I THE CONCEPT OF THE DISTINCTIVE FEATURE

1.1 RESOLVING SPEECH INTO ULTIMATE UNITS.

In a typical test of the intelligibility of speech, an English speaking announcer pronounces isolated root words (bill, put, fig, etc.), and an English speaking listener endeavors to recognize them correctly. For the listener this situation is in one sense simpler than normal speech communication because the word samples with which he deals cannot be broken up into shorter meaningful entities and are not grouped into higher units. Thus the division of sentences into words and of words into their grammatical components does not concern this listener. Nor need he account for the interrelation of words within a sentence and of various grammatical components within a complex word (ex-port-s, im-port-ed, re-port-ing, mid-night).

In another sense, however, this test is more complicated than normal speech communication. Neither the context nor the situation aids the listener in the task of discrimination. If the word bill were to appear in the sequence one dollar bill or as a single word said to a waiter after a meal, the listener would be able to predict its appearance. In such a situation, the sounds which compose this word are redundant to a high degree, since they "could have been inferred a priori"(1). If, however, the word is deprived of any prompting context, either verbal or non-verbal, it can be recognized by the listener only through its sound-shape. Consequently, in this situation the speech sounds convey the maximum amount of information.

The question arises: how many significant units, i.e., units relevant for the discrimination of the samples, do the sound-shapes of the samples contain? Upon perceiving syllables such as bill and pull, the listener recognizes them as two different words distinguishable by their initial part /bi/ and /pu/ respectively. This distinctive fraction, however, may be decomposed in turn. The listener, and any member of the English speech community, has in his vocabulary words such as pill and bull. On the one hand, identical means are employed for distinguishing bill from pill and bull from pull. On the other hand, the distinction between bill and bull is the same as that between pill and pull. Thus to distinguish between bill and pull a double operation is necessary. The fraction /bi/ in bill proves capable of being split into two segments /b/ and /i/, the first exemplified by the pair bill - pill and the second by bill - bull.

Each of the two segments derived serves to distinguish the word bill from a whole series of vocables, all other things being equal.* For each of them a set of other segments can be substituted. This substitution of one segment by others is called commutation.

* Henceforth we shall use the more condensed Latin equivalent of this formula: ceteris paribus.

We can list one whole commutation set. Commuting the first segment we obtain the series bill - pill - vill - fill - mill - dill - till - thill - sill - nil - gill /gil/ - kill - gill /ǯil/ - chill - hill - ill - rill - will. A closer examination of such a series permits certain inferences.

For some pairs of words in this set the discriminatory minimum is identical; hence one is warranted in saying that bill is to pill, as vill is to fill, or dill to till, or gill to kill, etc. or, for the sake of a more graphic presentation: bill:pill ≃ vill:fill ≃ dill:till ≃ gill:kill etc.

By the same token,
1) bill : vill ≃ pill : fill ≃ till : sill etc.
2) bill : mill ≃ dill : nil etc.
3) bill : dill ≃ pill : till ≃ fill : sill ≃ mill : nil etc.

A distinction is called minimal if it cannot be resolved into further distinctions which are used to differentiate words in a given language. We owe this term to Daniel Jones, from whom we also borrow the following definition*: "Wider differences may be termed duple, triple, etc., according to the number of minimal distinctions of which the total difference is composed. Duple distinctions are the result of two minimal distinctions." (2)

The distinctions between bill and pill, or bill and vill or bill and dill are minimal distinctions since they cannot be resolved into simpler discriminations, which are, in turn, capable of differentiating English words. On the other hand, the relation of bill to till is a duple distinction, composed of two minimal distinctions: 1) bill - dill (which is equivalent to the distinction pill - till) and 2) bill - pill (equivalent of dill - till). The relation of bill to sill is a triple distinction: in addition to the two minimal distinctions cited, it includes a third one: bill - vill (equivalent to pill - fill and to till - sill).

The discrimination between the words bill and fell implies a duple distinction in their initial segments (/b/ - /f/), and a minimal one in the middle segments (/i/ - /e/). To discriminate between words such as bit and said, we need a triple distinction in their first segment and one minimal distinction in each of the two others.

Without further examples, it becomes clear that the listener of a speech sample is faced with a series of two-choice selections. To identify the message bill, he must decide for the non-vocalic inception against the vocalic and for the consonantal against the non-consonantal. By this double operation, vowels, liquids and glides are eliminated because if the word had begun with a vowel,

* We, alone, are responsible for the way in which these concepts are hereafter applied to the empirical material.

the inception would have been identified as vocalic and non-consonantal; if with a liquid, as both vocalic and consonantal; and if with a glide, as neither vocalic nor consonantal. (For the interpretation of these distinctions see Sec. 2.2).

