages, e.g., /k/ or $L_1$ and /k/ of $L_2$, /e/ of $L_i$ and /e/ of $L_j$, etc., may have different ideal target positions. The difference due to this factor will be automatically reflected, e.g., in Öhman's conjoining rule.

ii. The difference in the total phonemic structure in different languages may cause a difference in the degree of the influence of coarticulation. For example, it is not difficult to imagine that in a language in which there are palatal stops as well as velar stops (both phonemic) there would be less influence of front vowels on /k/ than in a language where there is no phonemic palatal stop, just as /a/ is likely to have less allophonic variations in a seven-vowel (i, e, æ, a, ɔ, o, u) language than /a/ in a three-vowel (i, a, u) language, such as Tausug, where allophones of /a/ may range from [æ] to [ɔ]. I believe that this "freedom" factor can be formalized and be incorporated in the conjoining rule without impairing the universal character of the rule.

If the above assumption is correct, then we may place the conjoining rule that describes the intrinsic allophonic phenomenon as a universal meta-rule and exclude it from rules of systematic synthesis.[5]

---

[5]The statement that intrinsic allophonic features are universal and the argument that languages may differ in the degree of coarticulation

This consideration that intrinsic allophonic features are universal leads us to an open question whether or not the set of intrinsic allophones defined in terms of the physiological constraints of articulation and the set of universal allophones defined in terms of distributional facts are coextensive. If there are cases in which a particular sound x always (universally) becomes allophone x' in the context y, and yet no reason can be found for this in terms of the intrinsic/extrinsic criterion, then the two sets will prove to be non-coextensive. In other words, there must exist universal allophones which are not intrinsic. The opposite case, i.e., the existence of non-universal intrinsic allophones is assumed by Ladefoged but is not so assumed here. Whatever the case may be, it is important to bear in mind that it is useful and well motivated to distinguish two kinds of allophones, one statable as meta-rules, and the other generatable only by rules of systematic synthesis of a given language.

(2) Rules of intrinsic allophones are not rules in the generative sense. That is, these rules do not generate allophones in the same sense that syntactic rules generate terminal strings. Rules of extrinsic allophones generate new allophones which are not predictable and hence not describable except via specific rules, but intrinsic allophones are predictable given the target values of phonemes that are adjacent to one another and a few other constants such as the "freedom" factor, the time factor, etc. In this sense, intrinsic allophones are best compared with universal redundancy rules which also are non-generative, predictable, and redundant. (It was precisely this non-generative character of redundancy rules that led Stanley (1966) to treat them as conditions rather than as rules.)

(3) I tend to think that intrinsic allophones are in general perceptually irrelevant but extrinsic allophones are not. For example, English hearers may or may not notice the intrinsic allophonic difference between two [r]'s in *drew* or *true*, but they are more likely to notice the difference between two [r]'s in *reed* and *deer*. For instance, we may picture an Englishman who, upon hearing a person uttering *deer* with [r] of *reed*, asks: "Where are you from, from Germany?"

The question of the threshold of discrimination in speech perception

---

depending on their specific phonemic patterns may seem contradictory and irreconcilable. They are not. It is to be noted that the conjoining rule would yield the same value if the constant k in the rule denoting the "freedom" factor is the same, and this factor would be the same for *every* language which has the same relevant phonemic pattern (e.g., phonemic palatal stops). That is, I am assuming here that the degree of coarticulation, say between a velar stop and the following front vowel, is the same for any language provided that it has the same degree of the freedom factor (and the same target values), although it would be different from those languages which have different degrees of freedom factor, e.g., no phonemic palatal stops. There is no *a priori* reason to rule out such context-sensitive universal rules as non-universals.

is a tricky and difficult question, but if we may simplify a generalization by saying that perceptually irrelevant phenomenon is also irrelevant to linguistic description, and that intrinsic allophones are perceptually irrelevant, then we are justified in excluding the intrinsic allophonic phenomenon from the specification by rules of systematic synthesis.

We now come to the conclusion of this chapter: The scope of phonetic specification extends beyond the level of systematic phonetics into the level of systematic synthesis, rules of which specify the utterances up to extrinsic allophones which are defined as those allophones that are non-predictable from the intrinsic physiological influence of the neighboring sounds but are generatable only via rules whose environmental specifications involve units at a higher level.

## III

## THE ROLE OF A SPEECH SYNTHESIZER

It was mentioned in the preceding chapter that the process of actualization of speech is equivalent to speech synthesis. This permits us to picture the entire structure of systematic synthesis as a complex of a speech synthesis device in which rules of the systematic synthesis serve as instructions to the synthesizer the input to which is the output of systematic phonetics, and whose output is sound:

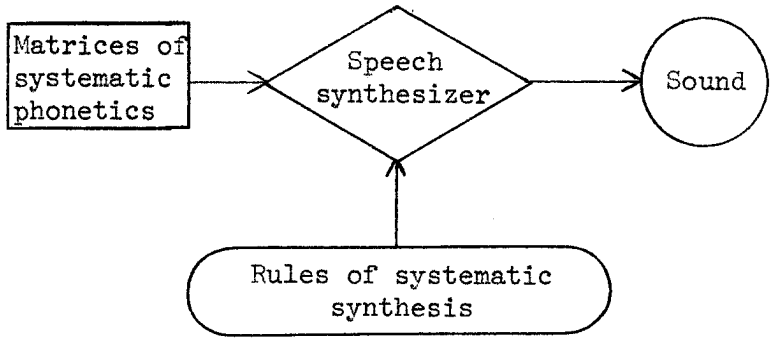



Figure 1. Diagram showing input and output of speech synthesizer

The function of a speech synthesizer is then, as diagrammed above, to carry out instructions (rules of systematic synthesis) on matrices of systematic phonetics and thereby producing utterances. This is the sound of the language. From a phonetic point of view, a phonological grammar is descriptively adequate if it generates actual audible (not written) utterances of natively acceptable character. Just as the syntactic component of English is inadequate if it generates a string of the form **many boy*, so is the phonological component of English if it generates, e.g., unaspirated stops in stress-initial position. In other words, if the output of the synthesizer is acceptable and agreeable to the native speaker-hearer, then we have the descriptively adequate phonological component. But if the output is not acceptable as "native" utterances, then we know that the computer program that provides performance instructions to the synthesizer or some other higher phonetic rules are not adequate, and that, therefore, revisions have to be made (assuming that the mechanism of the synthesizer is adequate enough for the purpose). In this respect, an important role of a speech synthesizer is to provide test utterances. In linguistic phonetics, this role is not trivial. If we are going to validate our phonological description, the assessment must be made of it in terms of "native" perception, and it is beyond question that this perception is aural in nature, and that aural perception is possible only with audible actual utterances.

(For detailed discussion on this, see p. 22ff) As Ladefoged (forthcoming) put it,

> We cannot test descriptions of a code [= a language] without reference to its manifestations; the only data we have for checking our descriptions of a language are the utterances of individual speakers.

In this sense, a speech synthesizer manipulable in terms of rules of systematic synthesis on the output of abstract, higher level phonological rules plays a fundamental role in the phonetic description.

Recently, however, there have arisen some doubts about the significance of a speech synthesizer in consideration of the fact that there exists no isomorphic relation between a physical stimulus and its aural perception. For example, Kelley, in a personal communication, expressed the view that creation of an acoustic speech synthesizer that synthesizes speech directly from fully specified matrices does not serve as a model of linguistic competence, and that, consequently, the non-existence of rules mapping fully specified distinctive feature matrices into input variables for an acoustic synthesizer is not significant.[6] This implies that the parametric values of a synthesizer need not necessarily represent or match perceptually relevant features and only those. Since this is an important issue, we will discuss it at length.

The issue hinges upon the fact that for a sound originated at the speaker's brain to reach the hearer, it has to travel several stages of different physical layers and that at each stage a non-linear, non-isomorphic transformation occurs so that there is no one-to-one correspondence between any two stages. Ignoring the initial stage of motor commands and the final stage of perception, three stages of the

---

[6]While this view does not imply that a message produced by a speaker via rules that generate motor commands does not have an acoustic representation nor that perceptually relevant features may not be present in the acoustic signal, it implies that the rules by which the speaker maps phonetic matrices into motor instructions that activate articulators and the rules that map phonetic representations into commands for an acoustic synthesizer are not comparable. The implication is correct, in so far as:

(1) the synthesizer is limited to an acoustic one, and

(2) phonetic matrices contain mentalistic elements of some sort which are not mechanically interpretable.

(1) is no longer tenable when the synthesizer is a dynamic analog of the vocal tract, and (2) is not permissible within the current format of generative grammar where there is no direct connection between the semantic component and the phonological component. Any relevant semanticism would have been taken care of by deep structures and semantic rules so that there will be no mentalistic element left at the final level of phonetic representation yet to be interpreted but not mechanically. Now, the operation of a speech synthesizer and the final stage of the speaker's production of sounds parallel each other.

phonic path are articulatory → acoustic → aural in that order. As is well known, the articulated sound is transformed into acoustic waves in the speaker's mouth, which are then again transformed at the listener's ear into some form of neurophysiological impulses whose arrival at the brain are directly responsible for perception, and that these transformations are non-linearly processed. The non-linearity of the ear is illustrated by the fact that equal increments in the intensity of a tone do not always correspond to equal increments in its loudness. A transformation in the signal occurs when the wave motion within the cochlea is converted to the form of nerve impulses. (It is for this reason that the usual spacing of the formant scales is according to the subjective pitch or mel scale. A mel is defined as the psycho-physical unit of pitch.) To cite an example, Lehiste and Peterson (1959) report that when listeners were asked to judge the relative loudness of vowels, they almost invariably identified the vowels that were produced with a greater subjective effort but with less inherent amplitude (such as /i/, /u/ recorded at zero VU) as louder than the vowels having greater intrinsic amplitude but produced with normal effort (such as /a/, etc.). For this reason, any observable phenomenon at a subsequent stage may but need not reflect a true and full image of the preceding, like a refracted light may not give a faithful reflection of reality.

The issue, then, is that since distinctive features are claimed to be perceptual features rather than acoustical, it is not really significant if a speech synthesizer which takes different acoustic values for different distinctive features as input variables produced sounds whose different perceptual quality did not match the acoustic differences.

I agree that acoustic definitions of perceptual features do not have to meet the "linearity" and "invariance" conditions (Chomsky 1964). But it is very important to note that the non-linearity (i.e., no one-to-one correspondence) in our case is invariably manifested in many-to-one, not one-to-many, relations. That is, the non-isomorphism exists only due to the fact that some variables or information-carriers in the antecedent stage are irrelevant or redundant for the establishment of the corresponding variables in the subsequence stage. That is, the following formula holds for the relation between the three stages of articulation(X), acoustics (Y), and perception(Z) in terms of the total amount of information needed to describe one and the same perceptual feature (not in terms of simplicity, or of the number of artificially categorized variables):

$$X \geq Y \geq Z$$

If this assumption is correct, then, it follows that the description of any stage in terms of the preceding stage must always be (more than) sufficient. That is, whatever indiscrete acoustic substance is present at the stage Y to be transformed into some discrete perceptual features at the stage Z, it must be the case that the information available at Y is at least equal to or greater than the amount needed to categorize the phenomenon at Z. That is;

> Each of the consecutive stages, from articulation to perception, may be predicted from the preceding stage. Since with each subsequent stage the selectivity increases, this predictability is irreversible and some variables of any antecedent stage are irrelevant for the subsequent stage. (*Preliminaries*, p. 12)

Thus, for any perceptual phenomenon, there must be, in principle, at least one or more than one corresponding acoustic phenomenon that is responsible for it; any perceptual distinction presupposes acoustic differences. The reverse case, i.e., the case in which several distinct perceptions are made from one and the same acoustic phenomenon is, in principle, impossible. As a system of communication, language includes redundancy so that it may be understood even with some loss of the original message through the channel noise. If the system of language operates in the reverse way, i.e., if the number of possible interpretations of the original message increases at every stage of its path from the speaker to the hearer, every act of communication would be like playing a chess game; trying to figure out the intended trick among many possible tactics at every move of the opponent. Language is not imagination or phantasy.

Jakobson and Halle (1956) themselves stress the importance of the definability of phonological units, and oppose the fictionalist's view of the phoneme as follows:

> When operating with a phoneme or distinctive feature we are primarily concerned with a constant which is present in the various particulars. . . . Phonemic analysis is a study of properties, invariant under certain transformations. . . . If the analyzer opposes the phoneme and its components to sound as a mere contrivance having no necessary correlate in concrete experience, the results of the analysis will be distorted through this assumption. The belief that the choice among phonemes to which we assign the sound might, upon occasion, be made arbitrary, even at random, threatens the objective value of phonemic analysis. This danger may, however, be avoided by the methodological demand that any distinctive feature, and consequently, any phoneme treated by the linguist, have its constant correlate at each stage of the speech event and thus be identifiable at any level accessible to observation. Our present knowledge of the physical and physiological aspects of speech sounds is sufficient to meet this demand. The sameness of a distinctive feature throughout all its variable implementations is now objectively demonstrable. (pp. 13-14)
>
> Each venture to reduce language to its ultimate invariants . . . with no reference to their empiric correlates is condemned to failure. (p. 15)

One might, however, challenge this dictum on the ground that sometimes we perceive something from nothing. As an example of a case in which a perception is nonetheless made from a nonexistent physical stimulus, Chomsky remarked (at the 5th Texas Conference on Phonology, January 1966) that an English listener can distinguish different degrees of stress even when their physical correlates (differences in amplitude, duration, or whatever they may be) are not actually present in acoustic waves. A more revealing experiment is reported in *Preliminaries*:

> Interference by the language pattern affects even our responses to non-speech sounds. Knocks produced at even intervals, with every third louder, are perceived as groups of three separated by a pause. The pause is usually claimed by a Czech to fall before the louder knock, by a Frenchman to fall after the louder; while a Pole hears the pause one knock after the louder. The different perceptions correspond exactly to the position of the word stress in the languages involved: in Czech the stress is on the initial syllable, in French, on the final and in Polish, on the penult. When the knocks are produced with equal loudness but with a longer interval after every third, the Czech attributes greater loudness to the first knock, the Pole, to the second, and the Frenchman, to the third. (pp. 10-11)

Thus, a perception was made, not through an extraction of acoustic stimulus which normally contributes to the listener as stress, but through a projection of the listener's internal knowledge about his language on the acoustic substance. That is, the listener projected his own grammar onto the input and made the judgement as his projection commanded. He simply heard the stress at the position where he expected it. This is how communication is often established between native and foreign speakers, and this is the perception by the so-called "analysis-by-synthesis" procedure (Halle and Stevens 1964).

Notice, however, how this projection or analysis-by-synthesis is possible, or further yet, how a speaker-hearer has internalized his grammar that enables the listener to make such a projection in a patterned and nonarbitrary way. This question is tied with the procedure of the child's language acquisition with which we are not overtly concerned here. But we can say in short that the child must have internalized his grammar (in a broad sense, including phonology) at least in accordance with a way which does not conflict with the pattern he found in the speech of people surrounding him. That is, in the speech that reaches the child's ear, the relevant data must have been present in a systematic way so that the child could extract its pattern and internalize it as a part of his competence. It is highly improbable that the child can perceive, say, a stress and internalize it as a perceptually discrete feature, when in none of the speeches of his playmates was

present acoustically distinctive stress.[7] From this point of view, the mechanism of speech synthesis is analogous to the child's learning process of speech production. That is, it is reasonable to assume that the child is constantly correcting and adjusting his "instructions" (= rules of systematic synthesis) whenever his speech output is corrected and rejected by his parents or playmates, until he perfects the instructions. At this stage, the set of rules of systematic synthesis as well as higher phonological rules is presumably fixed, i.e., the internalization of phonological grammar has occurred, so that a restructurization will be extremely rare. In this sense, if a listener judged acoustic nothing to be perceptual something, it is a matter of "naivete" of the judgement, not a proof that contradiction between the hypothesized perceptual categories and the corresponding observable phenomena in the acoustic or articulatory reality does not invalidate the hypothesis.

On the contrary, for a hypothesis to be a valid theory, it must be validated by the observable phenomena, or it must be able to explain the pattern of the behavior of the world that it hypothesizes. Einstein's General Theory of Relativity remained as a pure hypothesis until it was validated by the measurement of the deflection of the starlight in the gravitational field of the Sun on the day (May 29) of the solar eclipse in 1919; and Newton's Theory of Gravity would be an invalid hypothesis if it did not explain the planetary behavior in the universe. So will the perceptual features remain as a conjecture, unless they are acoustically validated. Thus, the authors of *Preliminaries* preface:

> We regard the present list of distinctive features, and particularly their definitions on different levels, as a provisional sketch which . . . requires experimental verification and further elaboration. (p. v)

To take an example, if the acoustic features, whatever they may be, of bilabiality and velarity do not have a greater perceptual similarity or psychological reality between them than those of alveolarity and velarity do, then either the claim that the former two articulatory processes yield the same perceptual feature [+gravity], while the latter two do not (i.e., [-grave] vs. [+grave]) is invalid, or the chosen acoustic cues as the common denominators of the one perceptual feature [+grave] are incorrect. Only an acoustic synthesizer can solve this kind of problem. We will see below in more detail in what sense this is true.

---

[7] I am not adhering here to the so-called empiricist's view that language is learned only by conditioning and external stimulation. I agree, with rationalists, that to a large extent the schema for grammar is given which will develop spontaneously in the mind under certain conditions (cf. Chomsky 1965, p. 49ff.). What is stressed here is these certain conditions that have to be initially presented to the child to set the language-forming process into operation.

Take the case of the "linguistic relevancy." The use of a speech synthesizer is the best and perhaps the only feasible way to discover the essential cue(s) in the phonemic distinction sorting out other redundant data. As Fant (1956) put it:

> It is evident that before we adjust our methods of specification in order to obtain statements that are optimal with regard to hearing and to the reception of a speech message, it pays to eliminate those redundancies that are due to an interdependence of the parameters of specification. (p. 109)

This is a problem that is frequently met in phonetic analysis, i.e., the question of which of several physical differences revealed between a pair of one phonemic opposition is the perceptually most relevant one. A decade of discussion on which of voicing, aspiration, or tenseness is the main perceptual cue in English stops is just an example. For a question of this kind, only one method of solution is feasible, namely an experiment using synthetic speech. It is so, because with a synthesizer which produces speech by combining artificially the different variables, it is possible to vary one (or any desired number of) feature(s) at a time, leaving the rest intact. Auditory judgements on the result of this kind of variation will tell us which factor is the most relevant to perception. This measure is infeasible and inadequate with the human vocal apparatus because there the variables cannot be controlled independently.

Needless to say, for a linguistic description to be economic and compact, and for the design in communication engineering, e.g., telephony, to be simpler, the question of "relevancy" is an important one. Contrary to the popular view that experimental phonetics is capable of solving this kind of question, all that instrumental analysis can do is to discover the physical facts corresponding to the linguistic entities or units. In fact, the more detailed the instrumental analysis becomes, the more numerous and complicated the physical data obtained become. Only synthesis methods can give a definite answer to questions of this type. By varying one feature at a time, it is possible to do what the human speaker cannot, i.e., to isolate one phonetic feature from another and examine its role in speech perception or in the communication processes. As Cooper (1962) put it:

> The use of an acoustic speech synthesizer enables us to decide what aspects of the acoustic pattern are significant carriers of information . . . and to convert the spectrum back into sound for phonetic evaluation by ear. Thus, the experimenter can test his hypotheses about significance by manipulating one spectrum and hearing the result. (p. 4)

By "significance" it is meant "linguistic relevance." An excellent example of this is provided by Malmberg (1963, p. 102f.):

Swedish has a phonemic word accent which is the sole distinguishing factor in such minimal pairs as *ânden* 'the duck' -- *ànden* 'the ghost,' *tânken* 'the tank' -- *tànken* 'the thought,' etc., which sound

> just different for a native speaker without any linguistic training. He normally hears the difference but is in most cases unable to give any meaningful description of what he hears or believes he does when he pronounces the words. That the phonetic description of this accent distinction has been no easier for native than for foreign linguists and phoneticians is easily illustrated by a survey of literature on the subject. (pp. 102-3)

In a series of experiments with speech synthesizers, Malmberg has found that the difference in neither intonation, nor duration, nor intensity, but in the pitch pattern was the absolute condition for the distinction of the two accents; accent 1 (´) being mainly Fall (◜) with the peak of pitch (150 cps) at the beginning of the vowel (within 25 msec), and accent 2 (`) being Rise-Fall (◝) with the pitch peak in the middle of the vowel (near 100 msec point from the beginning of the vowel). Malmberg therefore concludes that differences in other phenomena, though they "normally but not regularly accompany the pitch pattern, are 'redundant' in the proper sense of the term." (p. 110)

At this point, it is perhaps worth examing the notion "redundancy" in linguistic description, since it has an important bearing on speech specification depending on how it is viewed and defined.

Although there are indications that syntactic and semantic components also involve redundancies (E.g., [+Human] → [+Animate] → [+Concrete] → [+Countable], etc. Cf. Chomsky 1965; Katz and Postal 1964.), it is in phonology that redundancy plays the most important role in description, i.e., phonological redundancy rules. Redundancy rules have originally been motivated in order to define the ways in which language as a system of communication carries information which is unnecessary, and hence "redundant," for establishing intelligibility. For example, if, as a realization of a phoneme or a series of phonemes of a language (e.g., English stops), feature A (e.g., Voicelessness) is always found with feature B (e.g., Tenseness) and feature C (e.g., Aspiration), but the reverse is not true; that is, if features B and C are merely concomitant phenomena of feature A and their absence or presence does not affect the perception of the phoneme(s), but the reverse is not true, i.e., the absence of A bars the perception, then features B and C are called redundant, and this redundancy is reflected in a rule of the form:

$$(2) \qquad A \rightarrow \begin{bmatrix} B \\ C \end{bmatrix} \text{ or}$$

$$[\text{Voiceless Stop}] \rightarrow \left[\begin{matrix} \text{Tense} \\ \text{Aspirated} \end{matrix}\right] \qquad (\text{read} \rightarrow \text{"implies"})$$

This is done, of course, in order to simplify the description. (As a practical example, consider a dictionary in which every entry is specified with its phonetic shape. We can easily see a saving if the relevant

sound is specified A, instead of ABC every time.) In the present day linguistic literature, however, the term "redundancy (R) rule" is used in a wider sense.

R rules specify, on the one hand, the *inability* of some features to occur with each other due to the inherent physiological constraints. For example, when we say that all vowels are redundantly [+Continuant], its true implication is that there is, by definition, no non-continuant or interrupted vowel. In DF terms, the combination $*\begin{bmatrix}+\text{Vocalic}\\-\text{Continuant}\end{bmatrix}$ simply cannot occur. It seems that this kind of redundancy is universal due to the universal physiological structure of the human vocal tract.

R rules specify, on the other hand, *predictability* of some feature(s) given another feature or features. For example, [+Nasal C] → [+Diffuse] means that in this particular language all nasal consonants are articulated in the front part of the oral cavity. Let us call the former kind *restrictive R rules*, and the latter, *non-restrictive R rules*.

In a restrictive R rule, the redundant feature has nothing to do with the preceptual relevancy. The feature in question simply cannot be present or absent, whichever the case may be. In a non-restrictive R rule, the redundant feature is usually present in the sound in question, but is said to be "irrelevant" for perception. A logical corollary of this assumption is that a speech synthesizer need not take the redundant features into consideration, e.g., need not give values to them. It seems, however, that this is a gross misconception of the notion "redundancy" in linguistic context. I maintain that "redundant" does not mean "superfluous" which can be removed or left out of a speech synthesizer without consequence. On the contrary, redundant phenomena are very relevant for the establishment of communication. Very often a redundant feature is the sole criterion for the distinctive perception. For example, in the case of *rider/writer* where both /t/ and /d/ are a voiced flap [ɾ], the sole cue contributing to the distinctive perception is said to be the length difference in the vowel of the first syllable. But this durational feature is a redundant one as it is predictable from the voicing of the following consonant. Neglect or removal of redundant features may not only render an utterance poorly intelligible, as was in the example above,[8]

---

[8]For further similar examples, I cite the literature:

> The phoneme /i/ [in Russian] is implemented as a back vowel [ɨ] after non-palatalized consonants, and as a front vowel [i] in all other positions. These variants are redundant, and normally for Russian listeners it is the difference between the non-palatalized [s] and the palatalized [sʲ] which serves as the means of discriminating between the syllables [sɨ] and [sʲi]. But when a mason telephoned as engineer saying that the walls [sɨʂ'ejut] 'are getting damp' and the transmission distorted the high frequencies of the [s] so that it was difficult to comprehend whether the walls 'were getting damp' or 'turning

but also make the sound non-native, and even utterly silent. Let us consider an example or two.

Suppose that a language has only one phonemic nasal consonant /n/. Thus, in this language, [Nasal C] → [Alveolar]. This rule implies that alveolarity is predictable from the nasality of the consonant and that alveolarity conveys no other information than the one already present in the nasality. But the neglect of this redundancy may lead to the substitution of [m] or [ŋ] for /n/ which is normally realized as [n]. There is no doubt that this makes the utterance sound non-native. Similarly, aspiration is said to be a redundant feature of voiceless stops of English in stress-initial position. But its absence will make an English utterance outlandish. An analogous example may be cited from syntax. In *three boys*, *-s* would be a redundant morpheme of plurality, since a plural number explicitly specifies the plurality of the noun. But a neglect to express the redundant morpheme, e.g., **three boy*, would be non-English. We have argued earlier that an adequate phonetic theory must be capable of characterizing the sound of a language as being native, not foreign. If we hold the view, however, that redundant features may be dispensed from the synthesizer, our phonetic theory would become inherently incapable of fulfilling one of its important roles.

Take an extreme case. It is reported that standard amplitude values of formants are predictable from the frequency of formants. (cf. Fant (1956) and Chapter VIII for detail) If this assumption is correct, the amplitude would be a redundant feature of the formant in the proper sense. But if we neglect to assign amplitude values to formants in an acoustic speech synthesizer, all resonant phonemes might be utterly silent!

We, then, define the redundancy in phonology as those phenomena that are redundant or irrelevant to the establishment of the phonemic system of oppositions of a language at the systematic phonemic level, and only those. (For further discussion on phonological redundancy, see Chapter IV) At the lower level, redundancy rules must be interpreted in toto by a synthesizer.[9]

---

gray' [ʂiɻ'ejut], then the worker repeated the word with particular emphasis on the [+], and through this redundant feature the listener made the right choice. (*Preliminaries*, p. 8)

Although an English hearer will usually identify the consonants [/s/ and /z/ in final position] correctly, in spite of their resemblance to one another, the right identification is often facilitated by the concomitant difference in the length of the preceding phoneme: *pence* [peňs] - *pens* [pen:s]. (Jones 1950, p. 53, as was cited in *Fundamentals*, p. 9)

[9]An observant reader may ask why I excluded rules of intrinsic allophones from systematic synthesis on the grounds that they are reundancy rules, while maintaining here that all R rules must go through the

Concluding, Ladefoged's following remark is quite appropriate:

> If it [= a theory of phonetics] is to be interesting, the description of a language must also be testable; and the possibility of making a sufficient test must be inherent in the underlying theory. (forthcoming)

The establishment of the level of systematic synthesis as an extended part of the phonological component meets this condition, and as a converter of rules into testable sounds, a speech synthesizer plays a fundamental and indispensable role in the validation of linguistic description.

Deriving such testable utterances, however, is not quite so easy and simple as one might wish. At the level of syntax, it is relatively easy to generate a string of formatives, the grammaticalness of which is to be tested. In phonology, however, one could use only the human vocal apparatus until recently, and this has been found to be inadequate for the purpose because of its poor flexibility and controllability. Then, with the advent of magnetic tape recordings, there once was hope that one might be able to synthesize speech by cutting and resplicing prerecorded phonemic segments. That there is no isomorphic relation between the phonemic signals at the input and an inventory of prerecorded sound, and therefore this measure was also found not to be feasible is too well known. One only needs to look at spectrograms to see that speech varies continuously over stretches of greater than phonemic length.

It is my conviction, however, that newly developed techniques for synthesizing speech now make it possible to write a phonological description from which testable utterances can be made by precisely defined operations. On this basis, the phonology becomes, in effect, a set of rules for synthesis, with explicit procedures for going from a sequence of phonemes (or feature-complexes) to their realization as speech sounds.

Whether a simpler phonology will be achieved by stating rules for synthesis in articulatory terms or conversely in acoustic terms is still an open question. We take this up in the next chapter.

---

synthesizer. This is not a paradox. Note that I did not exlude intrinsic allophonic rules from going through a speech synthesizer, but only from a membership of systematic synthesis, as intrinsic allophonic features do not seem to serve to distinguish neither a language from another nor an idiolect from another. As lower level R rules, they go through synthesizer, as is maintained here, to the extent that they are minimally necessary for naturalness. Cf. Chapter IX.

# IV

# UNIVERSAL PHONETIC CATEGORIES

As stated in the preceding chapter the input to the speech synthesizer is matrices of systematic phonetics. In the present form of generative phonology, the rows of these matrices are Jakobsonian "distinctive features" (DF). In this chapter, we will critically examine (1) the binarity of DF's and (2) the properties of DF's. We shall justify the proposition that a model of universal phonetics whose features are non-binary articulatory categories[10] is built on more rational foundations and explains certain phonological facts in a more natural and intuitively correct way than the DF model.

The theory of DF has been proposed as a universal framework of phonological characterization of speech for a linguistic description. This chapter will be confined to the two aforementioned topics, topics that are the most controversial and have the gravest consequences in terms of the claims that the theory makes. The fact that a modification is proposed in the binary nature and the defining properties of DF's implies that we accept other important claims that the DF theory makes about phonology. In particular, we agree with the DF proponents

---

[10] A phonetic theory which is also based on physiological parameters has very recently been presented by Peterson and Shoup (1966), in which certain components of the speech mechanism are defined in a set of preliminary definitions, and assumptions about the actions of the vocal mechanism are given in a set of axioms. Unfortunately, the timing of the appearance of their paper was such that it was not available for writing this monograph, and it is regrettably not possible to discuss their paper in detail and compare their theory with the one which is presented in the second half of this Chapter IV. Yet, it is worth quoting from the opening page, the following statement by the authors about the requirements of a phonetic theory:

> A "phonetic" system must provide a means of describing the significant sounds found in the various languages. The system must provide sufficient detail so that the natural pronunciation of any particular language can be described rationally. It must also be sufficiently detailed so that pronunciations of different languages can be compared and related. An effective phonetic system must have universal application to spoken languages. If the system must be revised and reinterpreted for each different language, then it is not a phonetic system at all and it cannot serve the purposes of phonetic description. (p. 6)

that (1) some sort of subphonemic componential feature notations achieve a greater simplicity in description,[11] (2) that

> if we state rules strictly in terms of features, then we can propose an effective evaluation procedure which distinguishes true generalizations in terms of natural classes . . . from linguistically nonsignificant pseudo-generalizations (Chomsky and Halle 1965, p. 119),

and (3) that the featural notations enable the phonological description to meet the level of descriptive adequacy in that the theory makes the distinction between admissible and inadmissible phonological forms. We discuss in brief why these should be true before we move on to our critical review.

Economy of featural representation comes from the fact that P rules in general apply not to an isolated item or a group of unrelated disjunctive phonemes but to all members of the same natural phonological class. Thus if we crudely say that each DF represents a natural phonological class (a discussion and a more elaborate definition of "natural class" will be given later), then a rule in terms of DF would be simpler than a rule involving an enumeration of members of the class.

