LEIGH LISKER, FRANKLIN S. COOPER,
ALVIN M. LIBERMAN

The Uses of Experiment in Language Description*

ABSTRACT

To assess the validity of a grammar, one must derive utterances from it and then determine the acceptability of these utterances to native speakers of the language. At the level of syntax it is easy enough to isolate and recombine the relevant elements (words) and thus generate appropriate utterances, however difficult it may be to devise meaningful tests. In phonology, on the other hand, there should be no particular problem about designing tests; the difficulty is, rather, in deriving testable utterances, particularly when one insists, as he must, that the phonologically relevant elements (phonemes) be freely commutable. This paper attempts to deal with the latter problem.

Until recently the linguist who wanted to move from phonology to utterance could only use the human vocal apparatus, the flexibility and controllability of which is inadequate for the purpose. Instrumental methods seemed equally inadequate until, with the advent of magnetic tape recordings, it appeared that one might make commutation tests by cutting and rearranging tape segments of phoneme length. This has not proved to be feasible, however, and for reasons that have to do with certain very fundamental properties of speech.

Newly developed techniques for synthesizing speech make it possible now to write a phonological description from which testable utterances can be recovered 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 to their actualization as sound. Such rules are now available in acoustic terms, written at a phoneme or sub-phoneme level, and therefore appropriate as a basis for commutation testing. While these acoustic rules constitute a workable and testable phonology, they are not quite so simple as one might wish. Thus, in the case of at least one phoneme it is necessary to have two rules, one for each of two classes of vowel contexts; more generally, we must apply built-in "modifiers" to the various classes of phonemes in order to accommodate them to various positions in the syllable.

We believe that these complications can be reduced, and a simpler phonology achieved, by stating rules for synthesis in articulatory rather than acoustic terms. These articulatory rules should not, however, describe the changing shape of the vocal tract, since this

* This paper is based on research which has been supported since its beginnings by the Carnegie Corporation of New York. Much of the work of formulating the rules for synthesis was supported by the Research and Development Division of the Prosthetic and Sensory Aids Service, Veterans Administration, under Contract V1005M-1253.

would produce a result that is, like the acoustic rules, less than ideally simple; if greater economy of description is to be attained, the rules must rather be written in terms of the motor commands that actuate the articulators. Preliminary studies of muscle action potentials, recorded from several articulatory organs, tend so far to support this assumption. We are, however, a long way from having sufficient knowledge about these motor command patterns; moreover, there exists at the present time no synthesizer that will accept such commands as input and thereby permit a rigorous test of a phonology in these terms.

The procedures which lead to a phonological description in terms of acoustic rules for synthesis provide other data which should be of interest to the linguist, even though they are not critically relevant to his primary concern. One learns, for example, that various phoneme classes have very different perceptual properties; we are led then, to speculate that these differences may make the phonemes differentially efficient as vehicles of information and so determine their linguistic roles.

GRAMMAR AS TESTABLE DESCRIPTION

Whether the linguist prefers to believe that the business of grammar writing is one of discovery or invention, he is like other scientists in regarding his descriptions as tentative in nature. He must, in fact, be especially insistent on this point, for he is exposed to a hazard from which many other scientists are better shielded: a lay public with relatively easy access to observational data concerning which it has a stock of ready opinion. This ready opinion (to which the "educated community" is particularly prone) reflects the view that the grammar of a language is given, and that the "grammaticalness" of a speech event is derived therefrom:[1] indeed, it tends to go so far as to make this "grammaticalness" the necessary and sufficient condition for the acceptance of a speech stretch as an utterance of the language. For the linguist this is, of course, a matter of putting the cart before the horse. Since his purpose is descriptive and not legislative, the hypotheses which constitute the linguist's grammar must rather be tested to ensure that they conform to the observed behavior of speakers of the language. Only then can he arrive at a grammar which sets forth the pattern presumed to govern the production and acceptance of speech stretches as utterances in the language.

[1] If the attribute of 'grammaticalness' (or 'grammaticality'), as the term has been used recently, is to be taken as derivative, it is only so in that it presupposes a transcendental Grammar of the language. Despite the fact that linguistic discussion has been marred by an invidious, and odious, confrontation of "God's truth" and "hocus-pocus," it is lamentably true that the only kinds of grammars with which the linguist deals are the mundane products of his own activities. Therefore, in our discussion 'grammar' will refer to the linguist's output, while 'grammatical' and 'grammaticalness' will serve to characterize language samples according to the linguist's evaluation of certain kinds of associated behavior on the part of the native speaker.