The next decision to be made is between bill and gill /gil/ - diffuse or compact (see 2.41), between bill and dill - grave or acute (see 2.42), and finally, between bill and mill - non-nasalized or nasalized (see 2.44). A decision in favor of the latter of the two alternatives would leave no further selections, since /m/ is the only combination of grave and nasal in English. But the opposite choice being made, there inevitably follows the selection between bill and pill - weak or strong (in more general terms, lax or tense: see 2.43), and, finally, the selection between bill and vill - stop or constrictive (in more general terms, interrupted or continuant: see 2.311). An analogous sequence of operations treats the two succeeding segments of the sample /i/ and /l/. The set of selections to be made is, however, more restricted than for the initial segment. For example, when a sequence begins with a stop, as bill does, the option for vocalic is obligatory, since in English the initial stop may be followed only by vowels or liquids.

Any minimal distinction carried by the message confronts the listener with a two-choice situation. Within a given language each of these oppositions has a specific property which differentiates it from all the others. The listener is obliged to choose either between two polar qualities of the same category, such as grave vs. acute, compact vs. diffuse, or between the presence and absence of a certain quality, such as voiced vs. unvoiced, nasalized vs. non-nasalized, sharpened vs. non-sharpened (plain). The choice between the two opposites may be termed distinctive feature. The distinctive features are the ultimate distinctive entities of language since no one of them can be broken down into smaller linguistic units. The distinctive features combined into one simultaneous or, as Twaddell aptly suggests, concurrent bundle form a phoneme.

For example, the word bill is comprised of three consecutive bundles of distinctive features: the phonemes /b/, /i/ and /l/. The first segment of the word bill is the phoneme /b/ consisting of the following features: 1) non-vocalic, 2) consonantal, 3) diffuse, 4) grave, 5) non-nasalized (oral), 6) lax, 7) interrupted. Since in English 7) implies both 1) and 2), the latter two features are redundant. Similarly 3) is redundant as it is implied by 4).

A speech message carries information in two dimensions. On the one hand, distinctive features are superposed upon each other, i.e., act concurrently (lumped into phonemes), and, on the other, they succeed each other in a time series. Of these two arrangements the superposition is the primary because it can function without the sequence; the sequence is the secondary since it implies the primary. For example, the French words où /u/ "where", eu /y/ "had" (participle), y /i/ "there", eau /o/ "water", oeufs /ø/ "eggs", et /e/ "and", aie /e/ "have!", un /œ̃/ "one", an /ã/ "year", etc., each contains a single phoneme.

The difference between the distinctive features of contiguous bundles permits the division of a sequence into phonemes. This difference may be either complete, as between the last two phonemes /i/ and /ŋ/ in the word wing (which have no distinctive features in common) or partial, as between the last two phonemes of the word apt - /p/ and /t/ all of whose distinctive features are the same except one: /p/ is grave and /t/ is acute.

This suprasegmental extension of certain features such as interruptedness, diffuseness or non-nasality is selective: cf. such sequences as asp (continuant and interrupted), act (compact and diffuse) and ant (nasal and oral). On the other hand, strong (tense) and weak (lax) consonants cannot follow each other within a simple English word: cf. nabs/nabz/, nabbed/nabd/, and naps /naps/, napped/napt/. That is to say, in consonant sequences the tenseness and laxness features are suprasegmental.

Any one language code has a finite set of distinctive features and a finite set of rules for grouping them into phonemes and also for grouping the latter into sequences; this multiple set is termed phonemic pattern.

Any bundle of features (phoneme) used in a speech message at a given place in a given sequence is a selection from among a set of commutable bundles. Thus by commuting one feature in the first phoneme of the sequence pat we obtain a series bat - fat - mat - tat - cat. Any given sequence of phonemes is a selection from among a set of permutable sequences: e.g. pat - apt - tap. However, /tp'a/ not only does not, but could not exist as an English word, for it has an initial stop sequence and a single final vowel under stress, both of which are inadmissible according to the coding rules of contemporary English.

1.2 INVARIANCE AND REDUNDANT VARIATIONS

The consonants are quite different in the English coo and key or in the French coup and qui. In both languages a more backward (velar) articulation is used before /u/ and a more forward (palatal) articulation before /i/. The formants of the consonant are closely adapted to those of the following vowels, so that the frequency spectrum of /k/ before /u/ has a lower center of area and is closer to that of /p/ than is the case before /i/, where it has a higher center of area and is closer to that of /t/. Both in English and French, /p/ and /t/ are separate phonemes opposed to each other as grave and acute, whereas the two varieties of /k/ represent but a single phoneme. This seeming discrepancy is due to the fact that the opposition of /p/ and /t/ is autonomous, i.e. both /p/ and /t/ occur in identical contexts (pool - tool; pea - tea), while the difference between the two k-sounds is induced by the following vowel: it is a contextual variation. The retracted articulation and the low frequencies of one of these k-sounds and the more advanced articulation and high frequencies of the other are not distinctive but redundant features, since the distinction is carried by the subsequent vowels. In Roumanian, both k-sounds in question occur in one and the same context (e.g. before /u/: cu "with", with a backward articulation, and chiu "cry", with a more forward articulation) and, therefore, they represent two different phonemes.