The failure of taxonomic phonemics in the three categories mentioned above is, as Halle claims, due to the fact that the notion of "natural class" of phones has no significance when phonemes are viewed as indivisible units:

> We observe that the intuitively correct result is yielded by the proposed simplicity criterion in conjunction with a representation of phonemes as bundles of distinctive features, whereas the above counterintuitive result is obtained if phonemes are regarded as indivisible entities. The failure of the simplicity criterion in the latter case is due to the fact that the notion of natural class has no obvious meaning if phonemes are regarded as indivisible entities. (Halle 1964b, p. 337)

We conclude, therefore, that the conception of phonemes as indivisible units, whether or not the framework of universal phonetics based on this conception includes non-terminal phonological class-symbols, is inadequate for descriptively adequate phonological descriptions. This is not to accept the present form of the DF theory unconditionally, but only to imply that some sort of subphonemic componential notations are more consistent with the achievement of the level of explanatory adequacy

---

[11] One might argue against this on the grounds that the concept of simplicity is undefinable when one compares two different theories (cf. Chomsky and Halle 1965, p. 113). I believe however, that if generalizations are made about the same empirical data, there should be some meaningful relative measure of the degree of generalizations which is not just internal to a particular theory. See also fn. 15.

of phonology, in so far as a simplicity criterion and an evaluation measure are a fundamental part of the theory; and that, since the DF theory employs the concept of phonemes as divisible components, it is to that extent more adequate than the taxonomic theory of phonology. But the DF theory is not the only feasible system employing the concept of a phoneme as complexes of divisible components, and furthermore, as was mentioned in the beginning of this chapter, the DF theory needs a critical examination in some of its metatheoretical claims, in particular, regarding the questions of binary opposition and of featural properties. We will see how weakly the claims of DF proponents with regard to these two issues are supported, and in what ways a modified model of universal phonetics overcoming the weaknesses of the DF theory may be worked out. We will also discuss such crucial notions as natural class, featural hierarchy, phonological redundancy, etc.

With regard to the question of the binary character of DF's, a question we must ask is whether the binary scale is a mechanical measure that the analyzer profitably imposes on the linguistic code or whether this scale is inherent in the structure of sound (cf. Chao 1954). Interesting to note with regard to this question is that in *Preliminaries* (1951) one paragraph asserts that "the dichotomous scale is the pivotal principle of the linguistic structure. The code imposes it upon sound," whereas in *Fundamentals* (1956) there is a less dogmatic statement which now asserts that "there are several weighty arguments in favor of the latter solution" (p. 47). "Several weighty arguments" that Jakobson and Halle present are:

(i) A system of DF's based on binary opposition is the optimal code in encoding and decoding operations.

(ii) The binary opposition is a child's first logical operation.

(iii) Most of the DF's show dichotomous structure on the acoustic and motor level. (*Fundamentals*, pp. 47-49)

We will examine (i) and (iii) together first, and then (ii).

(i) seems to assert that a binary opposition is a mechanical scale that the analyzer imposes upon the code rather than its inherent structure. It is, however, unwarranted to assume that human brains are incapable of discriminating and perceiving the sound in a more complicated and less economic set of differential criteria than two, and to assume that, as Householder (1966) comments, the human brain necessarily functions in the most efficient, logical, and economic way, like a digital computer, with no room for "extravagant redundancy in our brain-storage." It might be a principle of science to assume that nature behaves in the most economic and efficient way, unless there are other factors that make this impossible. In the case of speech, there indeed is this factor, i.e., the fact that the primary physiological function of the organs of the mouth is mastication, not speaking. Thus, it would not be surprising if human speech does not use the optimal code. As for the argument (iii), it is true that "most" phonetic classes or DF's are inherently binary, e.g., voiced vs. voiceless, oral vs. nasal, etc. It is not known, however, how "most" leads to a logical conclusion "all." Classes such as tone, stress, place of articulation, vowel height, etc. may have more than two oppositional members, and it is difficult and just

as arbitrary to decide where the first binary division is and where the next one is. This is so because, as Ladefoged (1966; forthcoming) points out, these members are different manifestations of a variable which is physically a single continuum, and because different features of the same variable are distinguished from each other by the degree that they possess the property of the variable. To take concrete examples, we will examine the feature Diffuse/Compact and Acute/Grave, since they seem to be, among the twelve or so DF's that have so far been introduced, the most non-binary and the most controversial features. If we succeed in showing that these features are non-binary, then we will have proven that binarity is an arbitrary imposition of the DF theory on the phonetic structure.

But first, there is a minor but non-trivial matter to consider. Since the features Diffuse/Compact and Acute/Grave have different motor and acoustic manifestations which are incomparable on the same scale according to whether the feature is that of a vocalic or consonantal segment, we must consider first in what sense it is valid to assign the same feature to both vocalic and consonantal segments even when the manifestations of the feature in each case are incomparably different. It seems that this set-up was motivated by the desire for economy in the number of features, and made possible by the claims (1) that features are autonomous and independent of each other (*Preliminaries*, p. 41) so that any feature must logically occur with any other feature, (2) that, according to the principle of complementary distribution, two different manifestations can legitimately be combined into one feature:

> While the relational structure of these features, which are common to consonants and vowels, manifests a definite isomorphism, the variations are in complementary distritution. That is to say, they are determined by the different contexts in which they appear: the variations are dependent upon whether the gravity-acuteness and compactness-diffuseness features are superposed upon a vowel or a consonant. (*Preliminaries*, p. 7)

and (3) that there is a perceptual unity in the feature irrespective of whether it occurs in a vowel or a consonant, i.e., there is the same psychological "association" common to both Diffuse vowels and Diffuse consonants, etc.:

> On the perceptual level a distinct association links the consonantal and vocalic opposition of compactness and diffuseness. . . . The contact with [a], the most compact, and with [i] and [u], the most diffuse of the vowels, prompts the association of this stop with [k], the most compact, and with [p], the most diffuse of the stops, respectively. Similarly the scale of magnitude, i.e., the small-vs.-large symbolism, latently connected for the average listeners with the opposition of compact and diffuse, works alike for vowels and for consonants. (*Preliminaries*, p. 28)

I would like to argue that claim (1) is incorrect, and claim (2) invites an element of self-contradiction, and claim (3) has little experimental evidence. Claim (1) is incorrect in that some features are not autonomous at all. Universal restrictive redundancy rules are precisely statements about the inherent restrictions on featural combinations. For example, given a vowel ([+Voc, -Cons]), the following features are predetermined, i.e., there is no question of choice or option: [+Continuant, +Voice, -Strident, -Checked, -Sharp]. These amount to seven out of the original twelve distinctive features, leaving only five (Gravity, Compactness, Flatness, Tenseness, and Nasality) as capable of functioning independently. Further yet, $\begin{bmatrix} +\text{Voc} \\ +\text{Cons} \end{bmatrix}$ predetermines the values of most of the rest of DF's. In view of this, it is difficult to see in what sense DF's remain autonomous "despite their multiform interdependence within the phoneme and within the entire phonemic pattern" (*Preliminaries*, p. 41). Note that I am not arguing here that vowels and consonants should be specified with two different sets of features, but only that the hypothesis should be in accordance with the physical facts, which is not the case with the feature Diffuse/Compact as far as its physical manifestations are concerned. Notice that the features Voice, Nasal, Flat, etc., have the same physiological and acoustic correlates regardless of whether the segment is a vowel or a consonant. Thus the use of these features as defining categories of both vowels and consonants is justified.

Claim (2) is contradicted by an argument made elsewhere by Chomsky and Halle themselves, who vehemently denied and assaulted the significance of the principle of complementary distribution:

> The principle is apparently of no theoretical significance, and *should be dropped from linguistic theory altogether.* (Chomsky and Halle 1965, p. 129) [cf. also Chomsky 1964, p. 99, p. 103. Emphasis mine]

Although the above Chomsky-Halle criticism on complementary distribution may refer to just a particular form of this principle as used in a taxonomic discovery procedure, it is curious to see that a principle which was denounced with regard to one issue is called upon to justify another.

Claim (3) will be discussed in detail later when we talk about the property of DF's.

I see no justification or motivation to assign the same feature Diffuse/Compact to both consonants and vowels when its physical manifestations are so different from each other that whenever a P rule involving the feature Diffuse/Compact is stated, one must specify whether one is talking about Diffuse/Compact of a vowel or of a consonant. This unnecessarily complicates P rules. For example, a rule applying to a segment which is [-Continuant] (i.e., stop) does not have to specify that it is also $\begin{bmatrix} -\text{Voc} \\ +\text{Cons} \end{bmatrix}$, while a rule involving a segment which is

Diffuse or Grave must also justify the proposed measure. It is when one finds languages in which phonemes, regardless of whether they are vowels or consonants, having the feature Diffuse/Compact in common constitute a natural class, i.e., a class of /i, u, p, b, t, d/ (= [+Diffuse]) vs. a class of /æ, a, ɔ, k, g, č, ǰ/ (= [+Compact]), and its use simplifies P rules. In other words, if there are languages where a rule of the kind

(3) [+Diffuse] → [+Compact]/ ___X

applies to both Diffuse vowels and consonants, then we may say that grouping /i, u, p, b, t, d/ together under the one feature [+Diffuse] reveals a linguistically significant generalization. Otherwise, the measure has no theoretical significance.

It seems that the DF conception of speech sounds at the present moment is no richer than that of a set of classificatory features out of which sounds are in some way composed or combined. But since we know that *not* all possible combinations of the elements (features) are allowed as speech sounds, it follows that an exhaustive list of the features that make up the speech sounds would not in principle describe a human language, just as a list of grammatical categories would not describe a grammar. After all, it was the MIT linguists that first put out the "all and only" doctrine. But as it stands now, the DF theory may include "all,"[12] but not "all and only." There is an enormous redundancy, as Ladefoged (1965) and Householder (1965) points out. What is required in addition, therefore, is an explicit statement of the principles of formation and combination of features, i.e., phonological redundancy rules. Thus, the present form of the phonological component which employs the DF system includes R and MS rules. But as long as some of the R rules, e.g., restrictive R rules, are metatheoretical and universal, we might seek a framework in which the inherent restrictions are built into metatheory, not added on as a part of phonological rules of every language. We will see later in what ways this is possible and what advantageous consequences such a new framework has in the light of such notions as simplicity criterion, evaluation measure, phonetic similarity, etc.

We will now examine the validity of binarity in the case of the feature Diffuse/Compact in vowels. (The consonantal feature Diffuse/Compact will be discussed later in conjunction with the feature Acute/Grave.) In the present form of the DF theory, vowel heights are

---

[12]But Ladefoged (1965) doubts even this much capability of the DF system:

> Allowing for the stated combinatory restrictions, this apparatus [= DF system] generates 12,288 categories and even then probably does not account for all the 93 sounds [= consonants found in a number of West African languages described in Ladefoged 1964]. (p. 40)

specified as follows:

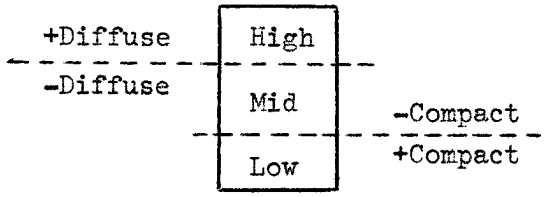



Figure 2. The relation between the DF's and a traditional classification of the vowel height

Let us for the moment assume that features are relative ("A DF is a relational property . . ." (*Fundamentals*, p. 14)) but autonomous, as is claimed. If these assumptions are correct, then it follows that the relative value of the feature Diffuse should remain invariant or independent of the relative value of the feature Compact. Thus, regardless of whether the relative values of High vs. Mid is [i] vs. [e], or [ɪ] vs. [ɛ], the relative values of Mid vs. Low, whether [e] vs. [æ], or [ɛ] vs. [a], should remain constant or intact, just as the value of Nasality should be the same for both Vocalic and Consonantal segments. This is a logical corollary of the autonomy of features. For example, it is claimed in Kim (1966) that tensity in Korean stops is autonomous, i.e., independent of aspiration, since the relative values of tensity remain constant while the relative values of aspiration fluctuate considerably, particularly in the case of lax stops, from voiced to aspirated. This was the essential reason for refuting Lisker and Abramson's (1964) one-dimensional categorization of stops in terms of the relative length of voicing. The two features Diffuse and Compact of vowels, however, are not analogous to Tensity and Aspiration in Korean stops, since values of the vowel height, whether Diffuse or Compact, swing on the same scale to the same degree. For example, Ladefoged and Broadbent (1957) showed that the identification of the vowel of a synthesized word, as /i/, /e/, or /æ/, was dependent on the vowel values of the introductory phrase "please say what this word is, /bit/"etc. This implies that when the formant values of /e/ and /æ/ shift downward to [ɛ] and [a] respectively, then /i/ also shifts to [ɪ] or [e], so that the former /e/ is now identified as /i/ when the shift is not known to the listener. This shows not only that the values are relative but also that vowel height is really a unidimensional scale, and that it is arbitrary to split three continuous items into two independent binary cuts.

It is to be noted, however, that three items on a continuum does not necessarily imply or presuppose a trinary division. If some combinations of two of the three items (with three items, a, b, c, there are three two-set combinations, ab, bc, ac) are found to behave together while the other combinations do not, this may be sufficient to suggest a binary cut. This, in fact, was the strong motivation, or, at least, the implication, to group High and Mid together as [-Compact], and Mid and Low together as [-Diffuse], but not to group High and Low under one feature. In particular, it was shown that in English, the first Vowel Shift Rule applies to non-Low, and the second VS rule to non-High, suggesting that High and Mid, and Mid and Low, but not High and Low, constitute natural classes respectively. But note Chomsky and Halle's (1965) answer in reply to Householder's

question: in what sense do /i, ʌ, æ, ɔ/ (the only phonemes occurring before /ŋ#/ in English, e.g., *sing, sang, sung, song*) constitute a natural class? They state that /ʌ/ is actually to be represented as /u/ in underlying systematic phonemic representation in the light of such alternation-pairs as *reduce - reduction, assume - assumption, numerical - number,* etc. and that, therefore, /i, u, æ, ɔ/ now constitute a natural class $\left\{\begin{matrix}[+\text{Diffuse}]\\[+\text{Compact}]\end{matrix}\right\}$ (Chomsky and Halle 1965, pp. 123-4). This, in fact, is to assert that just as High and Mid, and Mid and Low constitute natural classes, so do High and Low. In what sense then are High, Mid, and Low dichotomous? Notice furthermore that the natural classes of High and Mid, and Mid and Low are designated by a single feature [-Compact] and [-Diffuse] respectively, but the natural class of High and Low, by two features $\left\{\begin{matrix}[+\text{Diffuse}]\\[+\text{Compact}]\end{matrix}\right\}$. How does this differ from the others and what theoretical implications are there? Presumably Halle would say that a natural class defined with a single feature is more general (or more natural) than a natural class defined with two features, which is in turn more general than a natural class defined with three features, etc. But then the question is: where do we stop, and can any feature combine with any other feature(s) to constitute a natural class?

In the literature, natural class is defined as follows:

> A set of speech sounds forms a natural class if fewer features are required to designate the class than to designate any individual sound in the class. (Halle 1961, p. 91, and also Halle 1964a, p. 328)

This definition contains no measure to evaluate the degree of naturalness of natural classes. We cannot give our own definition of natural class here, because it involves more discussion which is yet to come, but it may be asserted that the definition of a natural class in terms of DF's is, at the present moment, not fully worked out, and that the binary division of High, Mid, and Low vowels (i.e., the bifurcation of the three units of the vowel height by reasoning that High and Mid constitute a natural class [-Diffuse], and Mid and Low constitute a natural class [-Compact]) is not well justified, either.

Not surprisingly, the DF proponents were also aware of the non-dichotomy of the feature Diffuse/Compact in vowels. For instance, we find such earlier statements as:

> The opposition compact vs. diffuse in the vowel pattern is the sole feature capable of presenting a middle term in addition to the two polar terms. On the perceptual level, experiments that obtained such middle terms through the mixture of a compact with the corresponding diffuse vowel seem to confirm the peculiar structure of this vocalic feature, which sets it apart from all other inherent features. (*Preliminaries*, p. 28)

> Among the inherent features, only the vocalic distinction compact/diffuse often presents a higher number of terms, mostly three. (*Fundamentals*, p. 48)

The above statements were made in 1951 and 1956 respectively. But a little later, it was apparently felt that, to achieve uniformity in the DF theory, every feature should be treated as binary, whether or not it is empirically so. Thus, we find in Halle (1957):

> Only in the case of the feature diffuse-nondiffuse has the *insistence* upon binary features led us to introduce a parameter which has an extremely restricted applicability and therefore may be said not to be optimal. It is for this reason that in previous formulations of the distinctive feature framework the feature compact-noncompact was defined as a ternary feature. In recent months we have been led to accept the more consistent solution of postulating two binary features in place of the ternary one, because in connection with our work on evaluation procedures for alternative phonemic solutions, we found that *the consistently binary system fitted our requirements better* than the mixed system previously used. (p. 71) [Emphasis mine]

This is to impose forcibly the analyst's view on the sound, not to describe its inherent structure. Grant that the binary code gives the simplest and the "most consistent" phonological description, but there is not the slightest reason to assume that facts may be dissected arbitrarily in order to make them "fit" the framework which was mechanically premolded. This point is stressed by none other than Chomsky and Halle themselves:

> Even if the absolute notion of 'simplicy' could somehow be justified, this would have little relevance to the problem of choosing among linguistic theories. Suppose it were true that a grammar X . . . is more 'complex,' in some sense, than a grammar Y . . . . This conclusion would still leave open the question whether the system used in natural languages is 'maximally simple' in this absolute sense. There is not the slightest reason to expect natural languages to be 'maximally simple,' assuming that some content can be given to this curious notion. *The relevant constraints are those of physical realizability, not 'absolute simplicity,'* whatever this may mean. (Chomsky and Halle 1965, p. 111, fn. 8) [Emphasis mine]

We will now turn briefly to the feature Acute/Grave. Positions of the tongue-hump relative to the pharynx are presently specified as follows:

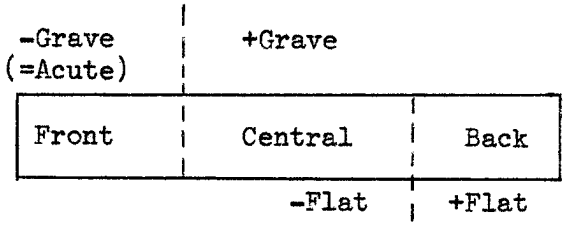



Figure 3. The relation between the DF's and a traditional classification of the vowel latitude

Unlike the case of vowel height where the originally single feature Diffuse/Compact was split into two independent features, the Front/Back dimension of vowels is specified with two primitively different features, Gravity and Flatness, as shown above. Implicit in this diagram is an assumption that, as far as classificatory matrices are concerned, there is no language in which Central is [+Flat] and/or Back is [-Flat]. This assumption seems hardly true. For example, there are strong grounds, in Korean, to set up both Unrounded Central and Rounded Central, as well as Rounded Back, as morphonemes (cf. Kim, forthcoming). The only way to specify these with different features would be, as Stockwell (1966) suggests, to make feature Flat independent of Gravity, and split Acute/Grave into two features, as in the case of Diffuse/Compact, so that

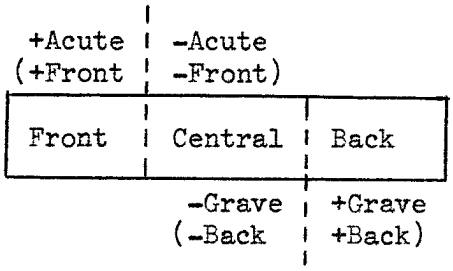



Figure 4. The suggested DF classification of the vowel latitude with Flatness independent of Gravity

Stockwell, further noting that front and back vowels form a class in many languages distinct from central vowels, proposes a feature Peripheral/nonPeripheral (i e æ u o ɔ/ ɨ ə a), and also a feature Opposite Rounding in an attempt to define the two-feature symmetry (Gravity and Rounding) found between front and back vowels in terms of one feature only. We will later discuss this latter phenomenon and its theoretical implications in more detail (cf. below, and Chapter VII), and return to the problem of the DF specification of vowel latitude. If the Korean case is correct and if we find other languages in which Central and/or Back agree with Flatness (For example, Westermann and Bryan (1952) cite several Benue-Congo languages having unrounded Back and/or Rounded Central vowels, e.g., Kum, Widekum, Mambila.), then Stockwell's suggested measure is inevitable. This makes the situation exactly analogous to the case of the feature Diffuse/Compact, and this is not surprising in view of the fact that Acute/Grave is merely the other scale, the abscissa, of the dynamic mechanism of the tongue whose ordinate is the scale of

Diffuse/Compact. Thus, the conclusion reached for Diffuse/Compact is also applicable here: the feature Acute/Grave defines unidimensional properties on a single continuum, and its scale is also non-binary.

We will now discuss the validity of the dichotomy of the place of consonantal articulation. Four main places of consonantal articulation are designated in DF terms as follows:

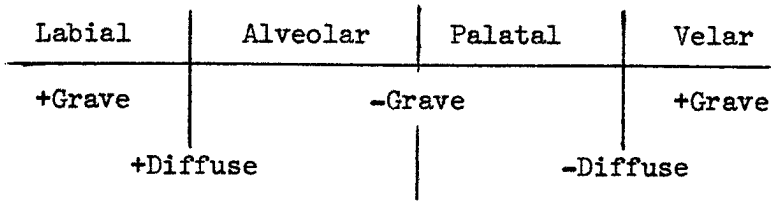



Figure 5. The relation between the DF's and a traditional classification of places of consonantal articulation

We will not engage ourselves here in the discussion of how many contrastive places of consonantal articulation a certain language has, etc. This aspect of the argument is given in detail in Ladefoged (1964; forthcoming). We will discuss here what seems to be an inconsistent and internally incoherent classificatory schema of the DF theory. To see this, we must compare the places of both consonantal and vocalic articulation together. From an articulatory point of view, it is true to observe that Front vowels fall under the Palatal region, and Back vowels under the Velar region, and that the difference between consonants and vowels in this respect lies, not in the places of articulation, but in the degree of oral constriction or the manner of production. A partial reproduction of the IPA chart in *Principles* (p. 10) clearly shows that founders of the IPA must have had this fact in mind, although it is nowhere explicitly stated in the *Principles*:

| Consonants | | Palatal | Velar | |
|---|---|---|---|---|
| Plosive | | c ɟ | k g | |
| . | . . . | | | . . . |
| . | | | | |
| . | | | | |
| Fricative | | ç ʝ | x ɣ | |
| Frictionless Continuants | | j (ɥ) | (w)ɣ | |

| Vowel | Front | Centr. | Back |
|---|---|---|---|
| Close | i | | u |
| Half-close | e | | o |
| . | | | |
| . | | | |
| . | | | |
| Open | | a | |

Figure 6. A partial reproduction of the IPA chart (*Principles*, p. 10) showing the common places of articulation for consonants and vowels

The network of this relationship is not entirely disregarded by the DF classification. Thus, all phonemes in the column Velar belong to [+Grave], and all the phonemes in the column Palatal belong to [-Grave]. But the classification of the degrees of oral openings, i.e., the vertical scale of the chart, with the DF Diffuse/Compact is incoherently made; namely, the first two rows are said to be [-Diffuse], the next two rows [+Diffuse], then the next row is simultaneously [-Diffuse] and [-Compact], and finally, the last row, [+Compact]. In vowels, it was the degree of the tongue-height or of the oral constriction that determined the degree of Diffuseness/Compactness. But when one extends the same notion to the consonantal case, one finds that, as far as Palatal-Velar region is concerned, the most constricted sounds are said to be [-Diffuse], instead of [+Diffuse] which is the value assigned to the most constricted vowels.

One might of course argue that DF's are neither physiologically nor acoustically definable categories but only perceptual categories, and that, therefore, inconsistencies of DF's with articulatory or acoustic data are irrelevant. This view was partly refuted in the preceding chapter in conjunction with the discussion on the use of an acoustic speech synthesizer, and more will be said later. But here, I would maintain that there are close correspondences between motor, acoustic, and perceptual categories, that there is no case where the correlation is unpredictably and incoherently unsystematic, and that a featural specification in an undefinable way, such as the unsystematic values of the feature Diffuse/Compact, is empirically unjustified. That there is a close correlation between articulatory categories and acoustico-perceptual DF's is amply attested by the exact correspondence between the two levels in some ten DF's. As was mentioned earlier, it is in the case of Diffuse/Compact and Acute/Grave that one finds the most discrepancy. I will attempt to show below that this discrepancy, in particular, two dichotomous divisions of the places of the consonantal articulation in terms of Diffuseness and Gravity, is not well motivated.

Consider the familiar case of Palatalization. In English, the phenomenon is shown in such diverse forms as:

/k/ → /s/, e.g., *electric - electricity, critic - criticism*
/k/ → /ʃ/, e.g., *magic - magician, music - musician*
/g/ → /dʒ/, e.g., *pedagogue - pedagogic, legal - legislature*
/t/ → /s/, e.g., *diplomat - diplomacy, democrat - democracy*
/t/ → /ʃ/, e.g., *act - action, correct - correction*
/d/ → /ʒ/, e.g., *divide - division, collide - collision*
/s/ → /ʃ/, e.g., *confess - confession, possess - possession*
/z/ → /ʒ/, e.g., *envisage - envision, televise - television*

What these examples show is that, disregarding the change in the manner of articulation and $\left\{\begin{matrix}/k/\\/t/\end{matrix}\right\}$ → /s/ change for the moment, both Alveolar and Velar consonants become Palatal due to the influence of the following /ɪ/ which we have seen as being definable also as Palatal. Thus, it is an evident case of assimilation, and I do not think that there is any doubt on this point whether from the point of view of historical sound change or of a synchronic phonology. Note, now, how this phenomenon may be

stated in DF terms. Since /i/ is $\begin{bmatrix}+\text{Diffuse}\\-\text{Grave}\end{bmatrix}$ and Palatal consonants are $\begin{bmatrix}-\text{Diffuse}\\-\text{Grave}\end{bmatrix}$, and since we know that the case in an assimilation, we must conclude that it is Gravity, not Diffuseness, that is assimilated. To assume otherwise is to say that it is a case of dissimilation, i.e., due to the influence of [+Diffuse] of /i/, non-Labial consonants become [-Diffuse]. We intuitively reject this assumption.

Consider now the rule Velar C → Palatal in DF terms:

$$(4)\qquad \begin{bmatrix}+\text{Grave}\\-\text{Diff.}\\ \text{C}\end{bmatrix} \rightarrow \begin{bmatrix}-\text{Grave}\\-\text{Diff.}\\ \text{C}\end{bmatrix} / \underline{\quad} \begin{bmatrix}-\text{Grave}\\+\text{Diff.}\\ \text{V}\end{bmatrix}$$

As far as the Gravity is concerned, the rule is expressed in an intuitively correct way; i.e., [+Grave C] changes into [-Grave C] in the environment of ___[-Grave V]. But note that we needed to specify one more feature, Diffuseness, throughout the rule, since Labial consonants which are also [+Grave] but differ from Velar consonants in the Diffuseness value do not become palatalized. This fact is unexplainable by the DF theory, since, when stated in DF terms, it should be the case that $\begin{bmatrix}+\text{Grave}\\+\text{Diff.}\end{bmatrix}$ need less assimilatory attraction to be assimilated than a segment of $\begin{bmatrix}+\text{Grave}\\-\text{Diff.}\end{bmatrix}$, as the environment which causes the assimilation is also a [+Diffuse] segment.

Examine next the rule Alveolar → Palatal in DF terms:

$$(5)\qquad \begin{bmatrix}-\text{Grave}\\+\text{Diff.}\\ \text{C}\end{bmatrix} \rightarrow \begin{bmatrix}-\text{Grave}\\-\text{Diff.}\\ \text{C}\end{bmatrix} / \underline{\quad} \begin{bmatrix}-\text{Grave}\\+\text{Diff.}\\ \text{V}\end{bmatrix}$$

This rule has nothing to do with Gravity, but is a case of a dissimilation of the feature Diffuse. That is, all segments involved are non-Grave and what changes is [+Diffuse] of Alveolar to [-Diffuse] due to the influence of [+Diffuse] of /i/. Thus, at best, the DF system must treat what is obviously one and the same assimilatory process as two unrelated processes; one, an assimilation of Gravity, and the other, a dissimilation of Diffuseness.[13]

---

[13]The similar point is briefly mentioned also in Ivić (1965) and Householder (1965). Chomsky and Halle (1965) attempt to describe these alternations by the following sequence of rules:

> The first rule applies to [k, g] . . . in certain contexts, changing the Gravity of the consonant to nonGrave; a second rule raises the nonGrave variant of /k/ to Diffuse; a third rule converts all of the nonGrave stops to Strident Continuants, in certain contexts. (pp. 122-123)

If this sequence of rules applies to underlying /k, g/ only, and another similar sequence is provided for palatalization of underlying

There is another theoretical aspect to consider regarding the present DF classification of places of consonantal articulation. It is a claim implicit in the notation that Labials and Palatals, and A Alveolars and Velars occur less often together, or constitute less natural "natural classes," than the rest of the two-feature combinations, since the members of the former set disagree from each other by two features, while those of the latter by only one feature. This is to say that, historically, there have been more cases of sound change of the latter type than the former, and that, perceptually, there is more confusion between the two items in the latter set than those in the first set. This kind of claim is in fact often made. For example, it is often cited, to justify the grouping of Labials and Velars together as [+Grave], that in English /x/ (spelt usually *gh*) changed to /f/, e.g., *laugh, tough, cough,* etc. But whatever the nature of this particular sound change may be, a mere fact that /x/ changed to /f/ in a number of cases does not provide a necessary and sufficient condition to assign a feature covering the two items involved. For instance, why isn't a DF provided covering Stops and Fricatives only, or Voiced and Voiceless only, phenomena much more abundantly attested as having common behavior? From Grimm's Law to the Palatalization rule, history

---

/t, d, s, z/, then, our argument still stands. That is, it is unmotivated to treat what is the same and one process in two separate ways. In this case, the collapsing is not a fortuitous simplification, but a statement of an empirical fact.

If, on the other hand, Chomsky and Halle's sequence of rules are to encompass the underlying /t, d, s, z/ as well in their path of derivations of the following form:

$$\begin{bmatrix} k \\ g \end{bmatrix} \rightarrow \begin{bmatrix} c \\ \text{ɟ} \end{bmatrix} \rightarrow \begin{bmatrix} t \\ d \end{bmatrix} \rightarrow \begin{bmatrix} s \\ z \end{bmatrix} \rightarrow \begin{bmatrix} \text{ʃ} \\ \text{ʒ} \end{bmatrix} \rightarrow \begin{bmatrix} \text{tʃ} \\ \text{dʒ} \end{bmatrix}$$

(1) (2) (3) (4) (5) (6)

it may be argued that

a. the environmental specification at each stage of derivation will probably become increasingly complex in order to filter out certain segments only, but not others; and this is not economical.

b. the sequence is circular without motivation. For example, (6) can be directly derived from (2) without having to go through the medial stages. Why go to Chicago from New York via San Francisco?

c. stage (3) has two kinds of /t, d/; one, underlying morphophonemic /t, d/, and the other, nonphonemic [t, d] derived from /k, g/ on their way to [ʃ, ʒ] or [tʃ, dʒ]. In this situation, I am not sure if the collapsing of this kind has any gains at all. For instance, are the environmental specifications of rules for both /t, d/ and [t, d] of /k, g/ the same? If different, the collapsing has no economy. If the same, then why not represent /t, d, s, z/ as /k, g/?

is full of instances of Stop → Fricative, or vice versa. But in the present form of the DF system, there is no feature which is common to them only: [+Cons.] includes other consonants, [-Nasal] includes vowels and liquids, etc. In fact, they differ in two features, Continuancy and Stridency, and need at least three features to set them apart from other segments: [-Vocalic, +Consonantal, -Nasal]. Then, what is the real significance of saying that, since /x/ became /f/, they show a mutual perceptual similarity, and that, therefore, we must provide a DF common to them? As far as the claim about the behavioral similarity of Labials and Velars is concerned, there are many counter-examples (not counter-examples in a strict sense, since the Gravity feature does not insist that Grave/nonGrave segments cannot be complementary allophones. Still, these examples fail to support the argument.). Among them:

(a) In Hawaiian, [t] and [k] are allophones of a non-Labial Stop contrasting with /p/.