In testing to determine the adequacy of his grammar, the linguist sometimes becomes the target of criticism by those among his scientific confreres who look with impatience on his habit of searching out the rare exception to the rule. The behavioral scientists, in particular, may suggest that it is misguided hubris on the linguist's part to insist that a statement in the grammar must either hold all the time or be rejected,[2] for it seems unreasonable to expect that human behavior can be predicted so completely. Now this suggestion may only be making the unobjectionable point that the cost of an increase in precision of description can sometimes be exorbitant, but it also implies that, at some stage in the process of grammar construction, recalcitrant data may be neglected as statically nonsignificant variations about some normal mode. This the linguist cannot allow, in principle.[3] It is true enough that he excludes observations of certain speech events (i.e., "lapses") from his working data, and his basis for doing so may be ultimately statistical;[4] but then the question is one of deciding whether a particular token is a proper "actualization" of an utterance type of the language, and on this decision the relative frequency of occurrence of the utterance type has no bearing. Just as the linguist does not require that his grammar enable one to predict the relative frequency of utterance types (and, in the limiting case, non-utterance types) in any finite corpus, so too he does not feel free to reject an utterance merely because his grammar, which otherwise provides for some arbitrarily high percentage of the utterance types in his corpus, fails to account for it. For though the relative frequency of an utterance type may possibly have relevance to the grammar, it will also depend on the circumstances in which the data were gathered;[5] the rare occurrence, or even the absence of a type, in one corpus may imply nothing about its frequency of occurrence in some corpus collected under different conditions.

[2] Thus M. Joos, in his "Description of Language Design," *Journal of the Acoustical Society of America* XXII (1950), 701–708, is possibly speaking for many linguists when he claims that "in principle every possible statement of ours must be either true or false—nothing halfway".

[3] Thus for example, in C. F. Hockett's "Note on 'Structure'," *International Journal of American Linguistics* XIV (1948), 269–271, "the one rule that all devotees of the game (that is, descriptive linguistics) agree on pretty well—is that all the utterances of the corpus must be taken into account."

[4] Lapses are informally defined, in Hocket's just cited "Note," as productions which the informant himself asks to be stricken from the record, or as features of utterances which he eliminates from alleged repetitions of those utterances.

[5] See L. Bloomfield, *Language* (New York: Henry Holt, 1933), p. 277: "The frequency of most lexical forms is doubtless subject to a great deal of superficial fluctuation, according to the practical circumstances."

Dual nature of language behavior

What we have just said means in effect that, with respect to the informant in his dual role of producing and responding to speech signals, we may not restrict our attention exclusively to his speech producing behavior; it is strongly asserted that we cannot dispense with observations of his responses to speech signals. In fact, when it is recalled that linguists have, from time to time, argued that grammars ought to (and perhaps, in theory, can) be constructed solely on the basis of a non-directively elicited corpus,[6] we are tempted to make the contrary assertion; namely, that in theory one might construct the grammar of a language solely on the basis of an informant's responses to speech or speech-like signals generated quite independently of any knowledge of the informant's behavior as a speaker.[7] It is unprofitable to take either proposal seriously as a working procedure, of course, for we are under no necessity of having to choose between them. Fortunately, the informant accepts most of the utterances he produces and can be induced to generate most of those he accepts.[8]

The need to test

If the grammaticalness of a speech event is independent of whatever grammar we may choose to construct, and if it is also not to be inferred strictly from the observed frequency of occurrence of an appropriate type in any corpus of material, then it can only be defined with respect to some procedure in which the degree of acceptability by native speakers is determined. Thus the determination of grammaticalness depends finally on the application of suitable tests of acceptability. Such tests are the link which connects the statement of the scope of a grammar with the process of validating that grammar or determining its actual domain.[9]

[6] This is presumably because, e.g., in Z. S. Harris *Methods in Structural Linguistics* (Chicago: University of Chicago Press, 1951), fn. 12 on p. 12, "it may not always be possible for the informant to say whether a form which is proposed by the linguist occurs in his language."