In the same way, the difference between the so-called "clear" and "dark" varieties of the English /l/ is redundant: the "clear" variant indicates that a vowel follows and the "dark" variant that no vowel follows; thus in lull, the initial /l/ is "clear" and the final, "dark". In Polish these two sounds may appear in one and the same context and form a distinctive opposition: cf. the "clear" /l/ in laska "cane" and the sound close to the English "dark" variety in ɣaska "grace."

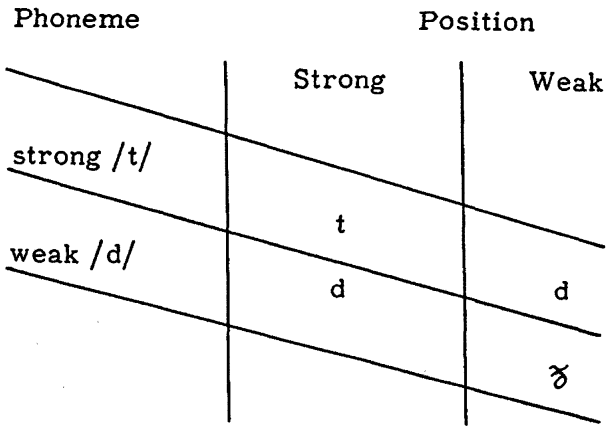
The relation between tart and dart, try and dry, and bet and bed represents in English one and the same minimal distinction regardless of the perceptible articulatory and acoustical difference between the three t-sounds cited. The invariant is the opposition of strength and weakness (for more precise data see 2.43). In English a regular concomitant factor of this opposition is the voicelessness of the strong consonants and the voicing of the weak ones. But this redundant feature may disappear occasionally; cf. the voiceless variants of /b d z/ observed by English phoneticians.

It is important to note that gradations in strength serve no distinctive purpose: they depend entirely upon the context. For instance, the heavy aspiration of the initial strong /p t k/ before a stressed vowel as in tart and, conversely, the lack of aspiration before other phonemes as in try are only contextual variants which cannot impede the identification of any /p t k/ as strong in contradistinction to the weak /b d g/.

Danish is another language that exhibits the opposition of strong and weak consonants. This opposition is implemented in different ways depending upon the position of the consonant in a word. Two positions are discernible in the Danish word - strong and weak. In monosyllabic words the strong position for a consonant is at the beginning of the syllable and the weak position, at its end. In strong position the strong stops are normally produced with a heavy aspiration, while their weak opposites appear as weak stops (differing from the English /b d g/ through voicelessness); e.g. tag "roof" - dag "day". In weak position the strong /t/ is weakened to the level of /d/, while its weak opposite is further weakened from /d/ to the weakest level /ð/ resembling somewhat the consonant of the English the; for example: hat [had] "hat" - had [hað] "hate". Consequently, the opposition of the strong and weak phoneme remains invariant in both positions; at the same time there is a redundant shift of both opposites induced by the weak position, which indicates that neither a stressed nor a long vowel follows. Although the weak phoneme in strong position and the strong phoneme in weak position overlap phonetically, in the strictly relational terms of distinctive features there is no overlapping:

"Two patterns are identical if their relational structure can be put into a one-to-one correspondence, so that to each term of the one there corresponds a term of the other" (3).

Hence, an automatic detector designed to distinguish between the two positions and between the two polar terms within each of them would unerringly "recognize" both the strong and weak phoneme:



The instances cited show how the invariance of the minimal distinctions can be separated from the redundant features that are conditioned by the adjacent phonemes in the sequence.