(b) It is reported that some English speakers substitute [t] for /k/ in env. #___/l/, e.g., [tlin tlɔθs] for *clean cloths* (cf. Jones 1960; Ward 1929). Grant that English has no initial /tl/- cluster so that there would be no confusion even if /kl/ were replaced by /tl/. But a still remaining question is: why is [tl] identified as /kl/ rather than /pl/ which is closer to [tl] in DF terms?

(c) In many languages [ŋ] is an allophone of /n/ contrasting with /m/. If [ŋ] is distinctive-featurally closer to [m] than to [n], why isn't [ŋ] usually an allophone of /m/ rather than of /n/?

(d) In Japanese, allophones of /h/ are [ɸ], [f], [ç], and [h]. Excluding [h], the remainders are Labials and a Palatal, a loosely related pair in DF terms.

All this of course is an empirical matter, but as Ivić put it,

> In the consonantal quadrangle, a decisive demonstration would require a considerable number of unambiguous instances of contacts between dentals and labials, and labials and velars, and a proof that valid examples of cross contacts (velars and dentals, palatals and labials) are at least much less frequent. (1965, p. 59)

Without such "a decisive demonstration," the present DF classificatory schema of the consonantal square cannot be claimed to be valid.

We will now consider briefly the second claim that was made to justify the binarity of DF's: the claim that features are binary because the binary opposition is a child's first logical operation.

Jakobson's research into the child's earliest speech and the aphasic's speech, which was found to be a mirror image phenomenon of the former, indeed gave an insightful aspect of language acquisition and some fundamental structure of speech sounds. That a child's universal vocabulary, *mama, papa,* etc., is not an accidental onomatopoeia, but has a phonetic explanation has no doubt an important theoretical implication, which we will further explore in Chapter VII. But it is difficult to understand in what way a child's first operation in terms of optimal

opposition or maximal differentiation justifies the binary nature of adult's speech, or, in other words, in what sense optimal opposition implies binary opposition. Black is the color optimally opposed to white, and it is conceivable that a child differentiates the two colors most easily in his earliest years, but does this justify or imply the binary structure of color? Furthermore, it is not clear how a triangular structure appearing in *Fundamentals*, p. 40

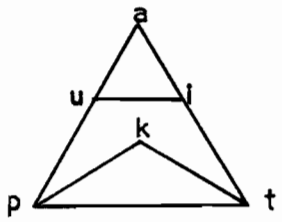



Figure 7. "The primary triangle" picturing the earliest oppositions acquired by the child (from *Fundamentals*, p. 40)

is to be interpreted as being a binary structure. Surely, nobody would claim that a triangle is a binary structure. Of course, it is not impossible to have a binary structure of a trinary set of items. An empirical evidence of the sort that Kim (1966) shows in the case of three series of Korean stops, two independent variables, at least, would be required to suggest or justify the binarity of a trinary phenomenon.

We examined above, in length, whether this was actually the case with certain phonetic categories, in particular, the DF's Diffuse/Compact and Acute/Grave, and found that these features are not independent variables. We thus formally reject the claim that "the inherence of the dichotomous scale in the linguistic system is quite manifest" (*Fundamentals*, p. 49).

It is now time to consider the defining properties of DF's.

Since the beginning of phonetics, a standard practice was to base phonetic categories on physiological and articulatory facts. Then, with the advent of acoustic phonetics, there appeared some considerations of a phonetic theory with categories based on the acoustic properties of sound. The DF theory is one such. As Ivić put it:

> Jakobson revolutionalized the approach to articulatory phenomena. He replaced the primitive and easy classifications in terms of place of articulations by those based on the elements truly relevant for the properties of the sound -- the shape and the size of resonators. (1965, p. 72)

Despite some claims that are made otherwise today, this was the foundation on which the DF theory was first built. That this is true is expressed in *Preliminaries*. I quote one paragraph in full:

> A distinctive feature cannot be identified without recourse to its specific property. . . . But to which of the consecutive stages of the sound transmission shall we refer? In decoding the message received (A), the listener operates with the perceptual data (B) which are obtained from the ear responses (C) to the acoustical stimuli (D) produced by the articulatory organs of the speaker (E). The closer we are in our investigation to the destination of the message (i.e. its perception by the receiver), the more accurately can we gage the information conveyed by its sound shape. This determines the operational hierarchy of levels of decreasing pertinence: perceptual, aural, acoustical, and articulatory (the latter carrying no direct information to the receiver). The systematic exploration of the first two of these levels belong to the future and is an urgent duty. (p. 12)

The last sentence in the quotation clearly indicates that the DF theory was originally intended to be an acoustic theory of phonetics. Several DF terms such as Flat, Sharp, etc. also suggest the acoustic properties of DF's.

But today, a voice is heard to the effect that DF's are actually perceptual categories, not acoustic, and that, since there is no one-to-one correspondence between perception and physical data, and since a single perceptual feature may be definable only with an awkward disjunction at the acoustic or articulatory level, it is not a matter of importance that DF's are not acoustically or physiologically definable in coherent terms. That this is a dangerous assumption was discussed in the preceding chapter and earlier in this chapter. Here, we examine another aspect of what it might mean to assert that DF's are perceptual features.

To say that DF's are perceptual categories begs a question: whose perceptual categories are they? That is, who is to say that [ø] is different from [e], a Frenchman, a Slovak, or a Russian? It is reported that a monolingual Slovak perceives French /ø/ as /e/, whereas a Russian perceives the same sound as /o/ (*Preliminaries*, p. 10). This shows that "the way we perceive speech sounds is determined by the phonemic pattern most familiar to us" (ibid.). What does it mean, then, to say that a perceptual DF theory is a framework of universal phonetics? If DF's are perceptual and perception depends on the particular phonemic pattern of languages, then to talk about DF's as a universal phonetic alphabet is nothing but imagination, unless the theory provides an explicit measure by which items in different languages can be equated or differentiated properly, whichever the case may be, independent of a perceptual pattern particular to a specific languages. For example, the theory must tell how French /ø/ is to be differentiated from Slovak /e/ or from Russian /o/, even though different perceptual groupings render two pairs of them as the same; or whether English /p/ and French /p/ are to be equated or differentiated, and on what grounds. This brings up the question raised in Chapter II: how do we identify phones cross-linguistically? We examine it here.

The DF theory provides one equation-formula for such a purpose:

> The fact that peoples who have no pharyngealized consonants in their mother tongue, as, for instance, the Bantus and the Uzbeks, substitute labialized articulations for the corresponding pharyngealized consonants of Arabic words, illustrates the perceptual similarity of pharyngealization and lip-rounding. These two processes do not occur within one language. Hence, they are to be treated as two variants of a single opposition - flat vs. plain. (*Preliminaries*, p. 31)

That is, whenever two phones are compared, if they occur as contrastive phonemes in a language, then they are to be specific with different features; otherwise, they are to be regarded as variants of the same feature. Thus, [ø] [e] and [o] are said to be manifestations of different feature combinations because at least French distinguishes all three, but labialization and pharyngealization are said to be mere variants of one feature because no language distinguishes them, etc. What this amounts to saying is that: find the phonemically richest language, probably in each DF, and use it as a model and a criterion in deciding the sameness or difference of two or more cross-linguistic phones. Essentially, this in an extension of the principle of complementary distribution, and extension of its application from the intralingual phonemic analysis to the interlingual universal phonetics, and as a first approximation, it may be a workable measure, just as the principle proved to be useful for a while in taxonomic phonology. But also just as the application of the principle to the logical extreme in the intralingual case would yield such an intuitively unacceptable result as grouping [h] and [ŋ] as allophones of one phoneme of English, so will the principle produce the similar results in interlingual cases. And, no doubt, this was the case when it was asserted that pharyngealization, labialization, velarization, retroflexion, all belong to the one and the same feature [+Flat] (Halle 1957, p. 67). But then again, just as taxonomic phonemicists soon realized that an additional criterion, namely, phonetic similarity, is needed to rule out such cases as [h~ŋ], so did the DF proponents, and we hear (from R. Wilson's correspondence with Halle) that the position of the archiphonemic character of Flat is no longer held. So we might modify the principle of complementary distribution and add a condition: if the phones are "phonetically similar." But this condition is vacuous and circular, because a set of criteria for interlingual "phonetic similarity" is precisely what we want to find.

Ladefoged suggests two criteria which are intended to define this interlingual "phonetic similarity":

> If two features are to be coalesced and regarded as variants of the same feature, they must both be members of the same type set. Only members of the same type set commute and can be guaranteed not to co-occur. I would also like to suggest a second criterion: two or

> more phenomena can be subsumed under a single feature if and only if they can be regarded as points on the continuum of that feature and can be described by numbers specifying the amount of the feature which they possess. (1965, p. 33)

A set of physiological type sets, such as nasal, stop, fricative, trill, tap, flap, etc., is assumed to be given as primitives, and the first condition for assuming the phone A of $L_a$ and the phone B of $L_b$ as two variants of one feature is that they belong to the same type set. The second condition amounts to the distinction between the difference in degree and the difference in kind (cf. Jakobson's "contrary opposition" vs. "contradictory opposition," 1939/1962a, p. 273). Two phones that meet the first condition will still be regarded as belonging to one feature only if the difference between the two is a matter of quantitative difference, not of qualitative difference. Obviously, these two conditions are more severe than the cited DF's criterion. For example, Ladefoged's second criterion will rule out the possibility of grouping labialization and pharyngealization as variants of one feature, since they are not neighboring points on a continuum, even though the two may belong to the same type set, secondary articulation, and hence meet the first condition.

At times, however, it is difficult to decide whether a given difference is quantitative or qualitative, in degree or in kind. For instance, is the difference between [t] and [k] qualitative or quantitative?; is the difference between [i] and [a] in kind or in degree? Whether we speak in terms of articulatory positions or in terms of acoustic loci of formants, the above sounds may be regarded as points on the continuum, and, hence, as manifesting quantitative difference, not qualitative.

We see thus that (1) there is a need to put a tighter constraint on the *criteria for interlingual phonetic similarity;* (2) the notion of "natural class" (and the notion of degree of naturalness) can not be adequately expressed in DF terms; and (3) the DF system has not yet *incorporated universal restrictive redundancy* as an inherent structure of the system, thereby yielding an enormous redundancy and making P rules unnecessarily complex. (1) is essential for a phonetic theory to serve as a universal framework, (2) is pivotal in the formulation of an evaluation measure, and (3) is necessary for the simplicity of the theory. A theory that does not provide proper and satisfactory ways to define any or all of these fails to that extent as an explanatorily adequate theory. I maintain that the current DF theory fails in this respect. This calls for a new modified theory of universal phonetics. We propose one such model of a first approximation in the following. I will see later whether and how the proposed model approaches the level of the explanatory adequacy better than does the DF theory.

For our model of universal phonetics, we make the following assumptions.

(i) There is a close correspondence between the articulatory, acoustic, and perceptual levels of sounds, so that a phonological description can be made at any stage provided that there are conversion rules.

(ii) Phonetic categories are not necessarily binary-structured.

(iii) Phonetic categories are ordered in such a way that it is possible to measure "the phonological distance," or the relative position of a given category in the phonological hierarchy, and to define the notion "optimal opposition" or "maximal differentiation."

Assumption (i) is valid, since, even though there is no isomorphism, "each of the consecutive stages, from articulation to perception, may be predicted from a preceding state" (*Preliminaries*, p. 12). As Ladefoged (1966) put it,

> Whenever man in general perceives linguistic items as belonging to the same group it is because these items have some common simple physical correlate. . . . Subjects usually consider an item which is, on a physical scale, in between two others, is also, on a psychological scale, correspondingly ordered.

Which stage of description, then, shall our model select? We have seen some difficulty in defining phonetic categories in terms of perceptual similarity. Until considerably more about the aural structure and neurophysiological behavior is known, a perceptual phonetic theory is a remote feasibility. A universal framework in terms of acoustic categories is now possible, as the development of modern acoustics is probably capable of describing even the most complex sound waves in terms of frequency, amplitude, and duration (cf. Fant 1960). We regard the DF theory as an essentially acoustic theory. It is of course not the only way. Vowels, for instance, may be categorized in terms of relative values of two formants as follows:

| | i | e | a | o | u |
|---|---|---|---|---|---|
| F1 | Low | Mid | High | Mid | Low |
| F2 | High | High | Mid | Low | Low |

Figure 8. A possible categorization of vowels in terms of the relative values of the first two formants

None the less, I would like to propose a universal phonetic framework whose categories are articulatory. It is true that sound is an acoustic phenomenon, and as such, it is most appropriate to describe it in terms of acoustic parameters. But it is also true that, while acoustic variability is infinite, the range of possible human speech waves is considerably smaller due to the inherently limited capabilities of the dynamics of the human vocal tract (and also of the aural structure). Furthermore, the unmistakable fact that the shape and the dynamics of the vocal tract are uniform for speakers of all languages makes the organization of "speech" sounds in terms of articulatory categories simpler, more practical, more convenient, and more reasonable than the theory dictates. And if any phonetic framework must be capable of discrete symbol representation of continuous reality, nowhere is the "discretization" more easily, more naturally, and less arbitrarily done than in the physiological structure of the vocal tract. For example, Bilabial, Dental,

Alveolar, etc., have more natural boundaries between them than, say, Diffuse/Compact, Acute/Grave, etc. Auditory perceptual categories are likely to be as discrete as physiological categories, but, as was said earlier, there is no easy way to establish universal perception to make perceptual categories a serious candidate for a universal phonetic alphabet. The development of neurophysiology and psychoacoustics will no doubt give some insights into the structure of sensation, but until then, talking about classifying speech sounds in terms of universal auditory categories is as ambiguous as talking about classifying the ranges of color in terms of universal categories of visual perception.

Assumption (ii) enables us to establish as few or as many categories as are empirically necessary. For example, Air-direction, Nasality, etc. will have two subcategories respectively, as they are binarily opposed in reality. But we will have to recognize more categories in the degrees of constriction, in the places of articulation, etc.

Assumption (iii) has never been explicitly adopted or stated in any phonetic theory so far proposed. But only its acceptance makes it possible to formalize and define such meaningful and important notions as "phonological distance," "featural hierarchy," "optimal opposition," etc. We will see later how our model incorporates these notions as an essential part of the theory.

Our model of universal phonetics will attempt to describe speech sounds with five articulatory parameters:

(1) the degree of aperture (D)
(2) the place of articulation (P)
(3) the manner of production (M, =secondary articulation)
(4) the glottal state (G)
(5) the air direction (A)

Each parameter is divided into several *Macrocategories*, and each macrocategory branches into *Subcategories* (or, simply *Categories*), which, in turn, may or may not have their own *Microcategories*.

The categories of the degree of aperture are:

| Macro-categories | Subcategories | | | Microcategories | |
|---|---|---|---|---|---|
| | degree of aperture | descriptive term[14] | phonetic term | degree | term |
| Consonantal | 0 | contact | stops | 01 | affricates |
| | 1 | occlusion | fricatives | | |
| Sonantal | 2 | obstruction | liquids | 12 | fricative liquids |
| | 3 | constriction | approximants | | |
| Vocalic | 4 | close | high vowels | 45 | half-close |
| | 5 | close-open | mid vowels | | |
| | 6 | open | low vowels | 56 | half-open |

Table 2. The categories of the degree of aperture

As is shown in the chart, the categories are ordered in terms of the degree of aperture, from 0 degree to 6. Consonants, liquids, the so-called semi-vowels, and vowels are distinguished by means of the different degrees of aperture.

Categories of the degree of aperture have the following hierarchy:

(6) (i) DEGREE → Consonantal ( Sonantal ) Vocalic
(ii) Consonantal → 0 ( 1 )
(iii) Vocalic → 4 ( 5 ) 6
(iv) Sonantal → 2 ( 3 )

This hierarchy implies "typological universals" that Jakobson (1958/1962c) referred to. That is, (i) asserts that every language must have at least two macro-degrees of aperture, Consonantal and Vocalic. The third macro-degree Sonantal may be chosen only if the other two have already been chosen. (This is the meaning of the parentheses.) (ii) asserts that there are two degrees of consonantal aperture, 0 and 1, but if a language has only one, it must be 0, not 1, i.e., no language has fricatives without stops. (iii) asserts that every language must have at least two degrees of vocalic aperture, 4 (close) and 6 (open), and that 5 (mid vowels) may be chosen only if 4 and 6 have already been chosen. Thus, it asserts that no language may have mid vowels without having high and low vowels. (iv) asserts that, of two possible Sonantal degrees of aperture, 3 presupposes 2, but not vice versa. (This seems to be true. For example, Korean, Japanese, Tausug, etc. have liquids, but no approximants (or semi-vowels).)

The following are categories of the place of articulation. Each Subcategory will be designated with a capital letter, and each Macrocategory with two capital letters of which the first one will be the symbol of a Subcategory to which a given Macrocategory belongs.

---

[14]The terms "contact, occlusion, obstruction, constriction" are borrowed from Halle (1964a) with a slight modification, and the term "approximant" is first suggested in Ladefoged (1964).

| Macrocategories | Subcategories | | Macrocategories | |
|---|---|---|---|---|
| | term | symbol | term | symbol |
| Labial | Labial | L | bilabial | LB |
| | | | labio-dental | LD |
| Lingual | Alveolar | A | dental | AD |
| | | | A-proper | AA |
| | | | post-alveolar | AP |
| | Palatal | P | pre-palatal | PR |
| | | | P-proper | PP |
| | Velar | V | pre-velar | VP |
| | | | V-proper | VV |
| | | | uvular | VU |
| Laryngal | Glottal | G | pharyngal | GF |
| | | | G-proper | GG |

Table 3. The categories of the place of articulation

Categories of the place of articulation have the following hierarchy:

(7) (i) PLACE → Labial + Lingual (Laryngal)
(ii) Lingual → Alveolar (Palatal) Velar
(iii) Alveolar → AD + AA (AP)
(iv) Velar → (VP) VV + VU
(v) Labial → LB + LD
(vi) Palatal → (PR) PP
(vii) Laryngal → (GF) GG

The implication of this hierarchy is the same as in the case of the degree-hierarchy. For example, (i) asserts that every language must distinguish at least two places of articulation, Labial and Lingual (e.g., Hawaiian), and (ii) asserts that Palatal presupposes Alveolar and Velar, but not vice versa, etc.

The categories of the manner of production or the secondary articulation, which we will symbolize with lower case letters, are:[15]

[15] I am not at all sure what the hierarchical structures of these categories would be like, except that the primary opposition should

| Macrocategories | | Subcategories | |
|---|---|---|---|
| term | symbol | term | symbol |
| nasal | n | | |
| oral | o | (spread) neutral | s |
| | | labialized | l |
| | | retroflexed | r |
| | | palatalized | p |
| | | velarized | v |
| | | pharyngalized | f |
| | | glottalized | g |
| | | tense | t |
| | | lax | x |

Table 4. The categories of the secondary articulation

The categories of the glottal state are:

| Macrocategories | | Subcategories | |
|---|---|---|---|
| term | symbol | term | symbol |
| voiced | + | voiced-proper | d |
| | | creaky | c |
| | | murmur | m |
| | | whisper | w |
| voiceless | - | aspirated | a |
| | | unaspirated | u |

Table 5. The categories of the glottal state

---

be oral vs. nasal. I am also undecided as to whether "lateral, trill, tap, flat" should be included here or in Degree as microcategories of Liquid. It is also uncertain how "nasal stop" and "nasalized stop" are to be distinguished. Another feasible alternative to deal with nasals and laterals would be to set up another macrocategory called "secondary aperture," analogous to secondary articulation, whose subcategories are nasal and lateral. In any case, the assertion made in this proposal should be regarded as tentative. Hierarchy typology is, as many other linguistic aspects are, an empirical matter, and an explicit formalization must wait for investigations of a number of more languages.

The categories of the air direction involve only two Macrocategories: egressive (e,←) and ingressive (i,→).

A stretch of speech sounds may be represented in a matrix form where columns represent segments and rows categories, as is shown in the following, or in a linear form in which every numeral signals the beginning of a segment, e.g., /OLo- 6VPx OLn OVo- 4Px OAn/ 'pumkin'. In the following are given matrices of English phonemes and some selected sounds from various languages (especially from African languages. For detailed phonetic descriptions, see Ladefoged 1964, and forthcoming.), with redundancies omitted.

| | p | b | t | d | k | g | m | n | ŋ | f | v | θ | ð | s | z | ʃ | ʒ | tʃ | dʒ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| D | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 01 | 01 |
| P | LL | LL | AA | AA | VV | VV | LL | AA | VV | LL | LL | AD | AD | AA | AA | P | P | P | P |
| M | o | o | o | o | o | o | n | n | n | | | | | | | | | | |
| G | - | + | - | + | - | + | | | | - | + | - | + | - | + | - | + | - | + |

| | h | l | r | w | y | i | ɪ | e | ɛ | æ | ɑ | ʌ | ɜ | ə | ɔ | o | ɷ | u |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| D | 1 | 2 | 2 | 3 | 3 | 4 | 4 | 5 | 5 | 6 | 6 | 6 | 5 | 5 | 5 | 5 | 4 | 4 |
| P | G | A | A | V | P | P | P | P | P | P | VP | VP | VP | VP | VV | VV | VV | VV |
| M | | s | r | | | t | x | t | x | | t | x | t | x | t | x | t | x |

Table 6. Matrix of English phonemes

| | n̥ | l̥ | t | t* | th | b̤ | a̤ | m̤ | b̰ | w̰ | o̰ | ʔ | ɾ̃ | ɣ | p' | t' | k' | ɓ | ɗ | ɠ | kp | gb |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| D | 0 | 2 | 0 | 0 | 0 | 0 | 6 | 0 | 0 | 3 | 5 | 0 | 4 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| P | A | A | A | A | A | L | P | L | L | V | V | G | P | P | L | A | V | L | A | V | L | L |
| M | n | | x | t | t | o | o | n | o | | | | n | l | g | g | g | g | g | g | v | v |
| G | - | - | | u | a | m | m | m | c | c | c | | | | - | - | - | + | + | + | - | + |
| | Bur-mese | | | Korean | | Gujer-ati | | Ndebele | Margi | | Langoc | Tagalog | French | | Amharic | | | Sindhi | | | Yoruba | |

| | ɻ | ɭ | ɬ | ɸ | f | t̪ | t | ʈ | ʕ | h | ɦ | ɮ | ɬ | ɕʷ | ʃʷ | tʲ | pʲ | zᵃ | sᵃ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| D | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 3 | 12 | 12 | 1 | 1 | 0 | 2 | 1 | 1 |
| P | AD | AA | P | LB | LD | AD | AA | AP | GF | GF | GF | A | A | P | P | A | A | A | A |
| M | v | v | v | | | | | | | | | | | 1 | 1 | p | p | f | f |
| G | - | - | - | - | - | - | - | - | + | - | m | + | - | - | - | - | | + | - |
| A | i | i | i | | | | | | | | | | | | | | | | |
| | Zulu | | | Ewe | | Malayalam | | | Arabic | | | Zulu | | Twi | | Russ. | | Tamazight | |

Table 7. Matrix of selected foreign phonemes

We will now see in what ways our model fulfils the assertions made earlier.

(1) How does the model provide a criterion for comparing cross-linguistic sounds?

We will say that phone A of $L_a$ and phone B of $L_b$ shall be regarded the same if and only if the union of phonetic categories of A and B coincides their intersection, i.e., if A and B have the same set of phonetic categories. Categories of a phone are those that are minimally needed to specify the phone. The criterion applies both at the systematic phonemic level and the systematic phonetic level. But at the first level, redundant categories are excluded, from comparison. For example, English /p/ will be regarded as oppositionally the same as French /p/, since both /p/'s have the same set of phonetic categories /OLo-/ at the systematic phonemic level. But Korean /pʰ/ will not be identified the same as English /p/, even though Korean /pʰ/ is phonetically closer to English /p/ than French /p/ is, since in Korean voiceless unaspirated and aspirated stops contrast, and the aspirated /pʰ/ has a specification /OLoa/ which is different from the English /p/ specification. But at the systematic phonetic level where redundant categories are all specified, English [pʰ] will now be equated with

Korean [$p^h$], as both have the same [OLoae], but French [p] will no longer be identified the same as English [$p^h$], as it has a different specification [OLoue].

(2) How does the model define the notion "natural class" and provide a measure of the degree of naturalness of the class?

We shall say that a set of phones having one or more phonetic categories in common belong to a natural class. The degree of generality or naturalness of natural class is determined (i) by the position of the category defining the class in the tree of the oppositional hierarchy, and (ii) by the number of categories needed to specify the natural class. The higher in the hierarchy, the more general, and the larger the number of categories, the less general the natural class. For example, a natural class of all Lingual phones is more general or more inclusive than a natural class of all Alveolar phones, since the category Lingual is higher up in the hierarchy than the category Alveolar, and a natural class of all contact sounds (stops) is a more general natural class than that of all voiceless contact phones, since the latter class is defined with two categories (0-), while the former with only one (0).

We maintain that this definition and the measure are more workable than those provided by the DF system. Consider an observation by Joos (1950), "/aw/ is never followed by /p, b, m, f, v/ in English." Clearly these phonemes constitute a natural class Labial. But consider how this class may be expressed by DF terms. $\begin{bmatrix}+\text{Grave}\\+\text{Diff.}\end{bmatrix}$ will not do, since it will include [+Grave +Diffuse] vowels. Thus, at best, it requires four features [-Vocalic, +Conson., +Grave, +Diffuse] to define a simple natural class Labial. We argue that this is counter-intuitive.

(3) How does the model define the notion "optimal opposition"? We will discuss this in more detail in Chapter VII.

(4) How does the model handle phonological redundancy, and how does it fare in terms of the simplicity criterion?

I propose that there be formalizations of two kinds of phonological redundancy rules: Universal Redundancy (UR) rules and General Redundancy (GR) rules. UR rules are those that were mentioned earlier and referred to as Restrictive R rules. In the sense that these rules are due to the inherent physiological restrictions, they are rather conditions than rules. That is, these rules have no role in the P rules except that they are metatheoretical conditions, since no P rule or sound change involves a change of these universally redundant categories. For example, an R rule

(8) Vocalic → Lingual

simply states an inherent condition that no vowel can be a Labial or Laryngal. Since no linguistic change, diachronic or synchronic, ever effects this condition, Stanley's (1966) proposition that "fully

specified matrices" be available before P rules is not justified as far as universally redundant features are concerned.[16] It seems that only at the level of systematic synthesis, the universal redundancy comes into playing a role. For example, a R rule

(9) Vocalic → Voiced

will be needed to turn on the switch of the fundamental tone control of an acoustic synthesizer in order to synthesize a normal vowel. This leads us to speculate whether UR is rather a part of systematic synthesis than a part of systematic phonemics. In the present form of the phonological component of a generative grammar, R rules are placed at the beginning of the component. But as was just observed, some of these R rules, namely, UR rules, have no theoretical motivation for being there, since no P rule will affect categories to violate the implications of UR rules. This means that UR rules may apply at any stage before systematic synthesis. Suppose, then, that we place UR rules at the beginning of systematic synthesis. One consequence of this is that, without inherent restrictions, P rules may generate combinations of features or categories which are quite meaningless and unphonetic. For example, the following P rule

(10) $\begin{bmatrix}\text{Degree 1}\\ \text{Voiced}\end{bmatrix} \rightarrow 2/X$

will produce z → l, ʒ → ɮ, j → ʎ, ɣ → ɬ, *and* *Labial liquid and *Laryngal liquid. How are we to handle these kinds of cases?

I suggest that UR rules, as a part of systematic synthesis, function as 'blocking rules' in such a way that any category-combinations, generated by P rules, that do not conform with UR conditions, will be blocked from going into the synthesizer as impossible speech sounds, analogous to a device in syntactic component where transformational rules function as a sieve filtering out only grammatical strings and blocking ungrammatical strings generated by context-free Base component. This is just a speculative suggestion requiring further investigation, but it seems that this set-up may have more theoretical motivation than the usual practice. Let us see in more detail how this might be true.

One consequence of placing R rules at the beginning of the phonological component before P rules is that there is no way to verify

---

[16]However, redundant features seem to be needed to specify environments that are necessary in the structural description of certain P rules. For example, a P rule that has Voiced as its conditioning environment, e.g.,

$\begin{bmatrix}\text{Fricative}\\ \text{Voiceless}\end{bmatrix} \rightarrow$ Voiced / ___Voiced (whether a vowel or a consonant)

will not work properly if a UR rule Vocalic → Voiced is to come after the P rules. I have no remedy to this problem at the present time.

whether or not R rules have been violated in the course of the application of P rules, unless we recycle R rules to the output of P rules. Consider, in particular, the following R rule in Korean stops:

(11) Lax → Unaspirated

Suppose now a P rule

(12) Tense → Lax

changes a Tense segment into Lax. By reapplying R rule (11) to the output of P rule (12), we will be able to specify the segment correctly regardless of whether the original Tense segment is Aspirated or Unaspirated. Thus, the measure of recycling R rules seems to achieve some economy, since without such measure we have to state P rule (12) as

(13) $\begin{bmatrix}\text{Tense}\\ \text{Aspirated}\end{bmatrix} \rightarrow \begin{bmatrix}\text{Lax}\\ \text{Unaspirated}\end{bmatrix}$

But this measure begs another problem. For example, consider another R rule in Korean:

(14) Consonantal → nonPalatalized

(or in DF terms, $\begin{bmatrix}\text{-Voc}\\ \text{+Cons}\end{bmatrix} \rightarrow$ [-Sharp])

and a later P rule:

(15) $\begin{bmatrix}\text{Consonantal}\\ \text{Alveolar}\end{bmatrix} \rightarrow \text{Palatal} / \_\_ \begin{bmatrix}\text{Vocalic}\\ \text{Palatalized}\end{bmatrix}$

(or in DF terms, $\begin{bmatrix}\text{+Diff.}\\ \text{-Grave}\end{bmatrix}$ C $\rightarrow$ [+Sharp] / ___[+Sharp V])

Now if we apply R rule (14) to the output of P rule (15), Palatal consonants will change back to nonPalatal. Since there is no way of knowing the derivational history of P rules just by looking at the output matrix, there is no way to prevent the reapplication of R rule (14) from nullifying P rule (15).

Stanley (1966), noting the problem, speculates that there might be a "natural breaking point" in the sequence of P rules which tells us just what the domain of R-rule recycling is, and that, if not, each P rule might be marked as to the reapplicability of R rules to its output. By the latter measure, (12) will be marked positively, but (15) negatively as to the R-rule reapplicability. But this is an arbitrary measure.

I feel that by distinguishing R rule into two kinds, Universal and General (of which more will be said soon), and by placing UR rules at the end of P rules, i.e., at the beginning of systematic synthesis, we may overcome this problem. Placing certain R rules at the end of P rules is not as revolutionary as it might seem, since "recycling" means, after all, reapplication of R rules to the "output" of P rules.

If the above speculation is correct, then we have found one more important role of the level of systematic synthesis: blocking a string whose segment(s) contain combinations of phonetic categories that violate UR conditions, i.e., those combinations non-convertible to speech sounds.

GR rules are those that apply in general and hence require no particular specification, e.g., Segment → egressive, Vocalic Velar (back vowels) → labialized (rounded), etc. These rules are, however, not universal, since there is no inherent reason why these should always be so. In fact, some sounds violate these rules, e.g. clicks, unrounded back vowels, etc.

We shall say that phonetic categories which are implied by GR rules will not be specified, but only those phones which do not accord with GR rules will be explicitly specified in the phonemic matrices. Thus, for example, egressive alveolar stop will be /0A/, but ingressive alveolar stop /0Ai/; rounded close back vowel, /4V/, but unrounded close back vowel /4Vs/, etc.

This convention enables us to achieve simplicity and to apply the evaluation measure in an intuitively more correct way.

To take an example, consider the case of the "Diphthongization rule."

$$(16)\quad \emptyset \rightarrow \begin{bmatrix} -\text{Voc} \\ -\text{Cons} \\ \alpha\text{Grave} \\ \alpha\text{Round} \end{bmatrix} \quad / \quad \begin{bmatrix} +\text{Voc} \\ -\text{Cons} \\ +\text{Tense} \\ +\text{Diff} \\ \alpha\text{Grave} \\ \alpha\text{Round} \end{bmatrix} \text{———}$$

That is, if values of Gravity and Rounding of a high tense vowel agree, then the vowel takes a diphthong whose Gravity and Rounding also agree with the vowel, i.e., /i/ takes /y/, and /u/ takes /w/. It looks superficially as if the agreement condition is very specific and delicate. But actually, back vowels are generally rounded, and front vowels unrounded; that is, Rounding is a GR feature of Gravity. If this phenomenon is agreed upon, then the feature Rounding can be left out of the rule, and, hence, we achieve an economy.

To see how this is related to evaluation measure, consider a case of the so-called "alpha-switching," which is to explain such alternation pairs as *goose - geese, tooth - teeth, foot - feet, mouse - mice* (/mūs - mīs/) *louse - lice, ring - rung, sing - sung,* etc. Features whose polar values are switched by the rule are, as in the case of Diphthongization, Gravity and Rounding, since the alternation is between /i/ and /u/. Compare now this rule with the following which also involves two-feature change:

$$(17)\quad \begin{bmatrix} \alpha\text{Grave} \\ \alpha\text{Nasal} \end{bmatrix} \rightarrow \begin{bmatrix} -\alpha\text{Grave} \\ -\alpha\text{Nasal} \end{bmatrix}$$

If the segment is C, the switching is between /m, ŋ/ and /t, c/, and if V, between /ũ, õ/ and /y, œ/. It is evident that there is a great difference in the degree of change between the two cases, even though both involve two-feature change. The reason is, needless to say, Gravity and Nasality are two truly independent features, whereas Rounding is merely a concomitant feature, or a GR feature, of Gravity (in the case of vowels only, of course). Nevertheless, our evaluation measure in terms of the symbol-counting will evaluate the two rules as yielding the change of the same degree, despite an enormous difference in the "phonological distance" (of which we will have more to say below) in the changes produced by the two rules. This counterintuitive measure can be corrected if we say that the first rule involves only one-feature change, the GR feature being unspecified in the rule, hence, not counted as a symbol in symbol-counting evaluation.[17]

One might ask where we draw a boundary line between General Redundancy and other language-specific redundancy. What is the criterion? Is it purely a statistical matter? Or is there any theoretical motivation? To take an example, if we say that an R rule "back vowels are rounded" is a GR rule, but that another R rule "fricatives are alveolar" (in Korean) is a language-specific R rule, is it because the former R rule is statistically more often found than the latter? If so, what is the reference point? Fifty percent of the languages of the world? Or, is there any other criterion for GR? Since the answer is tied with the notion "optimal opposition," we will discuss the question in Chapter VII.

We will now ask the final question of the chapter.

(5) How does the new model define the notion "phonological distance," and employ it as an evaluation measure?

We will define the phonological distance between phone A and phone B as the difference of union and intersection of the two sets of phonetic categories of the two phones, i.e.,

$$d = ( A \cup B ) - ( A \cap B )$$

The larger the difference, the greater the phonological distance, and vice versa. When the difference is zero, that is, when the intersection equals the union, there is no distance between the two, i.e., the two phones are identical. This is the case of phonetic identity defined earlier. The exact computing system may be devised in a similar way to that suggested in Peterson and Harary (1961), allowing different scores according to whether the differing categories are Macro-, Sub-, or Microcategories, etc. But here, we omit the details, and instead, will examine an example or two to see some of the implications of the notion phonological distance.

---

[17] This measure seems to be equivalent to what is now occasionally mentioned as "Marked vs. Unmarked" feature, reportedly developed as a new evaluation measure at MIT. But nothing is available in print yet.

Consider the case of [l] → [j] → [i] and [ɬ] (velar l) → [w] → [u].

E.g.; Lat. *filia* → Fr. *fille* [fi:j] 'girl'; Lat. *pulmo* → Fr. *poumon* 'lung'
Cf. French *travail* [travaj] 'work' but plural *travaux*
*cheval* 'horse' but plural *chevaux*
English: *half* (/half/ → /haɬf/ → /hawf/ → /ha:f/)
Also cf. *folk, talk, should, could, would,* etc.
Low German: *hell* 'clear; light' /heɬ/ → /hew/
*Kalf* 'calf' /kaɬf/ → /kawf/ cf. English /ha:f/
Bavarian dialect of German: *Holz* 'wood' /holts/ → /hoits/
Italian: *planu* 'plain' → *piano; flamma* 'flame' → *fiamma*
West Polish: *lau* 'field' /ɬau/ → /wau/

There are excellent phonetic explanations for these changes. For an explanation from a genetic point of view, see Jones (1960), for an acoustic explanation, see von Essen (1964). The examples are from von Essen.) From the viewpoint of our model, this is a simple case involving a gradual change of the degree of aperture, from 2 to 3 then to 4, with the place of articulation and other phonetic categories largely intact. That is, the phonological distance between these three phones is minimal. The rule in DF terms, however, involves the reversing of polar values of two fundamental, hierarchically high features, i.e., for [l] → [j],

(18) $\begin{bmatrix}+\text{Vocalic}\\+\text{Conson.}\end{bmatrix} \rightarrow \begin{bmatrix}-\text{Vocalic}\\-\text{Conson.}\end{bmatrix}$

This rule expresses the process as a far more complex sound change than it really is, and we argue that it is intuitively incorrect.

That Liquids and Glides are phonologically similar, not optimally opposite as the DF specification suggests, is also evidenced by a MS structure in English where only Liquids and Glides can occur after a stop in the initial consonant cluster, i.e., $(s)\begin{Bmatrix}p\\t\\k\end{Bmatrix}\begin{Bmatrix}l\\r\\w\\y\end{Bmatrix}$ (except *pw- and *tl-); and by the following phonological pattern in Tamil (cf. Firth 1934; Troubetzkoy 1949, p. 160):

| | | | | |
|---|---|---|---|---|
| p | t | ṭ | c | k |
| w | l | ḷ | y | ʀ |

where the Sonantals corresponding symmetrically to stops include both Liquids and Glides. To express this natural class Sonantal, the DF system must use two features with the alpha-notation, i.e., $\begin{bmatrix}\alpha\text{Vocalic}\\\alpha\text{Consonantal}\end{bmatrix}$.

Consider now the Vowel Lowering rule in Vulgar Latin:

ĭ → e  ŭ → o

ĕ → ɛ  ŏ → ɔ

i.e., short stressed vowels are lowered by 1 degree of aperture. This phenomenon can be expressed by a single rule in our format:

$$(19)\quad \begin{bmatrix} \alpha\text{Degree} \\ \text{Vocalic} \end{bmatrix} \rightarrow \alpha - 1\ \text{Degree} \;/\; \begin{bmatrix} \overline{\text{Short}} \\ \text{Stress} \end{bmatrix}$$

This rule is statable in DF terms as follows:

$$(20)\quad \begin{bmatrix} +\text{Diffuse} \\ \text{V} \end{bmatrix} \rightarrow [-\text{Diffuse}] \;/\; \begin{bmatrix} \overline{-\text{Long}} \\ +\text{Stress} \end{bmatrix} \quad (\text{for } \begin{bmatrix} \text{i} \\ \text{u} \end{bmatrix} \rightarrow \begin{bmatrix} \text{e} \\ \text{o} \end{bmatrix})$$

$$(21)\quad \begin{bmatrix} -\text{Diffuse} \\ -\text{Compact} \\ \text{V} \end{bmatrix} \rightarrow [+\text{Compact}] \;/\; \begin{bmatrix} \overline{-\text{Long}} \\ +\text{Stress} \end{bmatrix} \quad (\text{for } \begin{bmatrix} \text{e} \\ \text{o} \end{bmatrix} \rightarrow \begin{bmatrix} \varepsilon \\ \text{ɔ} \end{bmatrix})$$

That this pair of rules lacks a generalization and elegance as compared to rule (19) is self-evident.[18]

The case of the so-called "rhotacism," [z] → [r], e.g.,

Latin: *temposis* → *temporis* 'temporal'
*amase* → *amare* 'to love'
English: *is* ∿ *are*, *was* ∿ *were*, *rise* ∿ *rear*, *lose* ∿ *forlorn*
German: *verliesen* > *verlieren* 'lose' ∿ *Verlust* 'loss'
M.H.G. *kiesen-* O.E. *ceosan* - Mod G. *küren* - Mod E. *choose*
Skrt. *šašá-* 'hare' - Gmc **hazan* - O.E. *hara* - Mod. E. *hare*

is also similar to the above case. It involves a minimal change in the degree of aperture, from 1 to 2.

The model of universal phonetics discussed above is a framework of a first approximation. Many details, no doubt, have yet to be worked out and formalized. Nevertheless, I maintain that, in essence, the new model overcomes certain weaknesses of the DF theory. Whether it fails in some other areas where the DF system succeeds is yet to be seen.

---

[18] I am in debt to John McKay for this Latin example.

Rule (21) may be stated in terms of Tense/Lax:

$$\begin{bmatrix} -\text{Diffuse} \\ -\text{Compact} \\ +\text{Tense} \end{bmatrix} \rightarrow [-\text{Tense}]\ \ldots$$

But to say that the $\begin{bmatrix} \text{e} \\ \text{o} \end{bmatrix} \sim \begin{bmatrix} \varepsilon \\ \text{ɔ} \end{bmatrix}$ variation is that of Tenseness while the $\begin{bmatrix} \text{i} \\ \text{u} \end{bmatrix} \rightarrow \begin{bmatrix} \text{e} \\ \text{o} \end{bmatrix}$ variation is that of Diffuseness lacks a parallelism and makes two rules (20) and (21) seem unrelated. For a detailed discussion on this, see Chapter IX.

# V

# THE NATURE OF RULES OF SYSTEMATIC SYNTHESIS

In the preceding chapter we argued that the rows of matrices of systematic phonemics and systematic phonetics might justifiably be articulatory categories. According to the block diagram (Figure 1) on page 16 these categories are to function as input to an acoustic speech synthesizer in which the categories will be converted into acoustic signals by the rules of systematic synthesis. We ask now what is the nature and form of these rules that assign acoustic values to articulatory categories.

This question, however, may be preceded by another question, namely: must the input to the synthesizer be articulatory categories? The answer is no. The input could well be phonemes (whatever these may be), acoustic cues, or the units of motor commands. We have seen, however, that these are ample reasons to regard phonemes as complexes of divisible phonetic categories, and it would be undesirable and uneconomical to go back to phonemes when subdivided categories are already made to be available for economy of rules, and when in fact this economy was precisely one motivation to regard phonemes as divisible entities. Furthermore, due to the untenability of the "biuniqueness principle," it is impossible to go back to the original matrices of systematic phonemics from matrices of systematic phonetics.

To take a concrete example, in Holmes, Mattingly, and Shearme (1964) where a phonemic transcription of an utterance is used as input to an electronic analogue synthesizer for synthesizing speech by rules, each phoneme is given a rank. "This rank is high if the transitions of the corresponding phoneme are characteristic of the phoneme itself; it is low if the transitions depend upon the character of the adjacent phonemes" (p. 1321). The following rank is given for each phoneme of English:

| | |
|---|---|
| /p, t, k, tʃ/ | 23 |
| /b, d, g, dʒ/ | 26 |
| /m, n, ŋ/ | 15 |
| /f, θ, s, ʃ/ | 18 |
| /v, ð, z, ʒ/ | 20 |
| /l, w, r, j/ | 10 |
| All vowels | 2 |

It is obvious that it is undesirable to go back to phonemes and assign a rank to each phoneme, thereby making the number of rules approximate the number of phonemes, when categories are available for a more economical way of assigning rank values.

Similarly, the work of the Haskins Laboratories has familiarized us with the following figure:

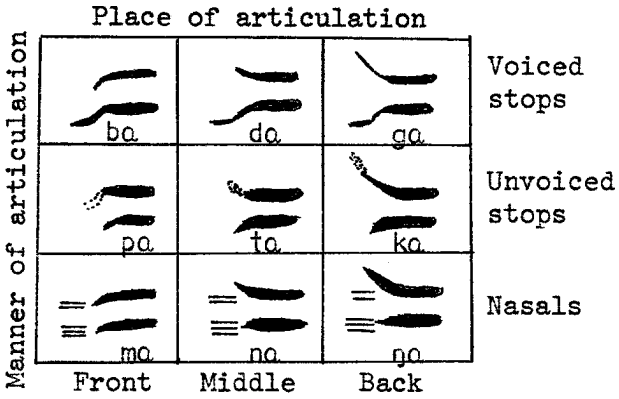



Figure 9. Spectrographic patterns illustrating the transition cues for initial stops and nasal consonants (from Liberman 1957, p. 120)

This figure, which illustrates the pattern of the acoustic cues for stop and nasal consonants of English, also clearly indicates the economy of synthesis by subphonemic rules, as all sounds having the same place of articulation have the same F2 locus, all sounds having the same manner of articulation have the same F1 locus, all nasals have the same nasal formants, etc.

The input to the synthesizer may be acoustic cues based on the spectral data. In fact, most of the speech synthesizers that have been developed so far for experimental use operate on the basis of information about the acoustic spectrum. That is, "the signals that control the synthesizer can be in the form of a spectrographic pattern or parameters derived from it." (Cooper et al. 1962, p. 6) Since sound is an acoustic phenomenon, this is most logical, and involves no transformation of levels when the synthesizer is an acoustic one. In our case, the problem is how to arrive at spectral data given matrices of articulatory phonetic categories.

One may be tempted to undertake synthesis by rules that operate directly in terms of motor commands, if one accepts the hypothesis that the signals at the level of motor commands provide a simple and direct representation of phonemes or some other linguistic units. In this case, one will not need the intermediate stage of articulatory phonetic categories before one arrives at the acoustic output. According to Cooper et al. (1962):

> The rules for converting linguistic units (phonemes) into machine control signals (motor commands) would be simple indeed; in fact, they might amount to no more than a look-up operation in a table that would contain about as few entries as there are phonemes. (p. 7)

The assumption that a great economy of description is to be attained

if rules are written in terms of motor commands that activate the articulators is interesting and ambitious. But it appears that any hope of exploring it further has to be abandoned at the moment, since, besides the fact that the use of electromyography in the study of speech is still at its beginning, there exists at the present time no synthesizer that will accept motor commands as input and thereby permit a rigorous test of a phonology in these terms.

Lastly, the conversion by rule from phonetic categories into acoustic waveform may proceed by way of articulatory configurations. That is, it is possible to control the synthesizer in terms of the changing shape of an equivalent vocal tract. The rules for synthesis in this case will convert articulatory specifications, such as our phonetic categories, into electrical signals that control the shape of a dynamic vocal tract analog such as DAVO.[19] Ideally, a dynamic vocal tract analog synthesizer should generate speech waves which are subject to inherent physiological constraints like those that limit the possible range of speech sounds produced by the human vocal mechanism. In the existing devices, however, this idealism is only partially realized. The difficulty is how to determine the shape exactly at every part of the tract and at each successive instant of time from phonetic categories that only grossly specify the articulatory configurations of the tract. Information such as cross-sectional areas of a particular segment, the dynamics of the tongue, etc. would be needed. A source of such information is X-ray investigation of the tract, but the technique is still cumbersome and not precise (cf. Ladefoged and Kim 1965; Vanderslice 1966).

An approach toward the formal interpolation of the tract configurations was made by Stevens and House (1955), and toward that of the dynamics of the tongue in time-sequence in coarticulation by Ohman (1966) and Linblom (1964). Accomplishments such as these and the building of a truly dynamic vocal tract analog synthesizer will no doubt yield exciting consequences both theoretical and practical. But until then, we will consider a device in which the rules of synthesis convert articulatory categories into signals that control an acoustic speech synthesizer as necessary for our purposes.

This device is actually quite harmonious with the nature of the speech process, which assumes that the units of an intended message are transformed into a set of motor commands which are then encoded into the changing configurations of the tract, and these in turn are further encoded to yield the acoustic waves (cf. Fromkin 1966). Since our phonetic categories are articulatory, our rules of synthesis would be equivalent to the last transformation, i.e., encoding of the tract configurations into acoustic waves. To continue to use the block diagram, the process of speech may be pictured as:

---

[19]An abbreviation of Dynamic Analog of the Vocal Organs. This synthesizer was built at MIT (Rosen 1958). It is a geometrical approximation to the shape of the human vocal tract realized in terms of an electrical transmission line. For experimental synthesis with DAVO, see Hecker (1962). Similar work at Bell Laboratories has been reported by Kelly and Gerstman (1961).

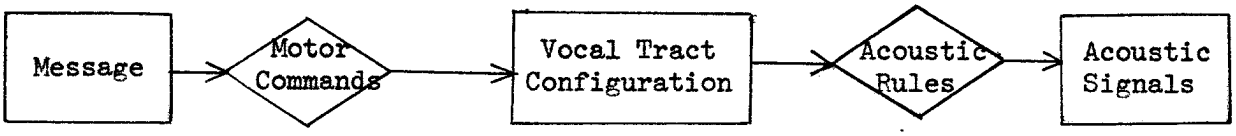



Figure 10. Schematic diagram of the encoding process of speech

"Acoustic rules" would be equivalent to our rules of synthesis, and it is the nature of these rules that we are concerned with in this chapter. Restating the question then: What is the form of the rules of systematic synthesis? Or how is the transformation mechanism from articulation to acoustic best stated?

At the present time, there are two basically different approaches. One is essentially a matter of compiling speech from a dictionary of recordings or a look-up table of values, and the other is a process of synthesizing speech entirely by rules without a table look-up procedure. In the latter case, a set of rules generates acoustic values of phonetic categories, and this set of rules may be regarded as a computer program instructing the synthesizer what parameters to operate, to what degree, for how long, etc., when given phonetic categories. In the first case, however, the pre-recorded fixed values are stored in a look-up table or in a computer memory, and values are drawn from this table or memory by a simple substitution procedure, not by any generative mechanism. In this case, synthesis "by-rule" is meant merely that transitional phenomenon between sequences of phonemes would be calculated by rules (or again by a computer program) from the information given for the relevant phonemes in the table. For example, in Holmes et al. (1964), twenty-seven items of information for every phoneme are entered in the table and a computer program calculates the appropriate transition values for each 10 msec unit of time between two phonemes by taking into consideration items such as rank, the standard duration of elements, the values of F1 - F3, the duration of the external transition of F1 - F3, the duration of the internal transitions of F1 - F3, etc.

This view of "rules of synthesis" is actually dominant in the literature. For example, when Liberman et al. (1959) illustrate the synthesis of /læbz/ by rule by saying:

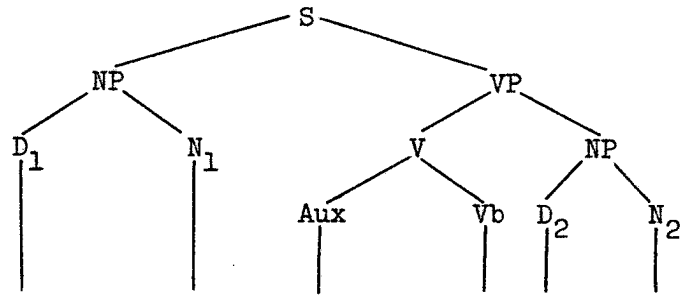
> The place rule for /l/ specifies locus frequencies at 360, 1260, and 2880 cps. . . . the place rule for /æ/ fixes formant frequencies at 750, 1650, and 2460 cps . . . . (p. 1497)

the procedure is not that of rule-generation, but that of substitution of pre-given values. Ladefoged, in a personal communication (a first draft of his forthcoming monograph, *Linguistic Phonetics*) also stated:

> Our description of a language would then include a table of values specifying ideal forms, and values accounting for the partial overlap or way of getting from one sound to another. The table of values might be expressed in

terms of numbers representing relative values of parameters for synthesizing speech, such as formant frequencies and durations.

This substitution process is analogous to the lexical look-up procedure in the syntactic component. We will carry the analogy further. In syntactic derivation, when a pre-terminal string of the following kind has been arrived at:




one goes to the lexicon (or dictionary) to find a word whose specification of grammatical category is not distinct from a node in the tree, and substitutes this item for the node in the pre-terminal string. Suppose one finds *the* for $D_1$, *girl* for $N_1$, *may* for Aux, *attend* for Vb, *the* for $D_2$, and *party* for $N_2$. Then one gets the sentence: *the girl may attend the party*. Actually, grammatical derivation is not as simple as this, but it is sufficient for our present discussion. Suppose now that one chooses *sky* for $N_2$, then the sentence will read, *the girl may attend the sky*. This is ungrammatical, the reason being that *attend* requires an Event Noun (e.g., *ceremony, festival, funeral, war, rodeo, gala*, etc.) as a direct object, and *sky* is not an Event Noun. Thus, as in phonology, the environment restricts the domain of possible items that may occur in that environment.

In the example given above, it appears that the selection of Verb governs the selection of Noun, but in the most recent treatment of this problem of lexical substitution (Chomsky 1965), Nouns are selected first, and Verb is inserted only if the selectional restriction of the Verb matches (or is not distinct from) features of Nouns that have already been selected. For example, *attend* (in the sense 'be present at') will be specified in the lexicon as [+V, +[____Event N], ...], and the insertion of *attend* will be made only when the selectional feature of the following Noun is [+Event].

Let us see now how the analogy fits in the case of systematic synthesis. In a synthesizer whose input elements are phonemes or very fine allophonic trasncriptions, one might think the analogous procedure feasible. To take an example, let us say that there are (among others) the following /t/-allophonic rules:

(22) /t/ → [ɾ]/ V́ ___ V

(23) /t/ → [tˡ] / ___/l/

(24) /t/ → [ʈ] / ___/r/

Implicit in the above rules is an assumption that /V, l, r/ are governing factors of /t/-allophones. In other words, a given value of /t/ can occur only when it conforms with the selectional restriction of the environment. That is, V, /l/, /r/ are substituted first, and values of /t/ depend on the already chosen phones /V, l, r/, just as in strict subcategorization rules in syntax, the insertion of appropriate verbs depends on the categorial configuration of other categories. Thus, we may specify in the "look-up table" different values of /t/ with their respective selectional restrictions, e.g.,

(25) [ɾ] = +/t/, +[V́ ___ V] . . .

(26) [t] = +/t/, +[___ l] . . .

(27) [ʈ] = +/t/, +[___ r] . . .

Thus, a complete inventory of possible English sounds with the specification of phonemic membership, selectional restrictions, etc. for each sound will make the substitution procedure feasible in the systematic synthesis. Of course, the inventory will be very large, but so is the dictionary of English lexical items. What is difficult in this procedure is not the size of the inventory but an explicit formulation of ranking of phonemes according to their influential ability. This formulation must be non-*ad hoc*, and supported with ample empirical evidence. In the above, we deliberately chose examples where the governing phonemes of the /t/-allophones were all [+Vocalic]. Can this fact be claimed to be universal, or is there any other such consistent ranking among phonemes or phonetic categories, as it seems to be the case in the lexical substitution, i.e., Noun is always a governing factor in the selection of Verb? Only when it is, "the look-up process" approaches feasibility.

But this assumption does not seem to hold at all. Consider another set of allophonic rules of the following:

(28) V → Ṽ / ___ /nC/

(29) /r/ → [ɽ] / /t/ ___

(30) /r/ → [ɾ] / #/θ/ ___

(31) /l/ → [ɬ] / ___(C)#

Contrary to the cases given earlier, these examples show vocalic phones being influenced by consonantal phones. If one follows the earlier assumption, one would substitute, in /trɪ̄/ *try*, [ʈ] on the basis of / __/r/. Then, the relevant environment for substituting [ɽ] is no

longer available, and one cannot apply rule (29). One may of course modify rule (29) as

(29') /r/ → [ɽ] / [t̠]___

But as soon as one does that one loses the generality of the rule, since the devoicing of /r/ depends on the voicelessness of /t/ not on the retroflexion of /t/. Furthermore, one now needs following separate rules, in addition to rule (29'):

(32) /l/ → [l̥] / [tˡ] ___

(33) /w/ → [w̥] / [tʷ] ___

(34) /y/ → [y̥] / [tʸ] ___

instead of one general rule

(35) Sonantal → [-Voiced]/ /t/___

We see thus that the procedure of substituting allophonic values according to the environment gives rise to an internal inconsistency in terms of the ranking of determining factors, and the rules lose generality if the inconsistency is made consistent.

Another aspect of the substitution procedure that must be considered is the significance of the "fixed" values in the look-up table. For example, in Holmes et al. (1964), the following values are entered for /a/:

Rank: 2
Standard duration: 15 (x 10 msec)
Duration in unstressed position: 15
F1: 790
F2: 880
F3: 2500
Duration of external transition: 4
Duration of internal transition: 4
Proportion of the steady-state value of the adjacent element which is added to the fixed contribution to derive the boundary value for F1 and F2: .5
Proportion of the steady-state value of the adjacent element which is added to the fixed contribution to derive the boundary value for F3: .5
Fixed contribution to the boundary value for F1: 410
" for F2: 470
" for F3: 1220
A1: 50.75 (db)
A2: 49
A3: 29.75
$A_{HF}$: 22.75

Fixed contribution to the boundary value for A1: 24.5
" for A2: 24.5
" for A : 10.5

Proportion of the steady-state value of the adjacent element which is added to the fixed contribution to derive the boundary value for A: .5

Duration of the external transition of A: 4

Duration of the internal transition of A: 4

We ask: what is the significance of these fixed values, when it is known that no linguistic values are absolute but relative, and that values fluctuate from time to time, from speaker to speaker, etc.? The real issue does not lie in the intra-idiosyncratic variations but in such significant differences as the different formant values of men, women and children, and the differences of durations dependent on the intended rate of speech, etc. We maintain that the rules of systematic synthesis should be capable of handling these variations, since, after all, the extension of the scope of phonetic specification to systematic synthesis was motivated precisely by these considerations. Quoting again the relevant sentence from Ladefoged (forthcoming):

> Thirdly, it [=a theory of phonetics] must lead to the specification of actual utterances by individual speakers of each language: this is physical phonetics. (cf. p. 8)

A look-up table procedure where pre-fixed absolute values are entered is inherently incapable of handling this.

In the light of this, we propose two measures to be used in systematic synthesis:

(1) the use of generative rules to assign acoustic values to phonetic categories, instead of using a look-up table,

(2) the use of the notion "degree" to specify relative values, instead of giving absolute numerical values.

We believe that the use of these two measures enables us to overcome the shortcomings of the substitution procedure via a look-up table, and to handle the relative nature of phonetic values in a more natural way. Examples showing how this claim should be true are given in Chapters VI, VIII, and IX, where detailed rules of assignment of values of formant frequency, amplitude, and duration, respectively, are formulated.

# VI

# FORMANT FREQUENCY ASSIGNMENT RULES

We will consider here how the formant frequencies of English vowel phonemes may be generated from phonetic categories given at the level of systematic phonetics, and how the generated values are said to be flexible and relative, not absolute as is the case if they were given in a look-up table.

As was mentioned, we assume that acoustic values are predictable from physiological categories, if not vice versa, and further that categories indicating physiological equidistance in general yield an acoustic equidistance also. The first assumption enables us to assign a uniform value to a given category, no matter what the other phonetic categories may be with which it is combined to form a phone. For example, we assume that if F1 of a Mid vowel is 500 cps, it is so whether the given Mid vowel is Front, Central, or Back; whether it is an oral or nasal vowel, etc. The second assumption is probably not true in a strict sense, and we will have to make some adjustments. But we assume that the principle is in general valid. For example, if it is assumed that High, Mid, and Low vowels are articulatorily equidistant from High to Low, then we will also assume that, if F1 of High is 300 cps and F1 of Low is 700 cps, F1 of Mid is 500 cps, a midpoint between the two. This phenomenon also enables us to use the notion "degree" in rule-formulation. Continuing and elaborating the above example, if Tense-High is 300 cps, Lax-High 400 cps, Tense-Mid 500 cps, Lax-Mid 600 cps, etc., then we can say that the distinction between Tense and Lax differs by 1 degree, and High, Mid, Low differ in 2 degrees, where 1 degree equals 100 cps. Flexibility of values is also seen in this example. That is, if we change the value of 1 degree, say, to 90 cps or 110 cps, we will have a formant chart different from, but still parallel to, the former one.