The sequential arrangement of distinctive features does not generate the only type of redundancies. Another less analyzed though very important class of redundancies is conditioned by the superposition of simultaneous distinctive features. There are languages in which the velar [k] is in complementary distribution with the corresponding palatal stop or with a still more advanced prepalatal affricate (pronounced as in the English chew). For instance, the velar sound occurs only before back vowels and the palatal (or prepalatal) sound only before the front vowels. In such cases the former and the latter are considered two contextual variants representing a single phoneme. By the same reasoning, if in French we find the velar stop /k/, the palatal nasal /ɲ/ (as in ligne) and the prepalatal constrictive /ʃ/ (as in chauffeur), we must consider the difference between this velar, palatal and prepalatal articulation as entirely redundant, for this difference is supplementary to other, autonomous distinctions. All of these consonants are opposed to those produced in the front part of the mouth as compact vs. diffuse (see 2.41). When the features of interruptedness (stop), nasalization and continuancy are superposed upon the compactness feature, they are accompanied, in the French consonants, by the redundant features of velarity, palatality, and prepalatality respectively. Thus the French /p b/ and /t d/ bear the same relation to /k g/, as /f v/ and /s z/ do to /ʃ ʒ/, and as /m/ and /n/ to /ɲ/.

The redundant character of the velar and prepalatal feature of the English compact consonants can be demonstrated in a similar manner. In Czech or Slovak, however, the analogous difference between velars and palatals (including the prepalatals in the latter class) is distinctive, since these languages have velars and palatals, ceteris paribus. The velar stop /k/ is opposed to the palatal stop /c/ and the velar constrictive /x/ to the (pre) palatal /ʃ/. Consequently, in these languages the opposition grave vs. acute characterizes not only the relation of labials to dentals but also that of velars to palatals: /k/ is to /c/ as /p/ is to /t/.

The multiplicity of distinctions traditionally accepted in the analysis of speech could be radically diminished were we to eliminate the redundancies linked to the relevant opposition of vowels and consonants. For example, it can be shown that the relation of the close to the open vowels, on the one hand, and that of the labials and dental consonants to consonants produced against the hard or soft palate, on the other, are all implementations of a single opposition: diffuse vs. compact (see 2.41); provided that the numerous redundancies contingent upon the fundamental difference between the vocalic and consonantal feature be eliminated. In their turn the relations between the back and front vowels, and between the labial and dental consonants pertain to a common opposition grave vs. acute (see 2.421).

While the relational structure of these features, which are common to consonants and vowels, manifests a definite isomorphism (one-to-one correspondence), the variations are in complementary distribution. 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.

By successively eliminating all redundant data (which do not convey new information) the analysis of language into distinctive features overcomes the "non-uniqueness of phonemic solutions" (4). This pluralism, pointed out by Y. R. Chao, interfered with the analysis as long as the phoneme remained the ultimate operational unit and was not broken down into its constituents. The present approach establishes a criterion of the simplicity of a given solution, for when two solutions differ, one of them is usually less concise than the other by retaining more redundancy.

The principle of complementary distribution, which has proven most efficacious in speech analysis, opens many new possibilities when its ultimate logical implications are made explicit. Thus if certain phonemic distinctions possess a common denominator and are never observed to co-exist within one language, then they may be interpreted as mere variants of a single opposition. Furthermore, the question can be raised whether the selection of a given variant in a certain language is not connected with some other features proper to the same linguistic pattern.

In this way the inquiry succeeds in reducing the list of distinctive features ascertained in the languages of the world. Trubetzkoy (5) distinguishes the following three consonantal oppositions: first, the opposition of strong and weak consonants, the former characterized by a stronger resistance to the air flow and stronger pressure; second, the opposition of a stronger and weaker resistance alone, without accompanying pressure differences; third, the opposition of aspirated and non-aspirated. Since, however, never more than one of these three oppositions has been encountered functioning autonomously within any one language, all three should be regarded as mere variants of a single opposition. Moreover, this variation is apparently redundant because it depends upon certain other consonantal features present in the same pattern (see 2.43).

The extremely limited set of distinctive features underlying a language, the restrictions on their actual combinations into bundles and sequences and, finally, a high amount of redundancy, lighten the load imposed upon the participants of the speech event.

In the hierarchy of the sound features the distinctive features are of paramount importance. However, the role of the redundancies must not be underestimated. Circumstances may even cause them to substitute for the distinctive features. In Russian the distinction between the palatalized and non-palatalized consonants plays a significant part in differentiating words. Palatalization produces a slight rise of the formants (see 2.423). The phoneme /i/ 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 [s̡] which serves as the means of discriminating between the syllables [sɨ] and [s̡i]. But when a mason telephoned an engineer saying that the walls [sɨr̡'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 gray" [s̡ir̡'ejut], then the worker repeated the word with particular emphasis on the [ɨ], and through this redundant feature the listener made the right choice. In S. S. Stevens' formulation:

> ".....the fact of redundancy increases the reliability of speech communication and makes it resistant to many types of distortion. By limiting the number of discriminations required of the listener and by assisting his choice through the redundant coding of information, we make talking to one another a reasonably satisfactory business" (6).

1.3 IDENTIFICATION OF DISTINCTIVE FEATURES

Any distinctive feature is normally recognized by the receiver if it belongs to the code common to him and to the sender, is accurately transmitted and has reached the receiver.