Keeping these considerations in mind, let us now look at various published data on formants of English vowel phonemes. We will not be concerned here with the second elements of diphthongs but only with steady-state values of monophthongs and initial elements of diphthongs. We will assume the following pattern of vowel phonemes of General American English, and, for the convenience of the reader, the traditional terms will be used:

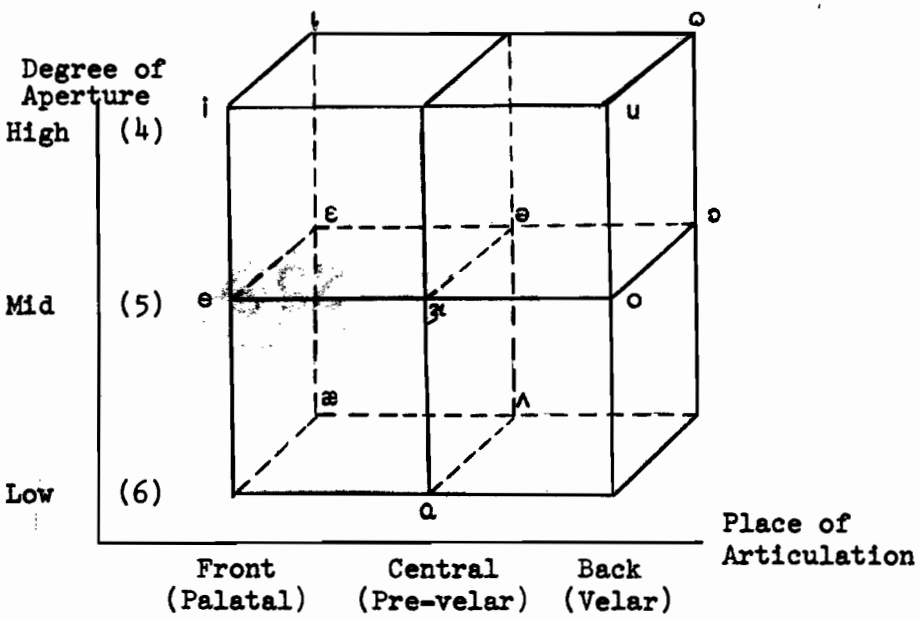



Figure 11. A schematic, three-dimensional quadrangle modelling vowel phonemes of General American English

Sample words of the form /h___d/ are:

heed hid who'd hood

heyed head heard hoed hawed

had hud hard

Table 8. Formant one frequency in English vowels

| Source | i | ɪ | e | ɛ | æ | ɑ | ʌ | ɜ˞ | ɔ | o | ʊ | u | ɐ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| (1) Peterson and Barney (1952) | 270 | 390 | | 530 | 660 | 730 | 640 | 490 | 570 | | 440 | 300 | |
| (2) Fairbanks and Grubb (1961) | 263 | 387 | | 493 | 733 | 775 | 588 | | 600 | | 392 | 279 | |
| (3) Householder (1956) | 300 | 400 | 500 | 600 | 800 | 750 | 600 | 520 | 625 | 475 | 380 | 250 | |
| (4) Lehiste (1964) | 276 | 406 | 506 | 606 | 718 | 725 | 606 | 500 | 612 | 587 | 581 | 293 | |
| (5) Peterson (1961) | 255 | 355 | | 560 | 750 | 750 | 650 | 450 | 625 | | 475 | 300 | |
| (6) Holbrook and Fairbanks (1962) | 272 | 422 | 520 | 592 | 752 | 630 | 475 | 630 | 535 | 465 | 342 | | |
| (7) Holmes et al. (1966) | 250 | 400 | 640 | 640 | 790 | 790 | 700 | 580 | 490 | 490 | 370 | 250 | 610 |

Continued on following page

Table 8 continued

| Source | i | ɪ | e | ɛ | æ | ɑ | ʌ | ɝ | ɔ | o | ʊ | u | ɒ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| (8) Arnold et al. (1958) | 315 | 380 | | 580 | 940 | 720 | 960 | 550 | 620 | | 500 | 250 | 720 |
| (9) Lehiste and Peterson (1961) | 315 | 415 | 360 | 570 | 640 | 645 | 610 | 475 | 505 | 495 | 450 | 355 | |
| (10) Lehiste and Peterson (1961) | 320 | 410 | 335 | 540 | 625 | 665 | 585 | 430 | 590 | 435 | 400 | 350 | |
| Average | 281 | 398 | 444 | 552 | 690 | 724 | 644 | 477 | 595 | 505 | 448 | 309 | |

Table 9. Formant Two Frequency in English Vowels

| | i | ɪ | e | ɛ | æ | ɑ | ʌ | ɝ | ɔ | o | ʊ | u | ɒ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| (1) | 2290 | 1990 | | 1840 | 1720 | 1090 | 1190 | 1350 | 840 | | 1020 | 870 | |
| (2) | 2378 | 2038 | | 1660 | 1654 | 1064 | 1199 | | 846 | | 1122 | 825 | |
| (3) | 2250 | 1850 | 1900 | 1800 | 1690 | 1125 | 1150 | 1380 | 870 | 870 | 1000 | 880 | |
| (4) | 2136 | 1943 | 1981 | 1818 | 1612 | 1300 | 1225 | 1337 | 850 | 925 | 993 | 787 | |
| (5) | 2300 | 2000 | | 1800 | 1600 | 1050 | 1250 | 1350 | 850 | | 1000 | 900 | |
| (6) | 2312 | 2025 | 2078 | 1925 | 1955 | 1245 | 1322 | 1310 | 902 | 848 | 1132 | 940 | |
| (7) | 2320 | 2080 | 1600 | 2020 | 1780 | 880 | 1360 | 1420 | 820 | 1480 | 1000 | 880 | 880 |
| (8) | 2200 | 2200 | | 2100 | 1750 | 1150 | 1750 | 1900 | 920 | | 1150 | 800 | 980 |
| (9) | 2200 | 1750 | 2015 | 1610 | 1570 | 1100 | 1185 | 1245 | 880 | 960 | 980 | 895 | |
| (10) | 2205 | 1755 | 2105 | 1705 | 1740 | 1145 | 1155 | 1255 | 985 | 905 | 1015 | 845 | |
| A | 2250 | 1922 | 2015 | 2770 | 1660 | 1103 | 1209 | 1318 | 878 | 901 | 1032 | 868 | |

Table 10. Formant three frequency in English vowels

| | i | ɪ | e | ɛ | æ | ɑ | ʌ | ɝ | ɔ | o | ʊ | u | ɒ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| (1) | 3010 | 2550 | | 2480 | 2410 | 2440 | 2390 | 1690 | 2410 | | 2240 | 2240 | |
| (2) | 3099 | 2591 | | 2444 | 2510 | 2614 | 2623 | | 2636 | | 2500 | 2496 | |
| (3) | 2750 | 2625 | 2500 | 2450 | 2540 | 2500 | 2500 | 1660 | 2600 | | 2450 | | |
| (4) | 2856 | 2544 | 2581 | 2587 | 2486 | 2637 | 2656 | 1644 | 2600 | 2737 | 2506 | 2152 | |
| (5) | 3000 | 2600 | | 2500 | 2500 | 2600 | 2600 | 2700 | 2600 | | 2500 | 2200 | |
| (6) | 2940 | 2710 | 2660 | 2610 | 2615 | 2500 | 2600 | 1650 | 2580 | 2435 | 2385 | 2302 | |
| (7) | 3220 | 2560 | 2500 | 2500 | 2500 | 2500 | 2500 | 2500 | 2500 | 2500 | 2500 | 2200 | 2500 |
| (8) | 3300 | 2700 | | 3000 | 2700 | 2450 | 2450 | 2500 | | | | | 2400 |
| (9) | 2700 | 2470 | 2510 | 2465 | 2460 | 2540 | 2565 | 1680 | 2525 | 2495 | 2360 | 2240 | |
| (10) | 2800 | 2415 | 2630 | 2415 | 2415 | 2520 | 2255 | 1575 | 2365 | 2435 | 2090 | 2105 | |
| A | 2894 | 2563 | 2576 | 2494 | 2492 | 2544 | 2523 | 1657 | 2539 | 2525 | 2379 | 2248 | |

Notes on the sources of data for Tables 8, 9 and 10:

(1) p. 183, Table II. Averages of 33 adult male speakers, uttering two tokens of each V in /h___d/.

(2) p. 210. Entered values are those of "preferred."

(3) p. 237. Mostly in env. /b___t/. Number of samples unknown.

(4) p. 25, Table VIII. Averages of 4 utterances of each V by GEP. /ʌ/ = /ə/.

(5) pp. 17-19, Tables 1-3. Averages of 4 male speakers, 2 utterances of each V by each speaker.

(6) p. 45, Table 2 (diphthongs) and p. 52, Table IV (monophthongs). Twenty male GAE speakers, in env. /h___/ for diphthongs, /h___d/ for monophthongs. Each subject uttered each vowel 3 times. Values of diphthongs are those of second (out of 5) sampling points.

(7) pp. 141-43. No information available on the source of data.

(8) p. 124, Table 6. Measurements of a single speaker of British English.

(9) p. 269, Table I. Averages of 1263 CVC words spoken by GEP.

(10) p. 269, Table I. Averages of 350 CVC words spoken by 5 speakers.

The average excludes rows (7) and (8), which are values for British English.

It may be seen that there is considerable variation in the formant frequencies reported. Even though the average has not much significance in this kind of case, we will attempt to formulate rules in such a way that they generate values close to the average values and, by a slight modification, can approximate any set of values.

We note now that, in the average values of F1, the difference between:

/i/ and /ɪ/ is 117 cps
/e/ and /ɛ/ is 108 cps
/ɑ/ and /ʌ/ is 80 cps
/o/ and /ɔ/ is 90 cps
/u/ and /ʊ/ is 139 cps
Average difference: 107 cps

and that the difference between:

/i/ and /e/ is 163 cps
/ɪ/ and /ɛ/ is 154 cps
/e/ and /æ/ is 246 cps
/u/ and /o/ is 194 cps
/ʊ/ and /ɔ/ is 147 cps
/ɑ/ and /ɝ/ is 247 cps
Average difference: 192 cps

That is, on the average, the difference in F1 between a Tense vowel and the corresponding Lax vowel is 107 cps, and the vowels differing in one degree of Aperture (i.e., High - Mid, Mid - Low) differ, on the average, by 192 cps. For simplicity, let us say that one degree equals 100 cps, and that Tense-Lax differ by one degree, and High and Mid, and Mid and Low differ by two degrees. That this is a reasonable assumption is readily seen by looking at the data of row (3) where F1 of /i/ = 300, /ɪ/ = 400, /e/ = 500, /ɛ/ = 600, /æ/ = 800, etc.

Our rules will be then a matter of assigning the "degrees" properly to a certain abstract formant value that every vowel is assumed to have at the beginning by virtue of being a vowel. Still there are many starting points from which the rule-formulation may proceed. For instance, one may start formulating rules, regarding the lowest formant as being a basic value. In this case, rules will be mainly a matter of operations of additions. Conversely, one may regard the highest formant value as the basic value; in this case, the rules will be so formulated as to deduct values from the given highest value. Or one may pick the midpoint as the starting value, and formulate rules of both addition and deduction.

We feel that the last starting point is intuitively more sensible and proper than the others in that the midpoint implies the average and the most neutral value. In vowels, the mid-values of F1, F2, F3 should approximate those of a neutral or mid-central vowel /ə/, and it is reasonable to assume that other vowels are deviations (in tract shape as well as in formant values) from this mid-central, the most neutral vowel. Let us define then the term "center frequencies" as the formant values of the abstract, neutral vowel. We believe that the center frequencies differ from person to person, however small the difference may be. But for our purpose, we will set the values of center frequencies at F1 = 550 cps, F2 - 1400 cps, and F3 = 2550 cps. (Cf. Holmes et al.'s /ɜ/, F1 = 580, F2 = 1420, F3 = 2500; Householder's /ɜ/, F1 = 520, F2 = 1380 cps. See also *Preliminaries*, p. 18, Section 2.13, Neutral position of the vocal tract.)

Formant assignment rules are then formulated on the basis of two given values: (1) values of center frequencies, (2) value of a degree. The rules for F1 are:

(36) F1 = 550 cps +

(a) if High, -2d, (where d stands for degree, and 1 degree = 100 cps.)
(b) if Low, +2d
(c) if αTense, α.5d / Low
(d) if Tense, -α.5d / High or Mid
(e) if Low-Central, -.5d

These rules should be read: "If the vowel is High, or if the matrix has the phonetic category High, decrease the given value by two degrees, etc." The use of "alpha"-variable notation is in accordance with the convention that is familiarly used in P rules in DF terms. It indicates the agreement in polar values between two or more variables. For example, rule (c) is to be interpreted as: if +Tense, +.5d; if -Tense, -.5d. Consequently, if a variable can assume more than two values, as we argued in Chapter IV, the alpha-notation is not usable. We will consider this aspect of variable notation in more detail in the next chapter. Notations such as / Low, in rule (c) should be read "if the vowel is further Low," etc. The rules for F2 are:

(37) F2 = 1400 cps +

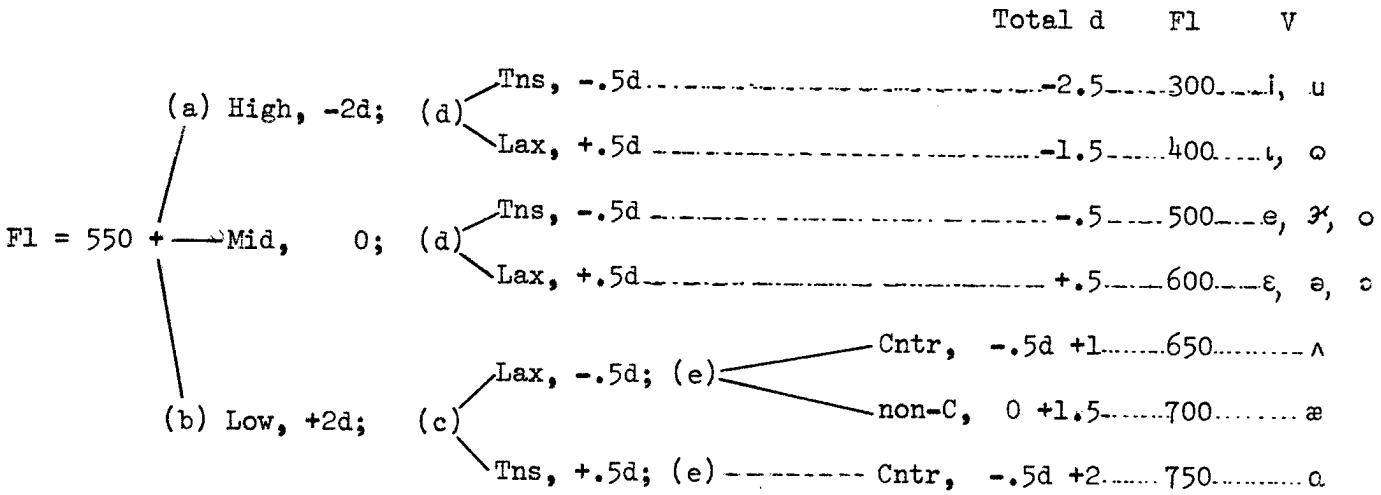
(a) if Front, +5d
(b) if Back, -5d
(c) if High, +1d
(d) if Low, -1d
(e) if αTense, α1d / Front
(f) if αTense, -α1d / Central or Back
(g) if Low-Central, -1d
(h) if High-Front-Tense, +1d

The rules for F3 are:

(38) F3 = 2550 +

(a) if High, +1d / Front
(b) if High, -1d / Back
(c) if Low, +1d / Back
(d) if Low, -1d / Front
(e) if High-Tense, +1d / Front
(f) if Low-Tense, -1d / Back
(g) if Mid-Central, -9d

As an illustration, we give below one derivation, that of F1:




In the following are comparisons of the values generated by rules with a few sets of data. At the top of the page is the set of values generated by rules, the next set is the average values from pp. 69-71. The third set is that of row (3); the fourth, (6), the fifth, (9). The values that deviate from the generated ones by more than 50 cps are underlined once, and those that deviate by more than 100 cps are underlined twice.

Table 11. A comparison of the generated formant frequency values with a few sets of data from Tables 8-10.

A. Generated

| | i | ɪ | e | ɛ | æ | ɑ | ʌ | ɝ | ɔ | o | ʊ | u |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| F1 | 300 | 400 | 500 | 600 | 700 | 750 | 650 | 500 | 600 | 500 | 400 | 300 |
| F2 | 2200 | 1900 | 2000 | 1800 | 1700 | 1100 | 1200 | 1300 | 1000 | 800 | 1100 | 900 |
| F3 | 2750 | 2650 | 2550 | 2550 | 2450 | 2550 | 2550 | 1650 | 2450 | 2450 | 2350 | 2250 |

B. Average of data in Tables 8-10

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| F1 | 281 | 398 | 444 | 552 | 690 | 724 | 644 | 477 | 595 | 505 | 448 | 309 |
| F2 | 2250 | 1922 | 2015 | 1770 | 1660 | 1103 | 1209 | 1318 | 878 | 901 | 1032 | 868 |
| F3 | 2894 | 2563 | 2576 | 2494 | 2492 | 2544 | 2523 | 1657 | 2539 | 2525 | 2379 | 2248 |

C. Data (3)

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| F1 | 300 | 400 | 500 | 600 | 800 | 750 | 600 | 520 | 625 | 475 | 380 | 250 |
| F2 | 2250 | 1850 | 1900 | 1800 | 1690 | 1125 | 1150 | 1380 | 870 | 870 | 1000 | 880 |
| F3 | 2750 | 2625 | 2500 | 2450 | 2540 | 2500 | 2500 | 1660 | 2600 | | 2450 | |

D. Data (6)

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| F1 | 272 | 422 | 520 | 520 | 592 | 752 | 630 | 475 | 630 | 535 | 465 | 342 |
| F2 | 2312 | 2025 | 2078 | 1925 | 1955 | 1245 | 1322 | 1310 | 902 | 848 | 1132 | 940 |
| F3 | 2940 | 2710 | 2660 | 2610 | 2615 | 2500 | 2600 | 1650 | 2525 | 2435 | 2385 | 2302 |

E. Data (9)

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| F1 | 315 | 410 | 360 | 570 | 640 | 645 | 610 | 475 | 505 | 495 | 450 | 355 |
| F2 | 2200 | 1750 | 2015 | 1610 | 1570 | 1100 | 1185 | 1245 | 880 | 960 | 980 | 895 |
| F3 | 2700 | 2470 | 2510 | 2465 | 2460 | 2540 | 2565 | 1680 | 2525 | 2495 | 2360 | 2240 |

The comparison shows that the generated values rather closely approximate those of the "Average" and of data (3), and somewhat loosely, those of data (6) and (9).

A few remarks are in order.

(1) It should be noted that, in F1 rules (c) and (d), the category Tense functions differently. That is, it adds .5d if the vowel is Low, but deducts .5d if the vowel is non-Low. Seemingly inconsistent rules of this kind are also found in F2 (e) and (f), where Tense adds 1d if the vowel is Front, but deducts 1d if the vowel is non-Front. This is however not an arbitrary measure. It shows the "centrifugal" character of the Tensity feature so that "tense phonemes are produced with more deviation from the neutral, central position than the corresponding lax phonemes." (Jakobson and Halle 1964, p. 60) This phenomenon gives another justification for setting the neutral center frequencies as the starting point. If the starting point had been somewhere else, say, a corner of a vowel triangle, the "centrifugal" character of the tensity feature would not have been expressed with any kind of generality.

(2) It should be clear now that one can modify the desired values by using any or all of the following measures:

(a) change the values of center frequencies
(b) change the value of a degree
(c) add, delete, or modify rule(s)

(a) enables us to move our vowel quadrangle or trapezoid along the frequency scale without altering the shape of the trapezoid. It is well known that children have generally higher formants than women, who in turn have generally higher formants than men (cf. Potter and Steinberg 1950; Peterson 1961). This variation is easily adaptable in our model of synthesis by using method (a). (b) affects the width of the vowel trapezoid. For example, for one whose F1 values range widely from 250 cps to 800 cps, the value of a degree will be greater than the one whose F1 values range narrowly, say, from 350 cps to 700 cps. Note that in this case all the rules are unaffected and the same. Still there may be cases where the differences in the vowel quadrangle are not all that symmetrical; for example, the /i/ corner may be jutting out more conspicuously than other trapezoids, etc. This, we believe, can be adjusted by measure (c), i.e., by a slight modification of rule(s). For example, we note that in data (6) F2 of Front and Low-Central vowels are very much higher than the generated values, and hence there are many underlines. But this will be amended if we delete F2 rule (37.g), and add instead a rule

(37.g') if Front, +1d

Similarly, data (9) shows very low higher formant values in Lax vowels. Again, many underlines will be eliminated by adding the rules

(37.a') if Lax, -1.5d

(38.h) if Front-Lax, -1d

These considerations show the flexibility of our model of systematic synthesis. That is, our model reflects the relative nature of phonetic values in a proper way. What Joos (1948) remarkably early noted and termed as "inconsequential personal peculiarities" (p. 60, p. 86), which he pictured as follows, can be adapted by our rule-schema, but not by a look-up table procedure in which pre-stored absolute values are given and therefore fail to model the following picture.

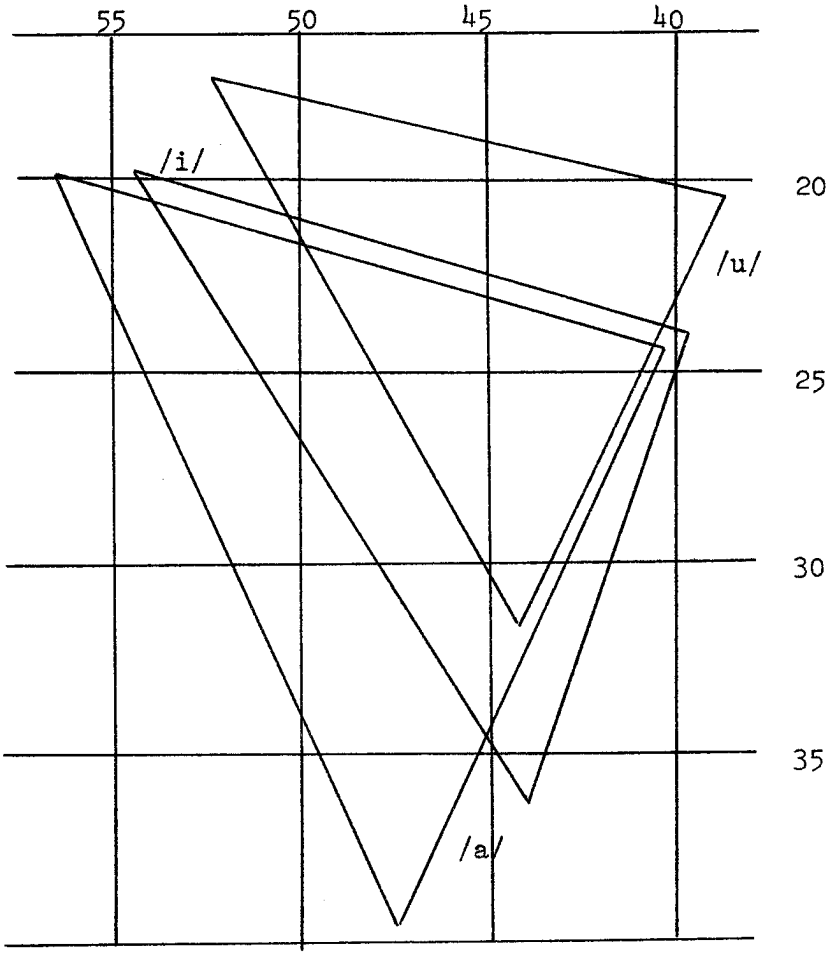



Figure 12. Vowel triangles of three different speakers illustrating Inconsequential Personal Peculiarity (from Joos 1948, p. 60, Fig. 30)

(3) The importance of ordering in phonological rules has been emphasized several times in the literature (Halle 1961; Chomsky and Halle 1965), and we are convinced that the claim is valid. But it is to be noted that in rules of systematic synthesis, ordering plays no role. All the formant assignment rules can be applied in any order. It will be seen later that the same is also true with amplitude assignment rules and duration assignment rules. This leads us to assume that internal unordering is a property of the rules of systematic synthesis.

(4) In the comparative chart (Table 11), we note that the phoneme whose formant values deviate most from the generated ones is /ɔ/. That is, while the rule predicts 1000 cps for F2, most attested values range from 850 to 900 cps. And we further note that the generated value of F2 of /o/, 800 cps, tends to be considerably lower than the actual data. The rules were constructed to predict these values since in non-Front vowels F2 of Lax vowels has in general a higher value than in the corresponding Tense vowel (cf. F2 values of /u/ and /ʊ/, /a/ and /ʌ/). But maintaining this generalization creates a problem in the values of /o/ and /ɔ/. Of course, we can solve the problem by adding the following rules:

(37.i) if Back-Mid-Tense, +1d

(37.j) if Back-Mid-Lax, -1d

But this kind of repair rule approaches the so-called "one rule for one phoneme" and makes the rules lose generality. Thus the problem and the peculiarity of /o/ and /ɔ/ still remain. We will return to this problem in Chapter IX, as the peculiarity also appears in duration.

(5) We note that there are several rules of the kind

(36.a) if High, -2d

(36.b) if Low, +2d

where the symmetry in phonetic categories involved and the assigned values is obvious, but where the alpha-variable notation cannot be used nevertheless. We will see in the next chapter the significance of this type of rule and a possible device by which the symmetry can be expressed with generality.

# VII

# THE NOTION OPTIMAL OPPOSITION

In the preceding chapter on rules of systematic synthesis of formant values of General American English vowels, we have noted such pairs of rules as the following:

(36) F1 (a) if High, -2d
(b) if Low, +2d

(37) F2 (a) if Front, +5d
(b) if Back, -5d
(c) if High, +1d
(d) if Low, -1d

(38) F3 (a) if High, +1d / Front
(b) if High, -1d / Back
(c) if Low, +1d / Back
(d) if Low, -1d / Front

These pairs of rules are those in which two phonologically most distant (within vowels) phonetic categories are assigned the same amount of opposite acoustic values, and suggest themselves that the pairs may be coalesced into one rule respectively, thereby achieving a linguistically significant generalization.

That this phenomenon is a linguistically significant one is evidenced by the fact that much linguistic literature has referred to the phenomenon with such terms as "maximal contrast," "optimal opposition," "differentiation maxima," etc. (cf. de Groot 1931; Jakobson 1942; Martinet 1955).

Jakobson and Halle (1956) and Jakobson (1959) especially showed that this is the principle by which a child first discriminates speech sounds, that, in particular, a child's universal vocabulary /pa/, /ma/, etc. are phonologically explainable in terms of the notion optimal opposition.

> Ordinarily child language begins . . . with what psychopathologists have terms the 'labial stage.' In this phase speakers are capable only of one type of utterance, which is usually transcribed as /pa/. From the articulatory point of view the two constituents of this utterance represent polar configurations of the vocal tract: in /p/ the tract is closed at its very end while in /a/ it is open as widely as possible at the front and narrowed toward the back, thus assuming the horn-shape of a megaphone. This combination of two extremes is also apparent

> on the acoustic level: the labial stop presents a momentary burst of sound without any great concentration of energy in a particular frequency band, whereas in the vowel /a/ there is no strict limitation of time, and the energy is concentrated in a relatively narrow region of maximum aural sensitivity. . . . Consequently, the diffuse stop with its maximal reduction in the energy output offers the closest approach to silence, while the open vowel represents the highest energy output of which the human vocal apparatus is capable. (*Fundamentals*, p. 37)

In the present DF theory of phonology, there is a notational device that enables us to express a rather superficially similar phenomenon in a general way. It is the use of the "alpha"-variable notation, e.g.,

(39) [αGrave] → [-αGrave] / X

The conventional interpretation of the rule is that "if [+Grave], change it to [-Grave]; and if [-Grave], change it to [+Grave], in the environment X." Since each DF may have two polar values only, plus and minus, the change of the value of a feature is equivalent to the optimal and maximal change. We cannot, however, apply this notational device to our case given at the beginning of the chapter. For instance, we cannot coalesce (36.a) and (36.b) as follows by using an alpha notation:

(36') if αHigh, -α2d

The reason is that when α is minus, -High does not mean Low, which is what we want, but non-High, i.e., Mid and Low. That is, "alpha" here does not mean "optimally opposite," but "non-". Since in the DF theory, a feature can assume only two polar values, the domain of "the optimally opposite" value of [+F] and non-[+F] coincides and has only one member, i.e., [-F]. But as soon as one accepts the assumption that phonetic categories are not necessarily binary, and hence a category can assume more than two values, one has to consider which of the two interpretations "alpha" implies, and whether both of them are linguistically significant notions.

The first question is unanswerable, since the DF theory has no need to distinguish the two, and there is no way to tell which notation was implied when the alpha-variable was first introduced. The second question is an empirical one. The answer depends on whether natural languages have both phenomena which may be generalized in a significant way by using special notations. For the moment, without exploring many languages, we will assume the independent existence of both phenomena, which we will henceforth designate by the symbols $\nu$ (nu) implying "non-", and ω (omega) implying "optimally opposite," respectively. For an example of a "non-" case, take the rule for the English Indefinite Article. It is *an* before a word beginning with a true vowel, and *a* otherwise. This "otherwise" means, in DF terms, true consonants, Glides, and Liquids,

which may be awkwardly specified in DF terms as $\left\{\begin{bmatrix}-\text{Vocalic}\end{bmatrix}, \begin{bmatrix}+\text{Conson.}\\+\text{Vocalic}\end{bmatrix}\right\}$. Thus,

(39) Ind Art → *an* / ___ # $\begin{bmatrix}\begin{bmatrix}+\text{Voc}\\+\text{Cns}\end{bmatrix} \ldots\end{bmatrix}_N$

(40) Ind Art → *a* / ___ # $\begin{matrix}[[-\text{Voc}]\ \ldots]_N\\ \begin{bmatrix}\begin{bmatrix}+\text{Cns}\\+\text{Voc}\end{bmatrix}\ldots\end{bmatrix}_N\end{matrix}$

(40) is in many ways *ad hoc*. First of all, the grouping of Consonants, Glides, and Liquids could be equally expressed as

$\left\{\begin{bmatrix}+\text{Conson.}\end{bmatrix}, \begin{bmatrix}-\text{Vocalic}\\-\text{Conson.}\end{bmatrix}\right\}$ instead of $\left\{\begin{bmatrix}-\text{Vocalic}\end{bmatrix}, \begin{bmatrix}+\text{Vocalic}\\+\text{Conson.}\end{bmatrix}\right\}$

Secondly, whichever way of grouping is chosen, it implies that there is a major binary break between the upper set of segments and the lower set, when in fact the involved segments are three major phonetic categories of equal status. Thirdly, (40) fails to show that the relevant environment is the exact complement of the environment of (39), that is, any other possible value-combinations of Vocalic and Consonantal except [+Vocalic, -Consonantal], i.e., non-$\begin{bmatrix}+\text{Vocalic}\\-\text{Conson.}\end{bmatrix}$. That this relation is significant and must be shown by the form of rules is seen by comparing (40) with the following:

(41) Ind Art → *a* / ___ #$\left\{\begin{matrix}[[+\text{Nasal}]\ldots]_N\\ \begin{bmatrix}\begin{bmatrix}+\text{Strd.}\\+\text{Grave}\end{bmatrix}\ldots\end{bmatrix}_N\end{matrix}\right\}$

Like (40), the environment of (41) does not overlap with that of (39), since no vowel is [+Nasal] or [+Strident] in English, and both (40) and (41) involve the same number of symbols. This fact would make both rules as having the same rank of evaluation. But the environment of (40) is the complement of that of (39), while the environment of (41) has nothing to do with that of (39). Thus we see that the environment of (40) must be stated in such a way that it shows the complement relation that it bears to the environment of (39). This is possible by using the notation ≁:

(40') Ind Art → *a* / ___ # $\begin{bmatrix}\not\sim \begin{bmatrix}+\text{Vocalic}\\-\text{Conson.}\end{bmatrix}\ldots\end{bmatrix}_N$

where ≁ means non-$\begin{bmatrix}+\text{Voc}\\-\text{Cns}\end{bmatrix}$, i.e., any combination of values of $\begin{bmatrix}\text{Voc}\\\text{Cns}\end{bmatrix}$

except $\begin{bmatrix}+\text{Voc}\\-\text{Cns}\end{bmatrix}$. Using the notation ( ), which as in mathematics implies that the operation inside ( ) must be carried out first, we may now collapse (39) and (40') into one rule in the following way:

$$(42)\quad \text{Ind Art} \rightarrow a\ (an)\ /\ \_\_\ \#\left[\nu\left(\begin{bmatrix}+\text{Voc}\\-\text{Cns}\ \ldots\end{bmatrix}\right)\right]\ \text{N}$$

Rules of the kind of (40') are the so-called "elsewhere" rules which are numerous both in syntax and in phonology. We believe that the use of the $\nu$ (nu) notation and ( ) will bring simplicity and generality into the rules in such cases.

To take another example. It seems that Stanley's (1966) "negative condition" is motivated by the same kind of consideration. For instance, he finds it more economical to state an MS condition in a language which has consonant clusters but no geminate clusters in negative terms, i.e.,

$$(43)\quad \text{C cluster} \rightarrow\ \sim\ +\ X\ \begin{bmatrix}+\text{Cons}\\ \alpha\text{Grave}\\ \beta\text{Comp}\\ \gamma\text{Cont}\\ \delta\text{Voice}\end{bmatrix}\ \begin{bmatrix}+\text{Cons}\\ \alpha\text{Grave}\\ \beta\text{Comp}\\ \gamma\text{Cont}\\ \delta\text{Voice}\end{bmatrix}\ Y\ +$$

where X and Y may be any segment(s), and + is a word boundary. $\sim$ will be equivalent to our $\nu$. As Stanley points out, the rule stands for $2^4 = 16$ negative conditions, since there are four variables each of which can assume + or - independently of the others, and it is much more economical than a rule stating positively what possible consonant clusters are.

We see thus that the notation $\nu$ is an empirically and linguistically significant one. We will now see whether the same is true with $\omega$, the notation for "optimal opposition," not "complement." Again, we will not explore into many languages for examples.

Jakobson and Halle's example of /pa/ as being two segments that are maximally opposed may be viewed as features of /p/ having "optimally opposite" values of those of /a/:

| p | a |
|---|---|
| -Voc | +Voc |
| +Cons | -Cons |
| +Diff | -Comp |
| -Cont | +Cont |
| -Voice | +Voice |

Here, /p/ may be said to have the feature composition $\omega\begin{bmatrix}+\text{Voc}\\-\text{Cons}\\+\text{Comp}\\+\text{Cont}\\+\text{Voice}\end{bmatrix}$[20]

Schane's (1965) truncation rule in French applies only to words ending in a true consonant or a true vowel, but not to Glides and Liquids. The rule may be expressed as

(44) $\omega\begin{bmatrix}+\text{Voc}\\-\text{Cons}\end{bmatrix} \rightarrow \emptyset \;/\; \_\_\_\; X$ (where X is env. not relevant here)

The pairs of rules given at the beginning of the chapter may now be collapsed as follows by using ω notation:

(45) F1 (a) if High, -2d
(b) if Low, +2d
} if ωHigh, -ω2d

(46) F2 (a) if Front, +5d
(b) if Back, -5d
} if ωFront, ω5d
(c) if High, +1d
(d) if Low, -1d
} if ωHigh, ω1d

(47) F3 (a) if High, +1d / Front
(b) if High, -1d / Back
} if High, ω1d / ωFront
(c) if Low, +1d / Back
(d) if Low, -1d / Front
} if Low, ω1d / ωBack

There are some interesting aspects to be considered with regard to this notion of optimal opposition in phonology.

The first consideration is the fact that the notion is equivalent to the so-called "principle of maximal differentiation," or "the maximal distance in phonological space." This notion or principle explains, besides the child's first vocabulary /pa, ma/, such universal phenomena as the vowels of 3-vowel languages always being /i, a, u/, not, say /i, e, æ/ or /æ, a, o/, etc., and stops being in a higher position in the oppositional hierarchy than fricatives, as stops are more optimally opposed to vowels than fricatives are, etc.

---

[20]It is interesting to note that the opposition between /p/ and /a/ is realized in five DF's (Vocalic, Consonantal, Diffuse, Continuant, Voice) only, values of seven DF's remaining the same, i.e., [+Grave, -Nasal, -Strident, +Tense, -Check, -Sharp, -Flat]. This is another indication that features are not really independent but that there is a certain hierarchical order among DF's, such that only certain features are relevant in defining "optimal opposition" but not others. Otherwise, given [+Voc, -Cons, +Grave, +Comp, -Diff, -Nasal, +Cont, -Strd, +Tense, +Voice, -Check, -Sharp, -Flat] for /a/, the corresponding optimally opposite segment, i.e., the segment every feature of which assumes the opposite value of the corresponding feature in /a/, would be *[-Voc, +Cons, -Grave, -Comp, +Diff, +Nasal, -Cont, +Strd, -Tense, -Voice, +Check, +Sharp, +Flat]. This segment is not /p/, but a glottalized, palatalized, rounded, voiceless, lax, strident, alveolar, nasal consonant, something like [ˈn̪ʷ]!

Secondly, consider the fact that the optimal opposition of High-Front-Tense (i.e., /i/) should be expected to give, on the same High level, High-Back-Lax (i.e., /ʊ/), while factually it is /u/. That is, Tense does not function in the opposition of Front and Back. An explanation is found in acoustics. That is, since the Tensity feature has a "centrifugal character, it will make formants of Tense vowels move farther away from the neutral value. Thus, formants of Tense-Front vowels are farther from Tense-Back vowel formants than from those of Lax-Back vowels. Therefore, the former case is the case of the maximal distance or the optimal opposition. This is a case where articulation and acoustics do not correspond in one-to-one fashion, and where acoustics sheds more light on the explanation of the case than articulation.

Thirdly, the notion optimal opposition explains a well-known phenomenon that Front vowels are normally unrounded and Back vowels are usually rounded. This phenomenon is so universal and so normal that the IPA chart of cardinal vowels does not deal with the dimension of rounding separately, but the third dimension of rounding is fused into the two-dimension chart. Instead of stating the fact in an uninteresting way, we may ask why this should be so. Again, the answer comes from the notion optimal opposition in acoustic terms in the following fashion.

From acoustics, we know that rounding (or the narrowing of the orifice of a resonance) causes a lowering of the higher formants (or resonance frequencies). Notice what this means in the case of vowel formants. It means that if Front vowels are rounded and Back vowels are unrounded, the values of higher formants (particularly F2 and F3) would be closer to each other, since the high F2 and F3 of Front vowels will be lowered and the low F2 and F3 formants of Back vowels will be raised, i.e.,

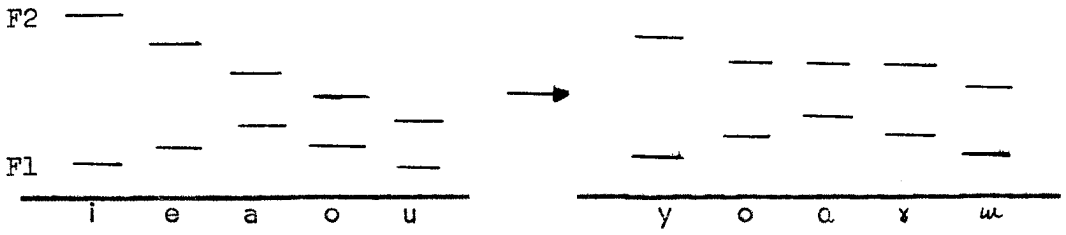



Figure 13. A schematic diagram showing the change in formant position when the feature Rounding is reversed, thereby violating the principle of maximal distance.

This is against the principle of optimal opposition or maximal differentiation.

This observation gives us an important insight into the nature of speech sounds in that it provides us with a criterion for General Redundancy. In Chapter IV, we raised a question: on what criteria do we say that a given redundant feature is a generally redundant one as opposed to a language-

specific redundant feature? It seems that the principle of optimal opposition gives one such criterion. That is, we may say that a redundant feature is a GR feature if it conforms to the notion of optimal opposition. Thus, for example, we can say that Unrounding of Front Vowels and Rounding of Back vowels are GR features, since Front and Back vowels are optimally and maximally opposed by being Unrounded and Rounded, respectively, but not otherwise.

On the other hand, some other cases of GR features seem to be strictly articulatory and physiological in nature. For example, the common phenomenon that when a velar or alveolar stop changes its place of articulation to palatal, the manner is also changed to affricate or fricative (cf. examples on page 38), is probably due to the physiological fact that it is more difficult for the tongue to make a complete closure in a small area in the middle of the hard palate. More obvious kinds of GR features are that most speech sounds are made with egressive air, that sounds with Aperture greater than degree 2 are generally voiced, etc. Thus, here again, we see that consideration of the articulatory mechanism, its capabilities and constraints as a physiological organ, is useful for the construction of a model of universal phonetics, this time by suggesting criteria for the formalization of General Redundancy.

As was noted earlier in Chapter IV, the formalization of General Redundancy is desireable, as it simplifies the phonological description and provides it with a workable, intuitively correct and natural metric in the evaluation of the degree of complexity of segments and rules. We gave an example of syncretism between the Front/Back dimension of vowel and Rounding, noting that /i/ and /u/, for example, are not distinctive in two "full" features but rather in one feature, when compared with such a pair /m/ and /t/ which differ from each other in two features Gravity and Nasality; and that redundancy of this kind is so general that it is desirable to formulate such redundancy as a meta-convention rather than specifying it as a redundancy rule in the phonological component of every grammar. Such a formalization will enable us to describe, e.g., the process of Palatalization with only one rule, the most relevant one that changes Alveolar or Velar stop into Palatal, and to dispense with a rule changing Stop to Affricate, for the latter is now interpreted with reference to a meta-convention of General Redundancy. In this way, Palatalization may be expressed with a single rule, compared with the three rules mentioned in Chomsky and Halle (1965, pp. 122-123. See also footnote 13 of this monograph.):

(48) Lingual Stop → Palatal / ___ Palatal Vowel

Note that this simplification is not made with an *ad hoc* measure but with an explicitly formulated meta-convention of General Redundancy which is linguistically significant.

Needless to say, General Redundancy is not meant to be an absolute condition. (Recall the distinction between Universal R and General R made in Chapter IV.) That is, it does not imply that there is no language

having the segment Rounded Front vowel and/or Unrounded Back vowel, or a genuine Palatal stop. The formalization of General Redundancy only implies that segments or rules having features not conforming with GR conditions will be evaluated as more complex to that extent, as these aberrant features need to be overtly stated, and the symbol-counting evaluation measure will rank the segments or rules more specific and complex than the corresponding segments or rules conforming with the GR conditions in which GR features are not explicitly stated but are implicitly understood with reference to the meta-convention.

In this respect, the formalization of General Redundancy (or "built-in redundancy," as it was called in Chapter IV) is not only a matter of simplification or generalization but is a matter of approaching the level of descriptive and explanatory adequacy in phonology.

# VIII

## AMPLITUDE ASSIGNMENT RULES

Available data on amplitude values of formants are not many, and furthermore, it is difficult to compare what there are, since each set of data has a different reference point relative to which amplitude values are given. Thus, in the sets of data given below (Table 12), values of set (1) are relative to A1 of /ɔ/ at 0 db; those of set (2), relative to A1 = 60 db for every phoneme; those of (3), relative to the baseline of each section; those of set (4), relative to A1 = 0 db for every phoneme; and those of set (5), relative to A1 = 50.75 db for every phoneme. In this kind of situation, it is meaningless to compute average values across different sets of data, as we have done in the case of formant frequency values, even if we shift the values of some sets proportionally according to a logarithmic scale toward an arbitrary uniform reference point, since some data, (2), (4), and (5), show values which are relative within a single vowel only, while others, (1) and (3), give values which are relative to all the other amplitude values in the entire vocalic system. The fact that average values which may serve as the output of our rules are not computable makes us choose only one set of data for such a purpose. We will select (3) as the reference data, the reasons being that

a. (3), together with (1), gives values which are relative within the whole set, not within a single vowel only,

b. (3) gives positive values, while (1) gives negative values,

c. (3) is more extensive than (1), in that (3) contains values of diphthongs, while (1) does not.

We will see below how rules may be inductively arrived at from the limited data.

First, we note that the average value of all A1 in (3) is 39.4 db, that of A2 is 25.2 db, and that of A3 is 15.3 db. We will therefore decide on A1 = 39 db, A2 = 25 db, A3 = 15 db as reference values or "center amplitudes," analogous to center frequencies.

Next, for A1, we note that the average difference in one degree of Aperture is 2.1 db, the values being smaller as the degree of opening increases from High to Mid to Low,

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| High | /i/ - 41 | 3 | /u/ - 41 | 3 | /ɪ/ - 40 | 0 | /ʊ/ - 42 | 2 |
| Mid | /e/ - 38 | | /o/ - 38 | | /ɛ/ - 40 | | /ɔ/ - 40 | |
| Mid | /ɛ/ - 40 | 2 | | | /ɜ/ - 40 | 2 | | |
| Low | /æ/ - 38 | | | | /ɑ/ - 38 | | | |

Table 12. Some published amplitude values of English vowels

(1) Peterson and Barney (1952)

| | i | ɪ | ɛ | æ | ɑ | ʌ | ɔɹ | ɔ | ʊ | u |
|---|---|---|---|---|---|---|---|---|---|---|
| A1 | -4 | -3 | -2 | -1 | -1 | -1 | -5 | 0 | -1 | -3 |
| A2 | -24 | -23 | -17 | -12 | -5 | -10 | -15 | -7 | -12 | -19 |
| A3 | -28 | -27 | -24 | -22 | -28 | -27 | -20 | -34 | -34 | -43 |

(2) Peterson (1961)

| | i | ɪ | ɛ | æ | ɑ | ʌ | ɝ | ɔ | ʊ | u |
|---|---|---|---|---|---|---|---|---|---|---|
| A1 | 60 | 60 | 60 | 60 | 60 | 60 | 60 | 60 | 60 | 60 |
| A2 | 51 | 51 | 54 | 57 | 60 | 55 | 57 | 54 | 52 | 44 |
| A3 | 50 | 52 | 51 | 51 | 44 | 45 | 53 | 34 | 44 | 32 |

(3) Holbrook and Fairbanks (1962)

| | i | ɪ | e | ɛ | æ | ɑ | ʌ | ɝ | ɔ | o | ʊ | u |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A1 | 41 | 40 | 38 | 40 | 38 | 38 | 41 | 40 | 40 | 38 | 42 | 41 |
| A2 | 20 | 25 | 23 | 26 | 28 | 30 | 26 | 26 | 33 | 25 | 23 | 20 |
| A3 | 18 | 22 | 18 | 23 | 24 | 16 | 16 | 20 | 12 | 7 | 14 | 8 |

(4) Arnolds et al. (1958)

| | i | ɪ | ɛ | æ | ɑ | ʌ | ɜ | ɒ | ɔ | ʊ | u |
|---|---|---|---|---|---|---|---|---|---|---|---|
| A1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| A2 | -5 | -2 | -.5 | -.5 | -3 | -6 | -4 | -2 | -6 | -5 | -8 |
| A3 | -8 | -8 | -6 | -6 | -14 | -14 | -10 | -16 | | | |

(5) Holmes et al. (1966)

| | i | ɪ | e | ɛ | æ | ɑ | ʌ | ɜ | ɒ | ɔ | ɑ | ʊ | u |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A1 | 50.75 | 50.75 | 50.75 | 50.75 | 50.75 | 50.75 | 50.75 | 50.75 | 50.75 | 50.75 | 50.75 | 50.75 | 50.75 |
| A2 | 33.25 | 36.25 | 45.5 | 42 | 47.25 | 49 | 43.75 | 45.5 | 47.25 | 45.5 | 50.75 | 42 | 38.5 |
| A3 | 36.75 | 35 | 35 | 38.5 | 38.5 | 29.75 | 31.5 | 33.25 | 22.75 | 22.75 | 33.25 | 28 | 17.5 |

Thus setting 1 db as 1 degree,

(49) A1 = 39 db +

(a) if High, +2d (where 1d = 1 db)
(b) if Low, -2d

Next we note that there is little difference between Front and the corresponding Back vowels, e.g.,

Front /i/ - 41 } 0 /ɪ/ - 40 } 2 /e/ - 38 } 0 /ɛ/ - 40 } 0
Back /u/ - 41 } /ʊ/ - 42 } /o/ - 38 } /ɔ/ - 40 }

Thus, we will need no rule for A1 involving the Front/Back dimension, But there is a greater difference in the Tense/Lax dimension, e.g.,

Tense /i/ - 41 } 1 /u/ - 41 } 1
Lax /ɪ/ - 40 } /ʊ/ - 42 }

Tense /e/ - 38 } 2 /o/ - 38 } 2 /ɑ/ - 38 } 3
Lax /ɛ/ - 40 } /ɔ/ - 40 } /ʌ/ - 41 }

But note that if the vowel is High, it is the Tense counterpart whose A1 value is greater than the corresponding Lax vowel, while if the vowel is non-High, the situation is reversed, i.e., the Lax vowel has a greater A1 value than the corresponding Tense vowel. (This "away-from-neutral" phenomenon in the Tense/Lax feature was discussed in Chapter VI.) Thus,

(49.c) if αTense, αld / High

(49.d) if αTense, -αld / non-High

For A2, we note again that there is little difference in the Front/Back dimension, but that there is about 3 db difference for each one degree difference in Aperture (again, curiously enough, excepting the case involving /ɔ/), Low being greater in A2 value than Mid, which is in turn greater than High, e.g.,

High /i/ - 20 } 3 /ɪ/ - 25 } 1 /u/ - 20 } 5 /ʊ/ - 23 } 10
Mid /e/ - 23 } /ɛ/ - 26 } /o/ - 25 } /ɔ/ - 33 }

Mid /ɛ/ - 26 } 2 /ɜ/ - 26 } 4
Low /æ/ - 28 } /ɑ/ - 30 }

Hence,

(50) A2 = 25 db +

(a) if High, -3d
(b) if Low, +3d

The Tense/Lax difference also amounts to approximately 4 db, and here again, when Low, the Tense vowel has a greater A2 value than the corresponding Lax vowel, but when non-Low, it is the Lax vowel which has a

greater A2 value, e.g.,

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Tense /i/ - 20 | 5 | /u/ - 20 | 3 | /e/ - 23 | 3 | /o/ - 25 | 8 |
| Lax /ɪ/ - 26 | | /ʊ/ - 23 | | /ɛ/ - 26 | | /ɔ/ - 33 | |

| | |
|---|---|
| Tense /ɑ/ - 30 | 4 |
| Lax /ʌ/ - 26 | |

Hence,

(50.c) if αTense, α2d / Low

(50.d) if αTense, -α2d / non-Low

For A3, we find that there is a considerable difference, about 10 db, between Front and Back counterparts, e.g.,

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Front /i/ - 18 | 10 | /ɪ/ - 22 | 8 | /e/ - 18 | 11 | /ɛ/ - 23 | 11 |
| Back /u/ - 8 | | /ʊ/ - 14 | | /o/ - 7 | | /ɔ/ - 12 | |

This phenomenon is noted by Holbrook and Fairbanks (1962) who state that "the differences in A3 appear to characterize front and back vowels as classes" (p. 55). Thus,

(51) A3 = 15 db +

(a) if Front, +5d
(b) if Back, -5d

Instead, there is little difference in "height" dimension (except the /ɑ/, /ɝ/ case, which will be considered separately below), e.g.,

| | | | | |
|---|---|---|---|---|
| High | /i/ - 18 | /ɪ/ - 22 | /u/ - 8 | /ʊ/ - 14 |
| Mid | /e/ - 18 | /ɛ/ - 23 | /o/ - 7 | /ɔ/ - 12 |
| Low | | /æ/ - 24 | | |

But there is about 5 degree difference in the Tense/Lax dimension except in Central vowels, e.g.,

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Tense /i/ - 18 | 4 | /e/ - 18 | 5 | /u/ - 8 | 6 | /o/ - 7 | 5 |
| Lax /ɪ/ - 22 | | /ɛ/ - 23 | | /ʊ/ - 14 | | /ɔ/ - 12 | |

but /ɑ/ - 16 } 0
/ʌ/ - 16 }

Thus,

(51.c) if αTense, -α2d / non-Central

As was noted earlier, A3 value of /ɝ/ is considerably greater than that of /ɑ/, Other data confirm this. For example,

| | | | | |
|---|---|---|---|---|
| Data (1) | /ɑ/ - -28 | Data (2) | /ɑ/ - 44 |
| | /ɝ/ - -20 | | /ɝ/ - 53 |

Since the lowered F3 is a characteristic and vital cue for /ɝ/, a strong A3 is rather expected. Thus,

(51.d) if Mid-Central, +5d

All together, and using the omega-notation where applicable, we have the following amplitude assignment rules:

(49) A1 = 39 db +

(a) if ωHigh, ω2d (where 1d = 1 db)
(b) if αTense, α1d / High
(c) if αTense, -α1d / non-High

(50) A2 = 25 db +

(a) if ωHigh, -ω3d
(b) if αTense, α2d / Low
(c) if αTense, -α2d / non-low

(51) A3 = 15 db +

(a) if ωFront, ω5d
(b) if αTense, -α2d / non-Central
(c) if Mid-Central, +5d

These rules will generate the following amplitude values:

Table 13. Generated amplitude values of English vowels

| | i | ɪ | e | ɛ | æ | ɑ | ʌ | ɝ | ɔ | o | ʊ | u |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A1 | 42 | 40 | 38 | 40 | 38 | 36 | 38 | 40 | 40 | 38 | 40 | 42 |
| A2 | 20 | 24 | 23 | 27 | 26 | 30 | 26 | 23 | 27 | 23 | 24 | 20 |
| A3 | 18 | 22 | 18 | 22 | 22 | 15 | 15 | 20 | 12 | 8 | 12 | 8 |

These values show a close approximation to those of Data (3) which the rules modelled. Except /ɔ/, whose predicted value again differs considerably from the observed data, only two, A1 of /ʌ/ and A2 of /æ/, deviate as much as 3 db. (We will consider the peculiarity of /ɔ/ in the next chapter.)

When we examine the generated values of amplitude closely, we note an interesting phenomenon. That is, it seems that there is a possible connection between formant frequencies and amplitudes. For example, we note that A2 follows the pattern of F1, i.e., as F1 increases, so does A2; and that A3 is proportional to F2, i.e., as F2 decreases from /i/

to /u/, so does A3. We then ask: is this merely a fortuitous relation, or is this a systematically concomitant relation between the two acoustic variables? If it is the latter case, then we will be able to dispense with the rules given above. What we would need instead is a few correlation formulae that would give amplitude values from formant values. If this is possible, then we will simplify our theory of speech specification to that extent, since we eliminate those redundancies that are due to an interdependence of the parameters of specification. We will explore the possibility here.

Fant (1956) asks the same kind of question:

> Given the formant frequencies only, to what extent can the relative vowel intensities, the relative intensity levels of the formants, their bandwidths, and the particular shape of the spectrum envelope be predicted? (p. 109)

Fant then shows indications that there is a rather intimate relation between formant frequencies and formant levels (=amplitudes), and in particular, that "formant bandwidths generally are statistically well correlated with the particular pattern of formant frequencies, and formant levels can be calculated once the frequencies and bandwidths of the three or four first formants and the slope of vocal cord spectrum envelope are given" (p. 117).

Fant, however, does not give any formula(e) that would calculate amplitude values from the given formant values, but several scattered hints of the following kind (a more mathematically oriented reader is referred to Fant et al. 1963 and references 4 - 8 cited there):

(a) We may state that A3 and A4 are very low for the vowels [u] and [ɔ] because both F1 and F2 are low on the frequency scale. In other vowels with higher frequency positions of F1 and F2, A3 and A4 are higher. (p. 115)

(b) In general, when two formants approach each other their levels [=amplitudes] will both increase, e.g., the first and second formants of [a] and [ɔ] and the second and third formants of front vowels. In [i] there is a typical increase in levels A3 and A4 since F3 comes closer to F4 than to F2. (p. 115)

(c) In general, one octave decrease in F1 results in -12 db shift of the levels in all parts above the frequency of the first formant. The level of the first formant is also decreased, but to a lesser extent. (p. 116)

(d) For the American-English [i] A3 is 4 db lower than A2. A2 is at a maximum for [a]. A1 is highest for [a] and varies less than A2 and A3. With a few exceptions for A1, the levels of all three formants increase with i increasing compactness, i.e., increasing F1. (p. 116)

(e) The spectrum level is emphasized in regions where two formants approach each other and the levels of all parts of the spectrum as well as the total energy under the spectrum envelope are primarily determined by F1. (p. 117)

(f) The greater intensity found with the higher F1 is not an independent variable. It must occur because of the higher F1, everything else being equal. Similar relations hold for the gravity feature. When the second formant approaches either the first formant, as in grave vowels or the higher formants as in acute vowels there is a summation effect so that the level of the first formant gains from the second formant in grave vowels and the level of the third and higher formants gains from the second formant in acute vowels, thus contributing to shift the balance in the spectrum to the part occupied by the second formant. (pp. 118-9)

With these suggestions as guides, we will attempt below to compose formulae that would calculate approximate amplitude values from the given formant frequency values.

First, for A1, we note that there seems to be a correlation between F1 and A1 in such a way that the lower the F1, the higher the A1. That is, for /i/ and /u/ whose first formants are the lowest (300 cps), we find the highest A1 value (42 db), and for /a/ whose F1 is the highest (750 cps), we find the lowest A1 value (38 db), etc. Thus, still using the notions "center amplitude" and "degree," we will have the following formula for calculating A1 from F1:

$$(52)\quad A1 = 39 + \left(\frac{550 - F1}{100}\right) \text{ db}$$

Since the center frequency of F1 is 550 cps (cf. p. 72), F1 of 550 cps will give A1 of 39 db, but each 100 cps of F1 in excess of or short of 550 cps will result in the subtraction or addition of 1 db, respectively. The correlation of generated F1 and A1 using the above formula is as follows:

Table 14. Amplitude values of formant one calculated from F1

| | i | ɪ | e | ɛ | æ | ɑ | ʌ | ɣ | ɔ | o | o | u |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| F1 | 300 | 400 | 500 | 600 | 650 | 750 | 650 | 500 | 600 | 500 | 400 | 300 |
| A1 | 41.5 | 40.5 | 39.5 | 38.5 | 38 | 37 | 38 | 39.5 | 38.5 | 39.5 | 40.5 | 41.5 |

For A2, we note the relation between F1 and A2. That is, the higher F1, the greater A2. Thus, A2 of /i/ and /u/, whose F1 is the lowest, is also the least (20 db); and A2 of /a/, whose F1 is the highest, is the greatest (30db). Hence,

$$(53)\quad A2 = 25 + 2\left(\frac{F1 - 550}{100}\right) \text{ db}$$

Since the center frequency of F1 is 550 cps, F1 of 550 cps will give A2 of 25 db (as the value of the right hand term becomes zero), but each 100 cps of F1 above or below 550 cps will result in the addition or the subtraction of 2 db, respectively. The correlation of F1 and A2 using the above formula is as follows:

Table 15. Amplitude values of formant two calculated from F1

| | i | ɪ | e | ɛ | æ | ɑ | ʌ | ɝ | ɔ | o | ʊ | u |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| F1 | 300 | 400 | 500 | 600 | 650 | 750 | 650 | 500 | 600 | 500 | 400 | 300 |
| A2 | 20 | 22 | 24 | 26 | 27 | 29 | 27 | 24 | 26 | 24 | 22 | 20 |

For A3, the formula is more complicated. Firstly, following a general principle that "the levels of all three formants increase with increasing F1" (Fant's hint (d); cf. also (a), (c) (f)), we will write the following:

$$(54.a)\quad A3 = 15 + 2\left(\frac{F1 - 550}{100}\right)\ db$$

But on the other hand, we note a clear correlation between F2 and A3 in that the higher F2, the greater A3 (cf. hint (a)). Thus we rewrite the above as:

$$(54.b)\quad A3 = 15 + 2\left(\frac{F1 - 550}{100}\right) + \left(\frac{F2 - 1400}{100}\right)\ db$$

The formula, as it stands now, will add, to the center A3 = 15 db, 2 db per 100 cps increase of F1 in excess of 550 cps and 1 db per 100 cps increase of F2 in excess of 1400 cps (=F2 center frequency), and will subtract 2 db per 100 cps decrease of F1 and 1 db per 100 cps decrease of F2. It will give the following A3 values:

Table 16. Amplitude values of formant three calculated from F1 and F2

| | i | ɪ | e | ɛ | æ | ɑ | ʌ | ɝ | ɔ | o | ʊ | u |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A3 | 17 | 17 | 20 | 20 | 21 | 16 | 17 | 12 | 12 | 8 | 9 | 5 |

These values show a close approximation to the desired values except in the case of /ɝ/, /o/, and /u/. We note that these exceptional vowels are those that have lower third formants. Fant says that "when two formants approach each other their amplitudes will both increase" (hints (b) and (f)). Lower F3 means that it is to that extent closer to F2, and therefore to that extent has a greater A3. Hence, we will again rewrite the formula, incorporating this compensational factor:

$$(54)\quad A3 = 15 + 2\left(\frac{F1 - 550}{100}\right) + \left(\frac{F2 - 1400}{100}\right) + \left(\frac{2500 - F3}{100}\right)\ db$$

This formula will now give the following A3 values:

Table 17. Amplitude values of formant three calculated from F1, F2, and F3

| | i | ɪ | e | ɛ | æ | ɑ | ʌ | ɝ | ɔ | o | ʊ | u |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A3 | 17 | 17 | 20 | 20 | 21 | 16 | 17 | 20 | 12 | 8 | 11 | 8 |

Below, we will give the three sets of amplitude values,

A. Data (3) of Table 12
B. Values generated by rules (49) - (51)
C. Values calculated by formulae (52) - (54)

for comparative purposes.

Table 18. A comparison of three sets of amplitude values: A. observed data, B. rule-generated, and C. calculated from formant values

| A1 | i | ɪ | e | ɛ | æ | ɑ | ʌ | ɝ | ɔ | o | ʊ | u |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A. | 41 | 40 | 38 | 40 | 38 | 38 | 41 | 40 | 40 | 38 | 42 | 41 |
| B. | 42 | 40 | 38 | 40 | 38 | 36 | 38 | 40 | 40 | 38 | 40 | 42 |
| C. | 41.5 | 40.5 | 30.5 | 38.5 | 38 | 37 | 38 | 39.5 | 38.5 | 39.5 | 41.5 | 41.5 |

| A2 | i | ɪ | e | ɛ | æ | ɑ | ʌ | ɝ | ɔ | o | ʊ | u |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A. | 20 | 25 | 23 | 26 | 28 | 30 | 26 | 26 | 33 | 25 | 23 | 20 |
| B. | 20 | 24 | 23 | 27 | 26 | 30 | 26 | 23 | 27 | 23 | 24 | 20 |
| C. | 20 | 22 | 24 | 26 | 27 | 29 | 27 | 24 | 26 | 24 | 22 | 20 |

| A3 | i | ɪ | e | ɛ | æ | ɑ | ʌ | ɝ | ɔ | o | ʊ | u |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A. | 18 | 22 | 18 | 23 | 24 | 16 | 16 | 20 | 12 | 7 | 14 | 8 |
| B. | 18 | 22 | 18 | 22 | 22 | 15 | 15 | 20 | 12 | 8 | 12 | 8 |
| C. | 17 | 17 | 20 | 20 | 21 | 16 | 17 | 20 | 12 | 8 | 11 | 8 |

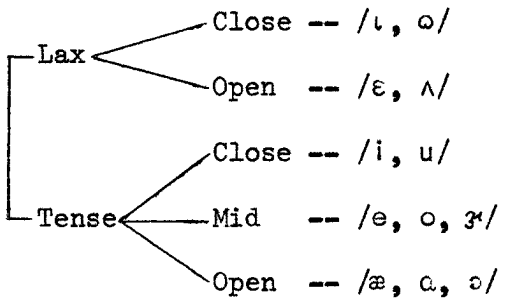
Except in the case of A2 of /ɔ/, the three sets of amplitude values for each formant of each vowel show a close similarity, the range being generally not more than 2 db. Needless to say, the values generated by either rules or formulae are not meant to be absolute but relative. As in the case of formant assignment, we can achieve flexibility and relativity of values by changing either or both the initial values of center amplitudes and the value of 1 degree. It should be also mentioned that the formulae are only a crude approximation, and that the final validity of them must wait for their extensive and successful application to an acoustic speech synthesizer controlled by a computer in whose program the formulae are integrated.