Suppose that both participants of the speech event use the same kind of standard English and that the listener has received the vocables gip, gib and gid, which are unfamiliar to him, as to many other English speakers. He does not know that gip means "to clean (fish)", gib, "castrated tom-cat", and gid, "an animal disease." Yet the information he obtains from these three samples is that they may be English words, since none of the features and feature combinations contained in them contradict the English code. Moreover, the three samples convey the information that, if they are words, then each of them has a different meaning, for there is a duple distinction between gip and gid and two different minimal distinctions separate gib from gip and gid. Were the English-speaking listener to hear the following highly improbable sentence: "The gib with the gid shall not gip it", he would know from his knowledge of the rules of the English code, that /gib/ ≠ /gip/ ≠ /gid/. Were the samples to

be transmitted in a German speech circuit, gib and gip would be identified as two optional variants of what is probably the same word, since in German the distinction of /b/ and /p/ is cancelled at the end of the word. The same identification would be made in a Finnish speech circuit, since in the Finnish code the difference between the sounds [b] and [p] has no distinctive value.

Information Theory uses a sequence of binary selections as the most reasonable basis for the analysis of the various communication processes (7). It is an operational device imposed by the investigator upon the subject matter for pragmatic reasons. In the special case of speech, however, such a set of binary selections is inherent in the communication process itself as a constraint imposed by the code on the participants in the speech event, who could be spoken of as the encoder and the decoder.

This follows from the fact that the sole information carried by the distinctive feature is its distinctiveness. The listener distinguishes the word /gib/ from /gid/ by one feature: the grave character of /b/ as opposed, ceteris paribus, to the acute character of /d/. The same word /gib/ is distinguished from /gip/ by a different feature: the weak character of /b/ as opposed to the strong character of /p/. In these two examples, pairs of words display one minimal distinction in corresponding segments, ceteris paribus. Other pairs of words can display a higher number of minimal distinctions either in one segment or in more than one segment. When we review these minimal distinctions used to discriminate between these pairs of words, we find only two possibilities: a) occurrences of the same opposition (gib:gid $\simeq$ fat:sat), and b) each of the two distinctions has a specific property of its own (gib:gid $\neq$ gib:gip).

To be sure, articulatorily, physically, and perceptually, there exists a continuous range of degrees from whisper to full voicing, but only two polar points - the presence and the absence of voice - are picked out as distinctive features. There is a continuous variation in the shape of the lips from a close rounding to spreading and in the corresponding acoustic effects; but the linguistic opposition flat vs. plain (e.g. German Küste "shore" - Kiste "box") is a linguistic assignment of distinctive value to two distant lip positions and to their contrastive acoustical effects (see 2.422). In general, no language possesses more than one minimal distinction based on the size of the lip orifice.

The dichotomous scale is the pivotal principle of the linguistic structure. The code imposes it upon the sound.

Only one phonemic relation presents a somewhat different aspect. This is the relation between vowels with a compact and those with a diffuse spectrum (open and close, in articulatory terms). In a language such as Turkish, the vowels are grouped into compact and diffuse pairs, other things being equal: /kes/ "cut!" is to /kis/ "tumor" as /kol/ "arm" is to /kul/ "slave". But a language such as Hungarian distinguishes, ceteris paribus, three degrees of compactness. Cf. /tår/ "bald" with an open rounded back vowel - /tor/ "feast" with the corresponding mid vowel - /tur/ "rakes up" with the close vowel, and, similarly, in the unrounded front series, /hæ/ "take it!" - /he/ "don't!" -

/ni/ "look!"*. The minimal distinction remains the same as in Turkish: /o/ and /e/ are opposed to /u/ and /i/ as relatively compact to relatively diffuse. In Hungarian, however, the same opposition (relatively compact vs. relatively diffuse) reappears in such pairs as /tår/ - /tor/ and /næ/ - /ne/; that is to say that /a/ : /o/≃/o/ : /u/. In this "phonemic proportion" /o/ (or /e/) functions as the "mean proportional." It carries two opposite features - compactness vs. the diffuse /u/ (or /i/) and diffuseness vs. the compact /a/ (or /æ/). (On ways of dealing with such bi-polar phonemes in analytical procedures see 2.414).

No other inherent phonemic oppositions exhibit such bi-polar complexes. There are, however, conjugate distinctions prone to merge, e.g., the pairs continuant - interrupted, and strident - mellow. When two conjugate oppositions merge the resulting opposition is maximally clearcut, optimal. Thus the optimal continuant consonants are strident; the optimal interrupted, mellow. A similar relation links the oppositions grave vs. acute and flat vs. plain. The optimal grave vowels are flat; the optimal acute are plain (concerning the reverse combinations of such features as interrupted strident, continuant mellow, flat acute or plain grave and ways of treating them in the analytical procedures see 2.324 and 2.4236).

It is the dichotomous scale of the distinctive features, in particular, and the whole patterning of the linguistic code, in general, that to a large extent determines our perception of the speech sounds. We perceive them not as mere sounds but specifically as speech components. More than this, the way we perceive them is determined by the phonemic pattern most familiar to us. Therefore, a monolingual Slovak identifies the rounded front vowel /ø/ of the French word jeu as /e/, since the only distinctive opposition in his mother tongue is acute (front) vs. grave (back) and not flat (rounded) vs. plain (unrounded). A monolingual Russian, on the contrary, perceives the same French vowel as /o/ because his native tongue possesses only the one of the two oppositions in question, namely, flat vs. plain. Even as expert a linguist as the Frenchman Meillet perceived the Russian sharpened /t̡/ as a sequence of /t/ and non-syllabic /i̯/ and not as a consonant with simultaneous, superposed sharpening (palatalization), for Meillet's judgment was based on his native French, which lacks the sharpening feature but possesses the non-syllabic /i̯/. Hence it is only to be expected that when nonsense syllables are used in intelligibility tests (traditionally called "articulation tests") the results depend upon whether or not these sequences are patterned in accordance with the rules of combination of the given linguistic code.

Interference by the language pattern affects even our responses to non-speech sounds. Knocks produced at even intervals, with every third louder, are per-

*The examples, which we have from John Lotz, belong to a colloquial variety of Standard Hungarian.

ceived 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.

If on the aural level too, speech analysis were to be conducted in terms of the binary phonemic oppositions, the task would be substantially facilitated and could perhaps supply the most instructive correlates of the distinctive features.

As to the acoustic investigation of the speech sounds, its whole development has been toward an ever more selective portrayal of the sound stimuli. Both the instruments used and the interpretation of the data recorded by them are progressively more oriented toward the extraction of the pertinent items. Investigators have come to see that the wave traces contain too much information and that means must be provided for selecting the essential information(8). As soon as it is realized that the proper criterion of selection is the linguistic relevance (expressed in binary terms), the acoustic problems of the speech sounds find a far more determinate solution. Correspondingly the articulatory stage of speech must be defined in terms of the means utilized to obtain any pair of contrastive effects. For example, as far as language uses an autonomous distinctive opposition of gravity and acuteness, we examine the acoustical correlates of the linguistic values in question and the articulatory prerequisites of these stimuli.

In short, for the study of speech sounds on any level whatsoever their linguistic function is decisive.

The interesting attempt, suggested by B. Bloch to decipher the phonemic pattern of a language from a mere examination of a sufficient number of recorded utterances (9) is onerous but feasible. It implies, however, two strictly linguistic assumptions. The first was formulated by Wiener (3): "In the problem of decoding, the most important information which we can possess is the knowledge that the message we are reading is not gibberish." This corresponds to the knowledge obtained by any listener upon reaching the so-called threshold of perceptibility, when the sounds heard begin to be perceived as speech sounds (10). Since it is speech, the second assumption follows as a corollary of the first: in its sound shape any language operates with discrete and polar distinctive features, and this polarity enables us to detect any feature functioning ceteris paribus.

Obviously such a task of deciphering becomes more difficult in the frequent cases called "switching code" by communication engineers (11) or "coexistent phonemic systems" by linguists (12). The Russian aristocracy of the last

century with its bi-lingual speech - switching continually from Russian to French and vice versa even within a single sentence - provides a striking illustration. Another example is set by some Mohammedan cultural languages with their Arabic interpolations. Two styles of the same language may have divergent codes and be deliberately interlinked within one utterance or even one sentence. For instance, urban colloquial Czech is a whimsical oscillation between the literary language and vulgar Czech, each of them displaying its own phonemic pattern.

The dichotomous scale is superimposed by language upon the sound matter much in the same way as the diatonic scale is superimposed upon the sound matter by the musical pattern (13). But just as a musical scale cannot be grasped without reference to the sound matter, so in the analysis of the distinctive features such a reference is inevitable. Knut Togeby eloquently demonstrated this by a consistent assumption of the contrary (14). A distinctive feature cannot be identified without recourse to its specific property.

Such an investigation is supplemented but not supplanted by an inquiry into the distribution of these features in the speech sequences. M. Joos has observed, that since the diphthong /au/ (spelled ou as in council) is never followed within a simple English word by [p b f v m], this distributional feature defines the labial class of English consonants (15). Such a statement, however, presupposes the identification of each of the consonants in its various occurrences. We must know that /t/ in rout is identical with /t/ in rite which is opposed to /p/ in ripe as grave vs. acute, ceteris paribus. Otherwise, we would not know that in rout the diphthong /au/ is followed by /t/ and not by /p/, and we could not prove the above statement.