# IX

# DURATION ASSIGNMENT RULES

As primary linguistic data from which the rules of duration of vowels of American English are to be induced, we will mainly rely on House (1961), since House's data have the advantage of being nicely organized in terms of phonetic categories that we are familiar with. The data, however, are a little defective in that the recorded materials are not real speech utterances but nonsense syllables, and that there is no measurement of diphthongs and the contextual influence of nasals, liquids, pause, etc. We will draw whatever supplementary information and supporting evidence is available from other data, such as Heffner (1937-1943), Rositzke (1939), Peterson and Lehiste (1960), Delattre (1962), Zimmerman and Sapon (1958), Fishcer-Jørgensen (1965), etc.

House's data is given below (Table 19), organized according to what House regards as variables of duration. The speech material were bisyllabic nonsense utterances of three adult males, and each nonsense word consisted of [həCVC] where the two C's were the same phoneme. The twelve vowels and fourteen consonantal environments tested were: /i, ɪ, e, ɛ, æ, ɑ, ʌ, ɔ, o, ʊ, u, ɝ/ and /p, t, k, f, θ, s, tʃ, b, d, g, v, ð, z, dʒ/. House's categorization of vowels is as follows:




Needless to say, House's "Close" is equivalent to our "High," and House's "Open" to our "Low." Yet, House's categorization of English vowels differs from ours (Figure 11) in three respects:

1. House collapses /ɛ/ and /ʌ/ under "Lax-Open,"; we distinguish them further, /ɛ/ as "Lax-Mid," and /ʌ/ as "Lax-Low."
2. House regards /æ/ as "Tense-Open;" we regarded it as "Lax-Low."
3. House regards /ɔ/ as "Tense-Open;" we regarded it as "Lax-Mid."

We will discuss implications of these discrepancies later.

What House's chart plainly shows is that, going from left to right, column by column:

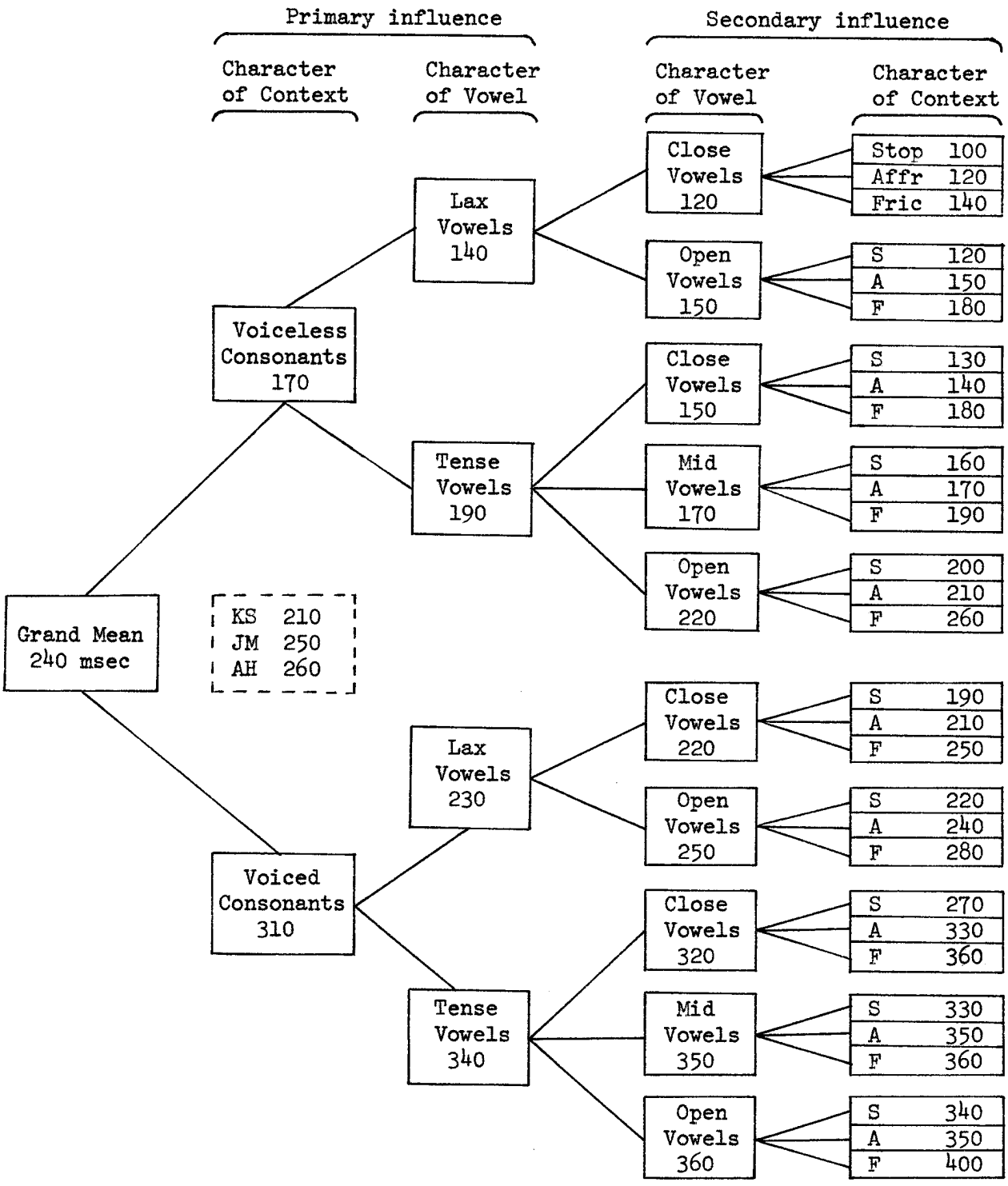



Table 19. Variations in vowel length in four different phonetic contexts (from House (1961), p. 1176. Times in msec.)

(1) a vowel followed by a voiceless consonant is shorter than a vowel followed by a voiced consonant;
(2) a lax vowel is shorter than its corresponding tense vowel;
(3) close (high) vowels are shorter than mid vowels, and mid vowels are in turn shorter than open (low) vowels;
(4) a vowel followed by a stop consonant is shorter than a vowel followed by an affricate, and this in turn is shorter than a vowel followed by a fricative consonant.

All the above phenomena are, of course, under the condition "ceteris paribus," and what is interesting is the degree of their interplay, and their relevancy to the systematic synthesis of English. We believe that House's presentation indicates a systematic behavior of some variables of vowel duration, and that it is worth examining these variables more closely to determine the relevancy of each variable.

As is shown, House selects four variables of English vowel duration. Is his choice valid? To answer this question, we will have to decide whether each and every variable reflects an "extrinsic" phenomenon. According to the argument given in Chapter II, "intrinsic" features are not to be included as variables in the systematic synthesis. Rephrasing the argument briefly here with reference to a practical situation, we will say that, if a phonetic feature of English is negligible in such a way that its neglect by a foreign speaker does not make his English sound outlandish solely for that reason, the negligibility being due to the inherent mechanistic behavior involved in the production of the phonetic feature in question, this behavior being the same for speakers of all languages, then this is an "intrinsic" feature; furthermore, we may say that, by definition, an intrinsic feature is not a part of systematic synthesis, since the function of systematic synthesis is to characterize different and specific phonetic characteristics of different languages and dialects, and an "intrinsic" feature plays no role in such a characterization. Likewise, we will say that, if a phonetic feature of English is not negligible in such a way that its neglect by a foreign speaker renders his English utterances as outlandish for that reason, then it is a fact of English to be characterized as such by the rules of systematic synthesis of English. We will call this latter kind of feature "extrinsic."

Thus, only if House's four variables are all extrinsic features, and, furthermore, only if there is no other extrinsic variable that House failed to detect, then our rules of duration and the values generated by them need to approximate those of House.

House thinks that what he calls secondary influences (two right columns) are intrinsic features and the primary influences (two left columns), extrinsic features:

> It is appealing to speculate that some inherent articulatory influences . . . are the manner of production of consonant contexts and the open-close dimension of vowel articulation. (p. 1176)

But before we take the statement for granted, let us further examine the precise characteristics of the four phenomena that seem to have an influence on the length of English vowels.

(1) The voicing of the following consonant.

Although all reports agree that the duration of vowel is significantly lengthened before the voiced consonants in English, it is not certain as to whether such a phenomenon is inherent and universal. Zimmerman and Sapon (1958) report that, from the point of view of the duration of a vowel preceding a voiced consonant, English and Spanish are qualitatively similar but quantitatively very different. That is, vowels in both languages are lengthened before voiced consonants, but the mean difference between a vowel preceding a voiced consonant and a vowel preceding a voiceless consonant in Spanish is found to be a mere 18 msec, while in English it is 83 msec. It should be further noted that in Spanish, voiced stops are phonetically fricatives. Thus, if we assume that the manner of articulation of the following consonant has a lengthening effect at all, as it seems to do (cf. House's rightmost column), then at least some of the 18 msec difference should be credited to this fricativeness. Kozhevnikov and Chistovich (1966) also report about 21 msec average difference in the preconsonantal vowel length in Russian (p. 107). Thus, it seems that if the lengthening effect of a voiced consonant on the preceding vowel is universal at all, the degree of the effect is language-specific. Furthermore, the fact that such English pairs as *rider - writer* have the same voiced flap [ɾ] and yet may differ in the length of the preceding vowels testifies that there is no inherent physiological reason why a vowel preceding a voiced consonant must be lengthened. In view of these facts, we will conclude that the effect of the voicing of the following consonant on the preceding vowel is not a universally intrinsic phenomenon but an extrinsic phenomenon of English.

(2) The manner of articulation of the following consonant.

The literature on this point is scanty, and what there is is often conflicting. Zimmerman and Sapon (1958) notes that "there appears to be no consistent pattern in terms of place or manner of articulation of the following consonant regarding its effect on vowel duration." (p. 153) On the other hand, Peterson and Lehiste (1960) report that in English "the voiced fricatives appear to have a further lengthening effect" (p. 701), although they find that affricates behave in the same manner as stops. Delattre (1962) is more inclusive, and states that a vowel is shorter before a liquid and longer before a "solid" consonant, and it is shorter before a stop and longer before a fricative; but he argues that all of these factors are physiologically conditioned, i.e., intrinsic, not "learned," i.e., extrinsic variations (p. 1142), without giving much evidence, however. It seems that the issue is still undecided, and calls for an experimental investigation. In this kind of situation, we have an option: either we may exclude the factor from the rules assuming that it is an intrinsic phenomenon, or we may incorporate it into the rules assuming that it is an extrinsic feature. We will do the latter just

for the reason that, even if the factor of the manner of the following consonant were an intrinsic one, our acoustic speech synthesizer does not have such a built-in restriction which a living human vocal tract has. This measure will thus produce output values closer to the real data.

Before we leave the discussion of the consonantal influence on the vowel length, let us briefly consider whether a preceding consonant affects the length of the following vowel, and whether the place of the final consonant has any effect at all.

There seems to be general agreement that the influence of the initial consonant upon the duration of the following vowel is negligible. For instance, Peterson and Lehiste (1960) state that "initial voiced-voiceless contrasts presented no discernible pattern" (p. 700), and Delattre's (1962) eight factors of vowel duration, intended to be an exhaustive list, contain nothing involving an initial consonant.

There seems to be, however, a durational factor in the place of articulation of the following consonant. Data of Lehmann and Heffner (1943), Zimmerman and Sapon (1958), Peterson and Lehiste (1960), Delattre (1962), etc. indicate that a vowel is longest before a velar consonant and shortest, in general, before a labial consonant. Lehmann and Heffner (1943) attempt to explain this phenomenon in terms of the facility of articulation of the consonant in question:

> The increased length of vowel is due to less skill in enunciating the consonants concerned. We assume that less skill means slower response to the stimulus evoking the movement. (p. 214)

Fischer-Jørgensen (1965) similarly hypothesizes that the delay in the execution of the command to articulate the following consonant causes a prolongation of the preceding vowel. But, in Fischer-Jørgensen, what causes the delay is not "less facility" in articulating an infrequently used consonant but the greater distance in places of articulation between the vowel and the following consonant. That is,

> The vowel is longer the greater the distance is between the place of articulation of the vowel and the following consonant. (p. 205)

Thus, she finds that [u] is shortest before [b] but longest before [d], while [i] is hortest before [d] and longest before [b] or [g].

It is transparent that, whichever assumption is correct, the phenomenon is due to the inherent dynamics of the tongue mechanism. It therefore need not be internalized as a rule by the speaker and by our systematic synthesis.

(3) The Tense/Lax dimension of the vowel.

If one of the primary articulatory correlates of Tenseness is the

overall higher relative amplitude or intensity, it may be speculated that the increase in effort affects duration proportionally. However, if we define Tenseness as being "away-from-neutral" articulation, and accept the hypothesis that a greater effort tends to shorten the length (i.e., the principle of the least effort), then we should assume that the two factors more or less compensate each other and there would not be as much as 100 msec difference between the pair Tense and Lax. We will thus agree with House that the "conditioned" explanation is "untenable as an overall explanation" (p. 1177). Delattre (1962) suuport House:

> Historical facts help confirm House's contention and indicate that the /i/I/ difference of length is *learned* whereas the /E/I/ is *conditioned*: /i/ is longer than /I/ today not because it is less open - due to an articulatory conditioning - but because of the survival of a former (Middle English) distinctive feature *long/short*/i:/i/ which gradually changed to a rather *less central/more central* articulatory distinction /i/I/ with attenuation (but not extinction) of the old *long/short* distinction. (p. 1142)

We will thus regard the Tense/Lax distinction as an extrinsic feature of English requiring a rule-generation.

(4) The open/close (Low/High) dimension of the vowel.

House's data on this are supported by the authors of *Preliminaries* who state that "the more diffuse [higher] vowels are, ceteris paribus, shorter than the more compact [lower]." (p. 36) A phonetic explanation for this kind of phenomenon is not given anywhere. House's following "hypothesis":

> it seems reasonable to hypothesize that the articulation of close vowels . . . may represent less muscular adjustment from a physiologic rest position of the vocal tract and may consequently require relatively less muscular effort than the production of sounds requiring more deviation from the rest position (p. 1177)

does not "seem reasonble"; our doubt is not that a physiologic rest position requires less muscular effort, but that high vowels require less muscular adjustment than low vowels. On the contrary, higher vowels seem to require a tenser articulation, as a certain part of the tongue must bunch up forming a more unnatural and distorted configuration of the tongue compared to that of lower vowels. If this speculation is correct, we may then say that, by the principle of least effort, tenser high vowels tend to be shorter than laxer low vowels, and that, therefore, the phenomenon is an intrinsic, physiologically determined one.

This phenomenon, however, raises an interesting problem, which is the fact that the two dimensions of vowels, High/Low and Tense/Lax,

function oppositely in their influence on the vowel length associated with the height of the vocalic articulation. That is, vowel duration increases as the tongue height decreases in one case (High/Low), but as the tongue height increases in the other (Tense/Lax, i.e., between a pair of a Tense and its corresponding Lax vowel, the Tense vowel is higher and longer than the Lax vowel). This problem has also been noted by Halle, and we find the following interesting statement in a footnote to one of his recent articles:

> I regard the distinction between the two types of /o/ [/o/ and /ɔ/] and of /e/ [/e/ and /æ/] as one of noncompact versus compact, rather than as one of tense versus lax. (1964b, p. 349, fn.16)

This means that Halle regards the difference between /i/ and /ɪ/, /u/ and /ʊ/ as that of Tenseness, while that between /e/ and /æ/, /o/ and /ɔ/ as that of Compactness. In terms of the intrinsic length of vowels, Halle's contention seems to be correct, since if we look at the two following figures (14 and 15) by Peterson and Lehiste (1960) and House (1961), both show that, while /ɪ/ and /ʊ/ are shorter in length than their Tense counterparts /i/ and /u/ respectively, /ɔ/ (a supposed Lax counterpart of /o/) and /æ/ are longer in length than /o/ and /e/ respectively.

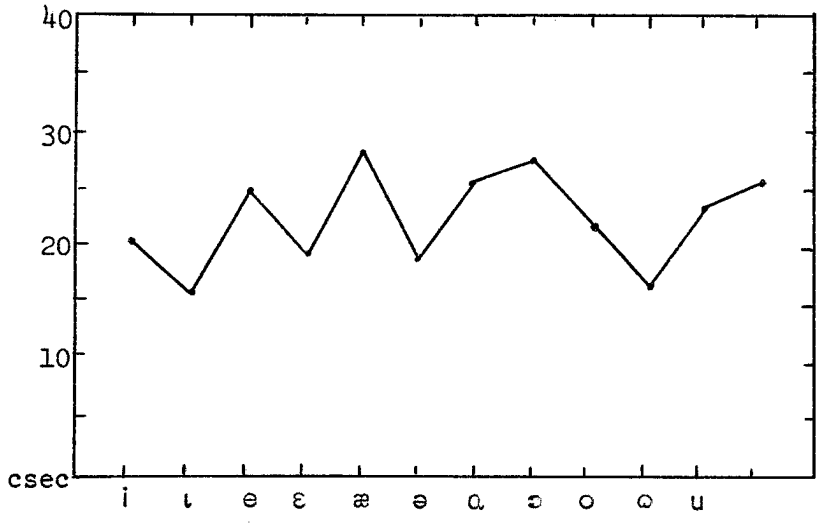



Figure 14. A graph of English vowel lengths showing /e/ being longer than /ɛ/, but /o/ shorter than /ɔ/ (from Peterson and Lehiste 1960, p. 701)

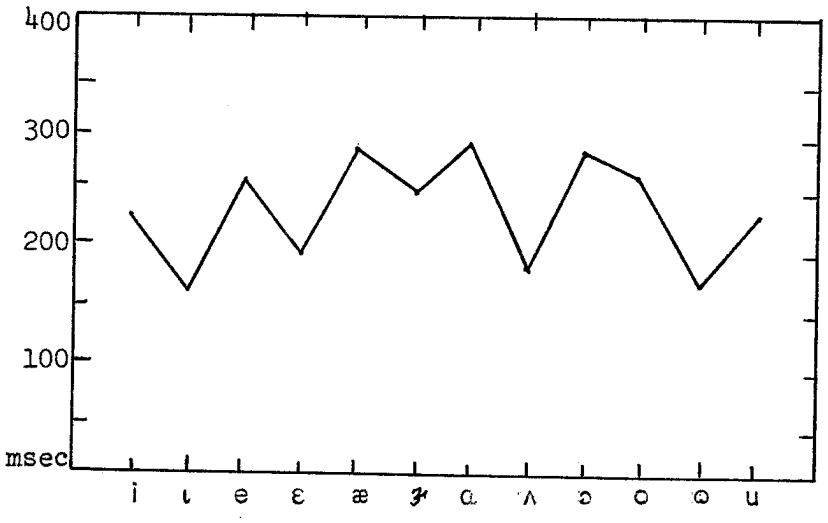



Figure 15. Same as Figure 14 from a different source (from House 1961, p. 1175)

If we assume that Tense vowels are redundantly long and infer that long vowels are therefore Tense vowels, then the above data show that /æ/ and /ɔ/ belong to the set of Tense vowels. This is not in conformity with our categorization of them, as we categorized them as Lax vowels. Other data support House and Peterson-Lehiste's observation. For instance, Heffner (1937) states that:

> [æ] is definitely at the long end of the series. (p. 132)

> The vowel of *bought, taught, caught,* etc., is essentially as long as the vowel of *boot, shoot, bait, date, feet, beat.* (p. 132)

Rosotzle (1939) also classifies both [æ] and [ɔ] as long vowels:

Long: [i, u, e, o, æ, ɔ, ɑ]
Short: [ɪ, ʊ, ʌ, ɛ]

There is however some difference in opinion about the intrinsic length of /æ/. For example, Meyer (1903) classifies British English vowels as:

Long: [i, u, e, o, ɔ]
Short: [ɪ, ʊ, ʌ, ɛ, æ, ɒ]

Heffner (1937) is also hesitant about /æ/:

> [æ] is found among the shorter vowels in one phonetic environment and among the longer vowels in another environment. (p. 131)

This environment is specified by Jones (1960) as:

> In the South of England, a fully long [æ:] is generally used in the adjectives ending in *-ad* (e.g., *bad, sad,* etc.) and is quite common in some nouns, e.g., *man, bag, jam,* etc. The [æ] appears to be more usually short in nouns ending in *-ad,* e.g., *lad, pad,* etc. (§874, p. 235)

Would it be that the short / æ / of British English has changed to long / æ / in American English, as Meyer and others imply? Or are there two distinctive forms of / æ /, Tense and Lax, as Jones seems to indicate? If so, would there be a difference in the formant quality also? We leave these questions unanswered and return to the case of /ɔ/.

There is another sort of evidence that supports /ɔ/ being a Tense vowel rather than a Lax vowel. It is a fact of English that only Tense vowels can occur in "unchecked" (or "open") monosyllabic words, and that /ɔ/ is one of them, e.g.,

/i/: *bee, fee, key, sea, pea, tea,* etc.
/e/: *bay, say, Kay, pay, ray, may,* etc.
/u/: *do, woo, two, who, you,* etc.
/o/: *go, toe, show, doe, no, row,* etc.
/ɝ/: *err, fur, her, blur, per, sir,* etc.
/ɑ/: *ah, bah, da, ha, ma, pa, rah, yah,* etc.
/aw/: *how, cow, bough, plough,* etc.
/ay/: *by, high, my, pie, thigh, sigh, tie,* etc.
/ɔy/: *boy, coy, hoy, joy, toy, soy,* etc.
/ɔ/: *raw, saw, law, paw, jaw, haw, maw, shaw, yaw,* etc.

(Note that / æ / does not occur in this context except in one onomatopeotic word *baa.*)

The kinds of data shown above seem to be convincing enough to indicate that Halle is correct in suggesting that /ɔ/ is opposed to /o/, not as Lax to Tense, but as Low to Mid (Compact to non-Compact in Halle's terms). We will therefore assign "Low-Tense" to /ɔ/.

This new categorization seems to remedy the "aberrant" behavior of /ɔ/ that has been noted on a few occasions earlier. For example, A2 of /ɔ/ will now be assigned the value 30 db, instead of 27 db, which is much closer to Holbrook and Fairbanks' (1962) value 33 db. Formant values of /ɔ/ also now seem to be closer to the observed data.[21]

---

[21]Formant values of /ɔ/ as "Low-Tense-Back" vowel will now be assigned by rules the following values:

F1 = 800, F2 = 700, F3 = 2550

It is interesting to note that the F2 value is predicted by rules to be lower than the F1 value. The phenomenon is known to occur in Low-Back vowel (cf. Stevens and House 1955), although the crossing is not observable due to the coupling effect. Here, we have to say only that the lowest formant, however produced, is F1 by definition.

Going back, then, to House's data, we find that the mean difference in length between:

| | | | |
|---|---|---|---|
| a Tense V and a Lax V | is | 30 msec, if / ___Voiceless C | (1) |
| " | is | 100 msec, if / ___Voiced V | (3) |
| a V followed by a Voiceless C and that followed by a Voiced C | is | 100 msec, if V is Lax | (3) |
| " | is | 160 msec, if V is Tense | (5) |
| a V followed by a Stop and that followed by an Affricate | is | 20 msec | (1) |
| a V followed by an Affricate and that followed by a Fricative | is | 30 msec | (1) |

For simplicity and elegance, we will regard the difference of 30 msec as one degree of difference and accordingly assign integers of the closest multiples of 30 to each mean difference. These integers are shown in the right hand column above. Our rules will be, then, a matter of assigning these integers properly to vowels which are assumed to have a certain basic and neutral length to start with (we will take the "grand mean" 240 msec as this value), and then converting the integers into actual length.

One further fact that should be mentioned here before formulating rules is that the voicing of the following consonant affects the preceding vowel disproportionately according to whether the vowel is Lax or Tense. That is, the lengthening effect of the voiced consonant is generally 2 degrees greater when the vowel is Tense than when it is Lax. The same phenomenon was also noted by Peterson and Lehiste (1960). In the following (Table 20), we find that in Tense vowels, the effect of the following voiced consonant is twice as great as that in Lax vowels. This fact should and will be incorporated in our rules.

Table 20. Vowel lengths in different contexts showing a greater influence of the voicing of the following consonant in Tense vowels than in Lax vowels (from Peterson and Lehiste 1960, p. 702. Times in csec. Differnces computed by me)

| Following C | Dur. of Lax V | Diff. | Dur. of Tense V | Diff. |
|---|---|---|---|---|
| p<br>b | 13.8<br>20.3 | 6.5 | 18.8<br>30.7 | 11.9 |
| t<br>d | 14.7<br>20.6 | 5.9 | 21.0<br>31.8 | 10.8 |
| k<br>g | 14.5<br>24.3 | 9.8 | 20.0<br>31.4 | 11.4 |
| f<br>v | 19.2<br>23.1 | 3.9 | 26.1<br>37.4 | 11.3 |
| θ<br>ð | 20.8<br>26.0 | 5.2 | 26.5<br>38.1 | 11.6 |
| s<br>z | 19.9<br>26.2 | 6.3 | 26.9<br>39.0 | 12.1 |
| tʃ<br>dʒ | 14.5<br>19.1 | 4.6 | 19.8<br>30.0 | 10.2 |

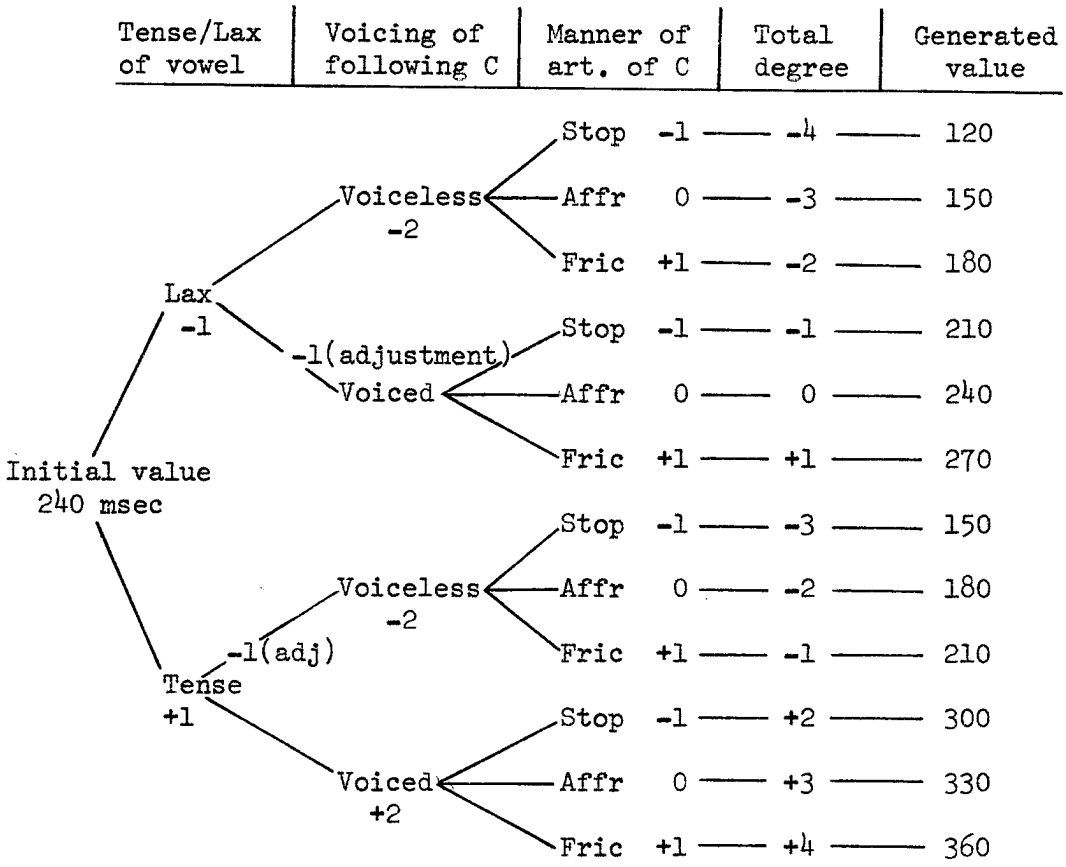
The following are, then, the rules of vowel duration:

(55) Duration = 240 msec +

(a) if Tense, 1d (where 1d = 30 msec)
(b) if / $\underline{V}$ αVoice C, α2d
(c) if / $\underline{V}$ Stop, -1d
(d) if / $\underline{V}$ Fricative, +1d
(e) if [αTense V] [-αVoice C], -1d (adjustment discussed in the preceding paragraph)

It is again easily seen that the use of this rule-schema allows for flexibility of values. For example, the initial value may be changed to 210 msec to model the speaker KS (see the dotted box in House's chart, Table 19), or to 260 msec to model the speaker AH, etc. The degree value may also be changed, either to 20, 25, 33 msec, etc. This way, we will be able to control the rate of speech. As was the case with the formant frequency and amplitude assignment rules, there is no internal linear ordering in the above rules.

The following is a derivational tree, the rightmost column showing the final values generated:




There is no direct way to compare these values with House's, as (1) House groups both /ɛ/ and /ʌ/ as "Lax-Open" while we distinguish them, /ɛ/ as Mid and /ʌ/ as Open-Low, and (2) we have decided that the length difference associated with the vowel height is an "intrinsic" feature, and therefore not to be rule-generated. Suppose, however, that for comparative purposes, we add the following:

(55.f) if ωHigh, -ωld

We can then make the following comparison:

Table 21. Comparison of vowel length values generatee by rules with those observed (House 1961).

(1) Tense/Lax of Vowel
(2) Voicing of the following Consonant
(3) Manner of articulation of the following Consonant
(4) The height of the Vowel
(5) House's (1961) values
(6) Values generated by rules (55.a - 55.e)
(7) Values generated with the additional rule (55.f)

| (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|
| Lax | Voiceless | Stop | High | 100 | 120 | 90 |
| | | | Mid | 120 | | 120 |
| | | | Low | | | 150 |
| | | Affr. | High | 120 | 150 | 120 |
| | | | Mid | 150 | | 150 |
| | | | Low | | | 180 |
| | | Fric. | H | 140 | 180 | 150 |
| | | | M | 180 | | 180 |
| | | | L | | | 210 |
| | Voiced | S | H | 190 | 210 | 180 |
| | | | M | 220 | | 210 |
| | | | L | | | 240 |
| | | A | H | 210 | 240 | 210 |
| | | | M | 240 | | 240 |
| | | | L | | | 270 |
| | | F | H | 250 | 270 | 240 |
| | | | M | 280 | | 270 |
| | | | L | | | 300 |
| Tense | Voiceless | S | H | 130 | 150 | 120 |
| | | | M | 160 | | 150 |
| | | | L | 200 | | 180 |
| | | A | H | 140 | 180 | 150 |
| | | | M | 170 | | 180 |
| | | | L | 210 | | 210 |
| | | F | H | 180 | 210 | 180 |
| | | | M | 190 | | 210 |
| | | | L | 260 | | 240 |
| | Voiced | S | H | 270 | 300 | 270 |
| | | | M | 330 | | 300 |
| | | | L | 340 | | 330 |
| | | A | H | 330 | 330 | 300 |
| | | | M | 350 | | 330 |
| | | | L | 350 | | 360 |
| | | F | H | 360 | 360 | 330 |
| | | | M | 360 | | 360 |
| | | | L | 400 | | 390 |

The comparison shows that the two sets of values approximate each other very closely, the mean deviation of generated values from House's measurements not exceeding more than 10 msec.