Thus for the identification of /p/, and of every other phoneme, a reference to the specific property of each of its distinctive features is imperative. 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 belongs to the future and is an urgent duty.

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. The exact measurement of the vocal tract permits the calculation of the sound wave (16), but the same acoustical phenomenon may be obtained by altogether different means. Similarly, any given attribute of the auditory sensation may be the result of

different physical variables (17) so that there is no one-to-one relation between the dimensions of the acoustical stimulus and the auditory attribute. The former cannot be predicted from the latter, but the totality of the dimensions of the stimulus renders the attribute predictable.

To sum up, the specification of the phonemic oppositions may be made in respect to any stage of the speech event from articulation to perception and decoding, on the sole condition that the variables of any antecedent stage be selected and correlated in terms of the subsequent stages, given the evident fact that we speak to be heard in order to be understood.

1.4 INHERENT AND PROSODIC DISTINCTIVE FEATURES

The distinctive features are divided into two classes: 1. inherent and 2. prosodic. The latter are superposed upon the former and are lumped together with them into phonemes. The opposition grave vs. acute, compact vs. diffuse, or voiced vs. unvoiced, and any other opposition of inherent distinctive features appears within a definite sequence of phonemes but is, nevertheless, definable without any reference to the sequence. No comparison of two points in a time series is involved. Prosodic features, on the other hand, can be defined only with reference to a time series. A few examples may clarify this statement.

A syllabic phoneme is opposed to the non-syllabic phonemes of the same syllable by a relative prominence. For the most part syllabicity is an exclusive function of the vowels. Cases when some vowels or liquids, ceteris paribus, carry the distinctive opposition syllabic vs. non-syllabic are particularly rare. For instance, the Old Czech sequence b r d u changes meaning depending upon the syllabic or non-syllabic character of the /r/ (see 2.226).

It is obvious that whether or not /r/ constitutes a maximum in loudness can only be determined by comparison with the loudness of the other phonemes of the same sequence.

In a sequence of syllables a relative prominence opposes one syllabic phoneme to the others of the same sequence as stressed vs. unstressed. In a number of languages words have, ceteris paribus, a different place of stress, for instance, English billow /b'ilou/ - below /bil'ou/. The greater and lesser prominence of syllabics is a relative notion which can be determined only by a comparison of all syllabics pertaining to the same sequence. The same holds when the distinctive role is played by the relation not of the loudness levels but of the pitch levels of the voice. In K. L. Pike's formulation, "the important feature is the relative height of a syllable in relation to the preceding or following syllables" (18).

When in place of or beside the level, the modulation plays a distinctive role, we identify the pitch or loudness contour of a phoneme by comparing two points in the time series. For instance, the Lithuanian falling pitch, which is opposed to the rising pitch and is due to a lowering of frequency, habitually accompanied

by a decrease of amplitude, is identified by comparing the initial and final fractions of the vowel affected. By a similar comparison we identify the Danish "falling loudness of the voice" (the so-called stød), which is due to a decrease of amplitude often accompanied by a decrease of frequency (19).

The prosodic opposition long vs. short (distinguishing either simple from sustained or simple from reduced phonemes) is based on the relative, not absolute, length of the phonemes in the given sequence. Their absolute duration is a function of the speech tempo. For instance, in the Czech pravá práva/prava: pra:va/ "true rights", the first vowel of the first word is identified as short in relation to the second, long vowel, while the second word displays the inverse relation.

1.5 THE DISTINCTIVE FEATURES COMPARED TO THE OTHER SOUND FEATURES

The smallest meaningful unit in language is called morpheme. A root, a prefix and a suffix are morphemes. A root word is a one morpheme word. The distinctive features and the phonemes possess no meaning of their own. Their only semantic load is to signalize that a morpheme which, ceteris paribus, exhibits an opposite feature is a different morpheme; cf. /gip/, /gib/ and /gid/. This discriminatory function may be assumed by more than one feature (and phoneme), as in the case of /bit/and /sed/.

There is no difference in function between diverse features (and phonemes). For instance, the question of what is the specific denotation of nasal consonants or, in particular, of /m/ in English, makes no sense. /m/ in map, mess, aim has on the semantic level no common denominator which would set it off from /n/ or from /b/. This lack of semantic difference between diverse distinctive features makes them purely discriminatory marks which are otherwise empty. It separates them from all other sound features functioning in language. Only these, purely discriminatory and otherwise empty units are used to construct the whole stock of morphemes of all languages of the world.

Configurational features are features which signal the division of the sound chain of the utterance into grammatical units of different degrees of complexity. For instance, in languages where the stress is bound to the initial (or final) syllable and, consequently, cannot serve as a distinctive feature, it functions as a border mark which denotes the beginning (or end) of the word. On the contrary, in a language where the stress is free (i.e. can fall on any syllable in the word), its place performs a distinctive function and contains no specific denotation.

From the various redundant and expressive features of English intonation, Z. S. Harris (20) has extracted three configurational units: "/?/ for rise, /./ for fall, /,/ for middle register (as against low register) base-line". /./ denotes the end of the sentence, /,/ the end of a phrase in a sentence to be continued, and /?/ the question, which in configurational terms means the end of a sen-

tence to be supplemented by an answer; i.e. the potential completeness of the utterance but incompleteness of the dialogue. When used as distinctive features, rise and fall have no other function than discrimination between morphemes, but when they serve as configurational features they carry a specific denotation; e.g. fall signifies the completeness of a sentence, and a rising intonation, even if superposed upon a mere nasal murmur, is immediately identified by English listeners as a question.

Expressive features are features which signal emotional attitudes of the speaker and the emphasis he puts on some of the particulars conveyed by his utterance. To use D. Jones' example (1), in the pronunciation of the English word enormous the emphasis may be effected "by an increase of strength coupled with an increase in the length of the vowel and the use of a special intonation" (a greater extent of the fall). In the expressive features, we deal with a special kind of relations. A neutral, unemotional variety is paired with the expressive variety which presents a "grading gamut" according to the term of Sapir, who defined this type of relation distinctly (21). Like the configurational features, the expressive features carry their specific denotation. In English the intensified stress, as opposed to the normal stress, denotes an emphatic attitude, and a further reinforcement of stress, a still more emphatic attitude.

The distinctive and the configurational features refer to the meaningful units of the utterance; the expressive features, to the speaker's attitude, and the redundant features (see 1.2) refer to other sound features: e.g. the redundant "clearness" of the English /l/ denotes that a vowel follows. Possession of a specific denotation unites the redundant features with the configurational and expressive features and separates them from the distinctive features. The "emptiness" of the distinctive features sets these apart from all other sound features.*

The following survey is confined to the inherent distinctive features. The prosodic features and other problems involving the sequential arrangement, in particular the segmentation of the sequence will be treated separately.

* In certain cases single distinctive features can assume an additional configurational function. In this function they obtain a positive denotation. For instance, in certain Scottish dialects where nasal vowels occur and are opposed to the oral vowels in the first syllable only (5), the occurence of a nasal vowel denotes the beginning of a word, but within the limits of the first syllable the opposition of nasal and oral vowels remains a "void" distinctive means.

II A TENTATIVE SURVEY OF THE DISTINCTIVE FEATURES

2.1 PREFATORY ACOUSTICAL REMARKS

In the sound spectrograms* the frequency-intensity pattern of speech is portrayed as a function of time. In this "running frequency analysis" the statistical properties of the speech wave are sampled within time intervals that are short compared to the duration of a phoneme. The spectrograms and the supplementary "cross sections" of intensity vs. frequency provide a source of information that may be rather confusing unless an optimal set of parameters is used in the analysis. These parameters can best be discovered by an analysis of language into distinctive features.

The speech wave may be considered as the output of a linear network; i.e., the vocal tract coupled to one or more sources. The speech wave has no other properties than those of the sources and the network. This relation may be written

$$W = T \cdot S$$

where W represents the speech wave, T the transfer function of the network, and S the source. Two simultaneous sources may be handled by superposition:

$$W = T_1S_1 + T_2S_2$$

Speech analysis shows that only a very limited number of characteristics of the source and of the transfer functions are utilized in the various languages of the world for semantic discriminations. These characteristics are described in the following paragraphs.

2.11 Properties of the Source Function Utilized in Language

2.111 Type of Source. There are basically two kinds of sources, periodic and noise sources. A periodic source is manifested by a characteristic harmonic structure in the spectrogram. A noise source, on the other hand, causes an irregular distribution of energy in the time dimension. These two sources can be simultaneously active in the production of a single phoneme.

2.112 Number of Sources. Some sounds such as [v] or [z] have two sources. One of these is located at a point of maximum stricture in the vocal tract, while the other, i.e. the so-called voice, is located at the larynx and is more or less periodic. A source which lies above the larynx in the vocal tract produces anti-resonances in the transfer function (cf. 2.122).

* The sound spectrograms to which reference is made in this report either are of the type produced by the Kay Electric Company Sonagraph or are from the book Visible Speech by Potter, Kopp, and Green (1).