Needless to say, the rules given above are not complete. Through the lack of adequate data, we have not considered several other factors of vowel duration. As more data are reported, we will have to incorporate them into our rules. For example, if Delattre's (1962) statement that Liquids shorten the preceding vowel further, while a Nasal consonant lengthens it beyond the degree that other voiced consonants do (cf. "A vowel preceding a nasal consonant is prolonged more than in articulation of a corresponding voiced consonant." (Kozhevnikov et al. 1965. English translation 1966, p. 165)), and if the degree of this "further" lengthening is assumed to be one, then we might expand our rules by adding the following:

(55.g) if / ___ Liquid, -ld

(55.h) if / ___ Nasal C, +ld

Stress is also likely to increase the duration of vowels as well as amplitude and formant frequencies (cf. Fry 1958; Lehiste and Peterson 1959), and so is pause, although in this case amplitude and frequency are likely to decrease.

Nonetheless, it will be interesting to apply and test the rules in the synthesis of running speech.

# X

# SUMMARY AND CONCLUDING REMARKS

In this monograph, an attempt has been made to demonstrate:

(1) that the scope of phonetic specification needs to be extended a step beyond what is presently known as the level of systematic phonetics for the reason that matrices at this level still lack some empirical content that is linguistically significant. For example, consider the much disputed case of "juncture." It is known that in such pairs as *my train - might rain, grey tie - great eye, a name - an aim, nitrate - night rate,* and scores of others, there are several consistent phonetic differences. In particular, in the pair *my train ∿ might rain.*

1. the vowel /ay/ in *my* is longer than the same vowel in *might,*
2. /t/ of *train* has a stronger aspiration than /t/ of *might,*
3. /t/ of *train* is retroflexed, and /t/ of *might* is not,
4. /r/ of *train* is (partially) voiceless, while /r/ of *rain* is fully voiced.

Linguists have been reluctant to recognize these facts, however, as these are subphonemic phonetic details, and the bricks (building blocks) of their linguistic structures were phonemes. Thus, they assign a collective term "juncture," which has no phonetic entity uniquely associated with it, to this phenomenon, as if it is nothing more than the mortar between the bricks. Needless to say, these facts are linguistically relevant and significant; hearers differentiate phrases with reference to these facts. But the significance will not be given substance without rules of "systematic synthesis," the term proposed for this third level of phonology (Chapter II).

(2) that a speech synthesizer plays a significant role in linguistic description in that it permits one to test and (in)validate a hypothesis made in an abstract higher level of phonology. For instance, if it is hypothesized that in English Tenseness is a redundant feature of Voicing or vice versa, the use of a speech synthesizer that has the possibility of independent control of the variables enables one to see whether or not the hypothesis is true (Chapter III).

(3) that there are several aspects in Jakobsonian Distinctive Feature theory that require serious reconsideration, in particular,

a) the claim that features are necessarily to be categorized in binary terms,

b) the claim that properties of DF's are perceptual,

c) the convention by which the features Gravity and Diffuseness are used for both consonants and vwoels,

d) the definition of natural class,
e) the treatment of redundant features.

It was argued that

a') it is not empirically desirable to describe certain phonetic categories, notably those of physical continua such as tone, vowel height, etc., in binary terms,

b') it is premature to talk about "universal perceptual categories" due to the fact that perception is largely dependent upon a specific phonological pattern of an individual language and that contributions from neurophysiology and psychoacoustics are yet to come,

c') an attempt to use certain features in categorizing both consonants and vowels has justification neither from considerations of economy nor from an empirical point of view; that is, the features Gravity and Diffuseness encompassing both vowels and consonants do not constitute natural classes so that any segment involving Gravity or Diffuseness must always be specified as to its vocality or consonantality (no economy), and that a DF specification of Palatalization requires two separate rules one of which implies a dissimilation rather than an obvious assimilation (counter-factual),

d') the present DF definition of "natural class" is not well-defined,

e') redundant features may be classified into General and Specific, and General Redundancy is best treated as a set of meta-conventions applying universally to all languages so that redundancy rules of the general nature, e.g., Front V → unRounded, Palatal Stop → Affricates, etc., will not appear in every phonological grammar (Chapter IV).

(4) that considerations of this sort have led to the proposal of an alternative framework of universal phonetics. We based the categories of our model on articulatory parameters for the simple reason that the limitations and dynamics of the vocal tract are uniform for speakers of all languages so that talking about universal phonetics is nowhere more natural, more practical, and simpler than in the area of physiology and in terms of articulatory parameters. The proposed parameters and their categories were:

Table 22. A proposed set of universal phonetic categories based on articulatory parameters

| Parameter | Macrocategory | Subcategory | Microcategory |
|---|---|---|---|
| Degree of Aperture | Consonantal | 0 (stop)<br>1 (fricative) | 01 (affricate)<br>12 (fricative liquid) |
| | Sonantal | 2 (liquid)<br>3 (approximant) | |
| | Vocalic | 4 (close V)<br>5 (mid V)<br>6 (open V) | 45 (half-close)<br>56 (half-open) |
| Plsce of Articulation | Labial | labial | bi-labial<br>labio-dental |
| | Lingual | alveolar | dental<br>alveolar<br>post-alveolar |
| | | palatal | pre-palatal<br>palatal |
| | | velar | pre-velar<br>velar<br>uvular |
| | Laryngal | glottal | pharyngal<br>glottal |
| Manner of Articulation | Nasal | nasalized | |
| | Oral | spread<br>labialized<br>retroflexed<br>palatalized<br>velarized<br>pharyngalized<br>glottalized<br>tense<br>lax | |
| State of Glottis | Voiced | voice-proper<br>creaky<br>murmur<br>whisper | |
| | Voiceless | aspirated<br>unaspirated | |
| Direction of Airstream | Egressive<br>Ingressive | | |

An attempt was made to give rules of hierarchy of categories within each parameter in order to describe certain typological universals such as "no fricatives without stops," "no palatal consonant without velar consonant," "no mid vowels without high and low vowels," etc. It was claimed that the new model provides a satisfactory way to define such notions as "phonetic similarity," "natural class," "optimal opposition," etc. (also Chapter IV).

(5) that at the level of systematic synthesis, values of sounds may best be generated internally using the notions of center values and degree, instead of giving values in a look-up table, the reason being that the former measure is more economical and reflects the relative and varying nature of speech sounds (Chapter V).

(6) that the formalization of the notion "optimal opposition" simplifies our phonological description, that A optimally opposed to B ($A = \omega B$) is a separate and independent notion from A is non-B ($A = \nu B$), and that this notion optimal opposition and some obvious inherent physiological constraints provide criteria for General Redundancy (Chapter VII).

Then, in accordance with principles stated in (5) above (Chapter V), we gave, for English vowels,

(7) rules of formant frequency value assignment (Chapter VI),

(8) rules of amplitude value assignment, and alternatively,

a set of formulae as a function of formant frequency values (Chapter VIII), and

(9) rules of duration value assignment (Chapter IX).

* * *

In the theory of phonology at the present state of linguistics, there are a number of questions to be answered and a number of problems to be solved. We have raised a few, and have attempted to answer them in this monograph. Many others lie outside the scope of our present consideration. Still, the monograph is not complete in itself, as there seem to be cases for which no definite answers can be given at the present time. We will consider a few of them before we close the book.

Firstly, we argued in Chapter V that the most economical and natural way to assign physical and numerical values to phonetic categories is via rules of the sort exemplified in Chapters VI, VIII, and IX, rather than via a customarily assumed look-up table. The examples showed that the new measure is feasible and desirable in many ways in the case of vowels. Yet we do not know whether the same measure will apply with equal desirability and advantage to consonants. Although acoustic data on English consonants are less readily available, it seems that the same measure may be feasible, judging from bits of information, e.g., nasal formants

are in the same position regardless of the place of articulation of a particular nasal consonant, and the acoustic locus for a given place of articulation is the same regardless of the manner or articulation, according to reports by the Haskins group. But some fundamental differences in phonetic nature between vowels and consonants make us doubt whether the new measure is profitable, though it may be workable, in the case of consonants. For example, consonants are in nature less continuous, less variable, and less flexible. Since one motivation for introducing the new measure was to accommodate the variability and flexibility of vowels, the question arises as to whether the measure is to that extent less appropriate for consonants. We do not know the answer at the present time, as we have not yet attempted to formulate consonantal rules equivalent to those for vowels given in Chapters VI, VIII, and IX. An eclectic and compromising approach might prove to be the best solution, as Cooper et al. (1962) put it:

> Some of the procedures that employ a combination of dictionary look-up and synthesis-by-rule may well prove to have important practical advantages; given a specific set of requirements, the highest quality in speech output for the lowest cost in instrumental complexity is more likely to be met by a hybrid system than by one limited to either [synthesis by] compilation or synthesis [by rule].

Secondly, we argued in Chapter IV that phonetic categories may best be represented by articulatory parameters, the argument being that if one requires a theory of phonology to provide a frame or a model of universal phonetics and to specify a set of possible human speech sounds, then the physiological model is the msot reasonable and workable one for the simple reason that the structure and the dynamics of the human vocal tract are largely uniform for speakers of all languages and that this simple fact explains why there is a limitation in the set of "possible" human speech sounds amidst an infinitely large acoustic range of sound.

There is, however, one irony involved in this line of argument. It is the belief that articulation and perception are rather distantly apart, and that, therefore, the phonetic specification in terms of physiological parameters is not directly relevant to speech perception. That is, if we view the communication process as a series of transformations of a sound event in the course of its travel from its source of generation (speaker's brain) to its ultimate destination (hearer's brain), and if we further assume that each transformation may introduce a channel noise and a non-linear distortion in such a way that there exists no one-to-one correlation between any two stages, then it follows that the closer the sound is towards the speaker, the less relevant it is to the hearer's perception. In this sense, the process is analogous to a syntactic transformation which may introduce an ambiguity, via deletion, substitution, inversion, etc., so that underlying sentences may not be inducible from surface structures. Likewise, what the hearer perceives (the final derivation) may not be inducible from what

or how the speaker articulates. This means that, since "we speak in order to be heard in order to be understood" (*Preliminaries*, p. 13), the hearer could not care less about what the speaker's vocal organs do as long as the output generated by them is understandable to him. Ventriloquy presents an extreme case in which different physiological processes may produce autidorily similar sounds. This kind of phenomenon leads us into looking for instances which clearly show that a description at a stage closer to the hearer is more relevant to, and throws more light on the nature of, speech perception. We saw in Chapter VII that in the case of syncretism between Rounding and the Front/Back dimension in vowels, the notion optimal opposition was definable in terms of acoustic distance more reasonably than in terms of articulatory distance. Going one step further, we may aruge that the change of [x] to [f] in English was due to perceptual similarity between the two sounds, though they are both acoustically and articulatorily different.

On the other hand, there are cases which seem to point in the opposite direction. For example, the nasalization of vowels in front of nasal consonants which later disappear (e.g., French), the palatalization of consonants before palatal vowels, etc., seem to be due to nothing but articulatory processes. That is, the nasalization of a vowel preceding a nasal consonant occur simply because of the so-called "coarticulation" phenomenon. At first, a nasal element in the vowel was undoubtedly a redundant feature, as it is in some words of present day English, e.g., *can't* [kæ̃nt], etc., but the switching of the redundant feature into the distinctive feature (i.e., the so-called "restructuring") must have occurred at one time, letting the now redundant feature, i.e., the nasal consonant, disappear. Probably, sound change is not a simple process, but a multiply complex one. For discussion, see Hockett (1965), Postal (1966).

There is another different but more important kind of phenomenon that indicates the important role of articulation in speech perception as well as in speech production. It is the so-called motor theory of speech perception (cf., Liberman 1957; Liberman et al. 1962; Denes 1965; Gulanov and Chistovich 1965). The theory, which we will not elaborate here, essentially hypothesizes that the points in the space of speech perception correspond to motor articulatory patterns and that the axes of the space correspond to the independent articulatory control parameters. In other words, the theory maintains that speech sounds are perceived by reference to the articulatory movements that produce them, and that "motor commands do stand in a very simple relation to the phonemes, and thus lend some further credence to the view that these commands provide a reference system in terms of which the complex acoustic signal is accurately and quickly identified" (Liberman et al., 1962). If this hypothesis (or theory) that categorial perception is made with reference to the corresponding articulatory categories is correct,[22] then a model of universal

---

[22]No doubt, some grave consequences ensue from this theory, the warrant for which is yet to be seen. For a critical review of the theory, see Lane (1965) and the bibliography cited there.

phonetics whose categories are articulatory parameters is the most rational and the most relevant one, as articulatory categories now refer directly to perceptual categories and thus only a minimum number of conversion rules is required.

Thirdly, it is mentioned here again that the model of universal phonetics proposed in this monograph is a crude one, and that further elaboration and explicit formalization are needed. For instance, we mentioned an undecided treatment of microcategories of Liquids (footnote 15). Rules of hierarchy of categories are perhaps to be reexamined seriously, especially the rules,

Alveolar → Dental + Alveolar-proper (Post-Alveolar)
Velar → (Pre-Velar) Velar-proper + Uvular
Laryngal → (Pharyngal) Glottal

Moreover, there has been no attempt to formalize and list the rules of General Redundancy. In this monograph, only a few examples, Front V → unRounded, Back V → Rounded, Palatal Stop → Affricate, etc., have been sketchily given. Undoubtedly, there are scores of others of this kind, e.g.,

$$(56)\quad \text{Vocalic} \rightarrow \left\{ \begin{array}{l} \text{Lingual} \\ \text{Egressive} \\ \text{Oral} \\ \text{Voiced} \end{array} \right\},$$

etc. (With regard to this, it is interesting to examine Ladefoged's (1965, p. 40) block diagram of a "finite state machine" generating the restrictions in combinations of manner of articulation and state of the glottis, which is an attempt to formalize some universal phonetic constraints.) Al of these, i.e., establishment of a given phonetic category, rules of hierachy, rules of General Redundancy, etc., are, needless to say, an empirical matter, and at least some of further formalization will have to wait until many presently unknown languages become known.

Finally, it should be said that we have not examined in this monograph the notion "phonological rule." This is as important an issue as any other in the phonology. That is, we must ask what types of operations are possible in phonology, just as we have asked what are the set of possible speech sounds. For example, is "transformational cycle" a part of universal phonetics, inherent in the nature of processes of speech sounds, or is it merely a convenient and perhaps economical device to describe a certain phonological phenomenon in English? We also must answer the question whether P rules are linearly ordered in empirical nature (e.g., Chomsky and Halle's "mutation"system), or whether they are unordered (e.g., Lamb's "realization" system). Even the use of such notation as parentheses, brackets, variables, must be carefully evaluated as to its significance in the phonological description.

This monograph has attempted to examine some issues in the theory of phonology, especially some aspects of the linguistic specification of speech. No doubt, the attempt is tinged by the color of the glasses through which the author is seeing. Yet, it is the author's humble hope that the color of his glasses has filtered out some insignificant distorted rays and allowed through some weak but important ones in a strengthened form, so that what may be seen now is a more refined and truer picture of speech sounds.[23]

---

[23]After the bulk of this monograph was written, I had the privilege of listening to Chomsky's lectures on English phonology at the Linguistic Institute (Summer 1966, UCLA), and of seeing, through his courtesy, a part of the manuscript of the forthcoming *Sound Pattern of English* (coauthored by Halle). As both lectures and the book are not yet in print, it is not possible to refer to specifically and discuss in detail some of the relevant issues dealt with in them, but suffice it to say that I was delighted and humbled to learn that Chomsky and Halle have been well aware of some of the issues that were raised and discussed in this monograph. Especially, a rather drastic reorganization of DF's with much reference to their empirical and physiological characters, and an extensive formalization of what has been called here Universally Restrictive and General Redundancies in terms of the "marking" convention show a certain similarity to some arguments presented in this monograph. Needless to say, particulars of the answers suggested in the two places differ from each other, but it is, at the moment, beyond the scope of this monograph to present a detailed discussion about their relative merits.

# BIBLIOGRAPHY

Abbreviations:

*Fodor and Katz* = J. A. Fodor and J. J. Katz ed., *The Structures of Language: Readings in the Philosophy of Language*, Prentice-Hall, Englewood Cliffs, New Jersey, 1964

*IJAL* = International Journal of American Linguistics

*JASA* = The Journal of the Acoustical Society of America

*JSHR* = Journal of Speech and Hearing Research


Arnold, G. F., P. Denes, A. C. Gimson, J. D. O'Connor, and L. M. Trim (1958), The synthesis of English vowels, *Language and Speech* 1.114-125

Austin, W. M. (1957), Criteria for phonetic similarity, *Language* 33.538-544

Chao, Y. R. (1954), Review of *Preliminaries*, *Romance Philology* 8

Chomsky, N. (1957), *Syntactic Structures*, Mouton, The Hague

----- (1962), Explanatory models in linguistics, in E. Nagel, P. Suppes, and A. Tarski ed., *Logic, Methodology, and Philosophy of Science*, Standford University Press, Palo Alto, California

----- (1964), Current issues in linguistic theory, in *Fodor and Katz*, pp. 50-118. (Originally, The logical basis of linguistic theory, in *Proceedings of the Ninth International Congress of Linguists*, Mouton, The Hague)

----- (1965), *Aspects of the Theory of Syntax*, MIT Press, Cambridge, Mass.

Chomsky, N. and M. Halle (1965), Some controversial questions in phonological theory, *Journal of Linguistics* 1.97-138

Chomsky, N. and G. A. Miller (1963), Introduction to the formal analysis of natural languages, in Luce, Bush, and Galanter ed., *Handbook of Mathematical Psychology*, Vol. II, John Wiley, New York, Chapter 11 (pp. 269-322)

Cooper, F. S. (1962), Speech synthesizers, in *Proceedings of the 4th Congress of Phonetic Sciences*, Mouton, The Hague, pp. 3-13

Cooper, F. S., A. M. Liberman, L. Lisker, and Jane N. Gaitenby (1962), Speech synthesis by rule, in *Proceedings of the Speech Communication Seminar*, Speech Transmission Laboratory, Royal Institute of Technology, Stockholm

Delattre, P. (1962), Some factors of vowel duration and their cross-linguistic validity, *JASA* 34.1141-1143

Denes, P. (1965), On the motor theory of speech perception, in *Proceedings of 5th Congress of Phonetic Sciences*, Basel, Switzerland, pp. 252-258

Essen, O. von (1964), Acoustic explanation of the sound shift [+] > [u] and [ɪ] > [i], in D. Abercrombie et al. ed., *In Honor of Daniel Jones*, Longmans, London, pp. 55-58

Fairbanks, G. and P. Grubb (1961), A psychophysical investigation of vowel formants, *JSHR* 4.203-219

Fant, G. (1956), On the predictability of formant levels and spectrum envelopes from formant frequencies, in *For Roman Jakobson*, Mouton, The Hague, pp. 109-120

----- (1960), *Acoustic Theory of Speech Production*, Mouton, The Hague

Fant, G., K. Fintoft, J. Liljencrants, B. Lindblom, and J. Martony (1963), Formant-amplitude measurements, *JASA* 35.1753-61

Firth, J. R. (1934), A Short outline of Tamil pronunciation, appendix to Ardens, *Grammar of Common Tamil*

Fischer-Jørgensen, Eli (1961), What can the new techniques of acoustic phonetics contribute to linguistics?, in S. Saporta ed., *Psycholinguistics*, Holt, Reinhart and Winston, New York, pp. 112-142. (Originally in *Proceedings of the 8th International Congress of Linguists*, Oslo, 1958, pp. 433-478)

----- (1964), Sound duration and place of articulation, *Zeitschrift für Phonetik* 17.175-207

Fromkin, Victoria A. (1965), *Some Phonetic Specifications of Linguistic Units: An Electromyographic Investigation*, Working Papers in Phonetics, No. 3, UCLA

----- (1966), Some requirements for a model of performance, *Working Papers in Phonetics*, No. 4, UCLA, pp. 19-39

Fry, D. B. (1958), Experiments in the perception of stress, *Language and Speech* 1.126-252

Groot, A. W. de (1931), Phonologie und Phonetik als Funktionswissehschaften, *Travaux du Cercle Linguistique de Prague*, 4.121

Gulanov, V. I. and L. A. Chistovich (1965), Relationship of motor theory to the general problem of speech recognition (Review), *Akusticheskii Zhurnal* 11.417-426. Translation in *Soviet Physics - Acoustics* 11.357-365 (1966)

Halle, M. (1957), In defense of the number two, in E. Pulgram ed., *Studies Presented to Joshua Whatmough*, Mouton, The Hague, pp. 65-72

----- (1959), *The Sound Pattern of Russian*, Mouton, The Hague

----- (1961), On the role of simplicity in linguistic descriptions, in *Proceedings of Symposia in Applied Mathematics*, Vol. XII (*Structure of Language and Its Mathematical Aspects*), American Mathematical Society, Providence, Rhode Island, pp. 89-94

----- (1964a), On the basis of phonology, in *Fodor and Katz*, pp. 324-333. (Originally, Questions in phonology, *Nuovo Cimento* 13.494-517, 1959)

----- (1964b), Phonology in generative grammar, in *Fodor and Katz*, pp. 334-352. (originally in *Word* 18.54-72, 1962)

Halle, M. and K. N. Stevens (1964), Speech recognition: a model and a program for research, in *Fodor and Katz*, pp. 604-612

Hecker, M. H. L. (1962), Studies of nasal consonants with an articulatory speech synthesizer, *JASA* 34.179-188

Heffner, R-M. S. (1937), Notes on the length of vowels, I, *American Speech* 12.128-134

----- (1940a), Notes on the length of vowels, II, *American Speech* 15.74-79

----- (1940b), Notes on the length of vowels, III, *American Speech* 15.377-380

----- (1941), Notes of the length of vowels, IV, *American Speech* 16.204-207

----- (1942), Notes on the length of vowels, V, *American Speech* 17.42-48

Hill, A. A. (to appear), Non-grammatical prerequisites to phonological statement

Hjelmslev, L. (1943), *Omkring Sprogteoriens Grundloeggelse*, Ejnar Munksgaard, Copenhagen. English translation by F. J. Whitefield, *Prolegomena to a Theory of Language*, University of Wisconsin Press, Madison, Wisconsin, 1961

Hockett, C. F. (1965), Sound change, *Language* 41.185-204

Holbrook, A. and G. Fairganks (1962), Diphthong formants and their movements, *JSHR* 5.38-58

Holmes, J. N., I. G. Mattingly, and J. N. Shearme (1964), Speech synthesis by rule, *Language and Speech* 7.127-143

House, A. S. (1961), On vowel duration in English, *JASA* 33.1174-1178

Householder, F. W. (1956), Unreleased /ptk/ in American English, in *For Roman Jakobson*, Mouton, The Hague, pp. 235-244

----- (1965), On some recent claims in phonological theory, *Journal of Linguistics* 1.13-34

----- (1966), Phonological theory: a brief comment, *Journal of Linguistics* 2.99-100

International Phonetic Association (1949), *The Principles of IPA*, University College, London

Ivić, P. (1965), Roman Jakobson and the growth of phonology, *Linguistics* 18.35-78

Jakobson, R. (1962a), Observations sur la classement phonologique des consonnes, in *Selected Writings I*, Mouton, The Hague, pp. 272-279. (Originally in *Proceedings of the Third International Congress of Phonetic Sciences*, Ghent, 1939)

----- (1962b), Kindersprache, Aphasie und allgemeine Lautegesetze, in *Selected Writings I*, Mouton, The Hague, pp. 328-401. (Originally in *Uppsala Universitets arsskrift*, Uppsala, 1941, pp. 1-83)

----- (1962c), Typological studies and their contribution to historical comparative linguistics, in *Selected Writings I*, Mouton, The Hague, pp. 523-532. (Originally in *Proceedings of the Eighth International Congress of Linguists*, Oslo, 1958)

----- (1962d), Why "mama" and "papa"?, in *Selected Writings I*, Mouton, The Hague, pp. 538-545. (Originally in *Perspectives in Psychological Theory*, New York, 1960)

Jakobson, R., G. Fant, and M. Halle (1951), *Preliminaries to Speech Analysis*, MIT Press, Cambridge, Mass.

Jakobson, R., and M. Halle (1956), *Fundamentals of Language* Mouton, The Hague

----- (1964), Tenseness and laxness, in D. Abercrombie et al. ed., *In Honor of Daniel Jones*, Longmans, London, pp. 96-101. Reprinted in *Preliminaries*, 3rd printing, pp. 57-61

Jones, D. (1960), *An Outline of English Phonetics*, 9th edition, Dutton, New York

----- (1961), The Phonology of English, Edinburgh Phonetics Diploma Course

Joos, M. (1942), A phonological dilemma in Canadian English, *Language* 18.220-223

----- (1950), Description of language design, *JASA* 22.701-708. (Reprinted in Joos ed., *Readings in Linguistics*, American Council of Learned Societies, New York, 1958, pp. 349-356

----- (1948), *Acoustic Phonetics*, Language Monograph No. 23, Supplement to *Language* Vol. 24, No. 2

Katz, J. J. and P. M. Postal (1964), *Integrated Theory of Linguistic Description*, MIT Press, Cambridge, Mass.

Kelley, K. (1966), Some comments on n-ary feature systems, Summer 1966 Meeting of Linguistic Society of America, UCLA

Kelly, J. L. and L. J. Gerstman (1961), An artificial talker driven from phonemic input, *JASA* 33.835 (A)

Kim, C-W. (1965), Rules of vowel duration in American English, Winter 1965 Meeting of Linguistic Society of America, Chicago, Illinois

----- (1966), On the autonomy of the tensity feature in stop classification, to appear in *Word*

----- (to appear), Some phonological rules in Korean

Kozhevnikov, V. A. and L. A. Chistovich (1965), *Rech: Artikulyatsia i Vosprivatiye*, Moscow-Leningrad. English translation (by Joint Publication Research Service, U. S. Department of Commerce): *Speech: Articulation and Perception* (1966)

Ladefoged, P. (1964), *A Phonetic Study of African Languages*, Cambridge University Press

----- (1965), The nature of general phonetic theories, *Georgetown University Monograph on Languages and Linguistics*, No. 18, pp. 27-42, Georgetown University Press, Washington, D. C.

----- (1966), An attack on the number two, *Working Papers in Phonetics*, No. 4, UCLA, pp. 7-9

----- (forthcoming), *Linguistic Phonetics*

Ladefoged, P. and D. E. Broadbent (1957), Information conveyed by vowels, *JASA* 29.98-104

Ladefoged, P. and C-W. Kim (1965), Human, replica, and computer-generated formants, *Working Papers in Phonetics*, No. 2, UCLA, pp. 18-26

Lamb, S. (1964), On alternation, transformation, realization, and stratification, *Georgetown University Monograph on Languages and Linguistics*, No. 17, Goergetown University Press, Washington, D.C. pp. 105-122

Lane, H. (1965), The motor theory of speech perception: a critical review, *Psychological Review*, 72.275-309

Lees, R. B. (1960), *The Grammar of English Nominalizations*, *IJAL* Vol. 26, No. 3, Part II. Publication 12 in Anthropology, Folklore, and Linguistics, Indiana University, Bloomington, Indiana

Lehiste, Ilse (1964), *Acoustical Characteristics of Selected English Consonants*, *IJAL* Vol. 30, No. 3, Part IV. Publication 34 in Anthropology, Folklore, and Linguistics, Indiana University, Bloomington, Indiana.

Lehiste, Ilse, and G. E. Peterson (1959), Vowel amplitude and phonemic stress in American English, *JASA* 31.428-435

----- (1961), Transitions, glides, and diphthongs, *JASA* 33.268-277

Lehmann, W. and R-M. S. Heffner (1943), Notes on the length of vowels, VI, *American Speech* 18.208-215

Liberman, A. M. (1957), Some results of research on speech perception, *JASA* 29.117-123. (Reprinted in S. Saporta ed., *Psycholinguistics*, Holt, Reinhart and Winston, New York, pp. 142-153)

Liberman, A. M., F. S. Cooper, K. S. Harris, and P. F. MacNeilege (1962), A motor theory of speech perception, in *Proceedings of the Speech Communication Seminar*, Speech Trnasmission Laboratory, Royal Institute of TEchnology, Stockholm

Liberman, A. M., F. Ingemann, L. Lisker, P. Delattre, and F. S. Cooper (1959), Minimal rules for synthesizing speech, *JASA* 31.1490-1499

Lightner, T. M. (1963), A note on the formation of phonological rules, Quarterly Progress Report, No. 68, Research Laboratory of Electronics, MIT, pp. 187-189

Lindblom, B. (1963), Spectrographic study of vowel reduction, *JASA* 35.1173-81

----- (1964), Articulatory and acoustic studies of human speech production, *Quarterly Progress and Status Report*, Speech Transmission Laboratory Royal Institute of Technology, Stockholm, December 1964

Lisker, L. and A. S. Abramson (1964), A cross-language study of voicing in initial stops: acoustical measurements, *Word* 20.384-422

Malmberg, B. (1963), *Structural Linguistics and Human Communication*, Academic Press, New York

Martinet, A. (1955), *Économie des Changements Phonetiques*, A. Francke Berne

Meyer, E. S. (1903), *Englische Lautdauer*, Uppsala and Leibzig

Öhman, S. (1966), Coarticulation in VCV utterances: spectrographic measurements, *JASA* 39.151-168

Peterson, G. E. (1961), Parameters of vowel quality, *JSHR* 4.10-29

Peterson, G. E. and H. L. Barney (1952), Control methods used in a study of the vowels, *JASA* 24.175-184

Peterson, G. E. and F. Harary (1961), Foundations in phonemic theory in *Proceedings of Symposia in Applied Mathematics*, Vol. XII (*Structure of Language and Its Mathematical Aspects*), American Mathematical Society, Providence, Rhode Island, pp. 139-165

Peterson, G. E. and Ilse Lehiste (1960), Duration of syllable nuclei in English, *JASA* 32.693-703

Peterson, G. E. and J. E. Shoup (1966), A physiological theory of phonetics, *JSHR* 9.5-67

Pike, K. L. (1947), *Phonemics*, University of Michigan Press, Ann Arbor, Michigan

Postal, P. M. (1964), Boas and the development of phonology: comments based on Iroquoian, *IJAL* 30.269-280

----- (1965), On the mentalistic character of so-called 'sound change,' in Postal, *Two Studies in Phonology*, forthcoming

Potter, R. K. and J. C. Steinberg (1950), Toward the specification of speech, *JASA* 22.807-820

Rosen, G. (1958), A dynamic analog speech synthesizer, *JASA* 30.204-209

Rositzke, H. (1939), Vowel-length in General American speech, *Language* 15.99-109

Saussure, F. de (1915), *Cours de Linguistique Générale*, Paris. English translation by W. Baskin, *Course in General Linguistics*, Philosophical Library, New York, 1959

Stanley, R. (1966), Redundancy rules in phonology, to appear in *Language*

Stevens, K. N. and A. S. House (1955), Development of a quantitative description of vowel articulation, *JASA* 27.484-493

Stockwell, R. P. (1964), Historical realism in English phonology, Winter 1964 Meeting of Linguistic Society of America, New York

----- (1966), Problems in the interpretation of the Great English Vowel Shift, The 5th Texas Conference on Phonology

Troubetzkoy, N. S. (1949), *Principes de Phonologie*, French translation of *Grundzüge der Phonologie* by J. Cantineau, Klincksieck, Paris

Vanderslice, R. (to appear), Computed transfer functions for four vocal tract replica shapes

Wang, W. S-Y, and C. J. Fillmore (1961), Intrinsic cues and consonant perception, *JSHR* 4.130-136

Ward, Ida C. (1929), *The Phonetics of English*, Heffer and Sons, Cambridge, England

Westermann, D. and Margaret A. Bryan (1952), *Languages of West Africa*, Handbook of African Languages, Part 2, Oxford University Press, London

Zimmerman, S. A. and S. M. Sapon (1958), Note on vowel-duration seen cross-linguistically, *JASA* 30.152-153