[7] This is not to say that it would be unnecessary to use care in devising listening tests. See fn. 10.

[8] For an extensive discussion of eliciting methods see C. F. Voegelin, "The Notion of Arbitrariness in Structural Statement and Restatement I: Eliciting," *International Journal of American Linguistics* XXV (1959), 207–220. It is to be noted that all the methods of eliciting described require that the linguist observe informant responses to, as well as productions of, speech.

[9] If these tests are to define an interesting scope for a grammar they must satisfy several minimum requirements: (1) they must not be so lax that all conceivable speech stretches are equally acceptable; (2) they must not be so stringent that utterance types of frequent occurrence in any corpus non-directively elicited from an informant will be

In the literature of linguistics the role of testing has rarely been the subject of extensive discussion. Perhaps this is to some extent explainable on the ground that testing procedures are lumped with other tricks of the trade known as "field methods in linguistics," a subject often regarded as more fit for student instruction than for theoretical ventilation. Nevertheless, in the course of doing field work, linguists may often test their observations by attempting themselves to produce utterances which conform to (or perhaps violate) some theory derived from those observations. Our discussion so far has been to make the point that grammatical statements must be tested; the fact that the linguist's tests are carried out under far from rigorous conditions is perhaps deplorable, but that is another matter.

Not only must testing be conducted as a necessary part of the business of grammar construction, but this testing should be experimental in the full sense of the word—it should include the testing of speech stretches which represent previously unobserved sequences of the elements obtained by analysis at the various levels of description.

Very recently there has been some public discussion of the procedures of minimal-pair testing at the phonological level[10] and of tests of grammaticalness at the level of syntactic description.[11] It would seem that the difficulties encountered in trying to test descriptions at these two widely separated levels are quite different in nature. At the phonological level the problem is largely one of devising a reliable method of proceeding from the description of utterances as phoneme strings to the production of acoustic signals for presentation to native users of the language; once this is feasible it should not be a serious problem to develop tests by which phonologically relevant responses can be elicited. Experimental testing at the syntactic level, on the other hand, is not hindered by any difficulty in isolating elements (words) and recombining them at will, but it is no simple matter to decide how to go about instructing the subjects so as to elicit responses

rejected by him; they should yield a classification of speech stretches into acceptable and unacceptable categories that does not do violence to the informant's intuitive judgment.

[10] See, for example, Z. S. Harris, *Methods in Structural Linguistics* (Chicago: University of Chicago Press, 1951), p. 32; M. Halle, "The Strategy of Phonemics," *Word* X (1954) 197–209; M. Kloster Jensen, *Tonemicity* (Bergen: Norwegian University Press, 1961); A. Abramson, *The Vowels and Tones of Standard Thai: Acoustical Measurements and Experiments* (Bloomington: Indiana University Research Center in Anthropology, Folklore, and Linguistics, 1962).

[11] See, for example, H. Maclay and Mary D. Slator, "Responses to Language: Judgments of Grammaticalness," *International Journal of American Linguistics* XXVI (1960), 275–282; A. A. Hill, "Grammaticality," *Word* XVII (1961) 1–10; N. Chomsky, "Some Methodological Remarks on Generative Grammar," *Word* XVII (1961) 219–239.

which are reasonably consistent and clearly relevant to syntactic description. In what follows we shall not be concerned with the problem of testing syntactic statements; instead our attention will be confined to phonological testing. This restricting of attention is dictated, not by any conviction that the need to test language description is less crucial at other levels, but rather by our own special interest in the perception and physical specification of speech as sequences of phoneme-size elements, and by our belief that recent advances in the technology of speech processing can contribute most significantly to the development of procedures for testing the phonological statements of grammars.

Testing at the phonological level

Generally speaking, the statements at any level of description in a grammar are concerned with either identifying the members of a set of elements or describing the ways in which these elements combine to constitute utterances of the language. Within the phonology we may distinguish at least two kinds of segmental elements, and thus two levels of description: the phonetic and the strictly phonological. The purely phonetic statements specify a set of elements (the phones) on the basis of their articulatory and/or acoustic characteristics;[12] the strictly phonological statements arrange the phones into higher-order elements (the phonemes) on the basis of both phonetic and distributional considerations, and then go on to characterize the phonemes themselves phonetically and distributionally. Ultimately, statements of the last kind may treat of certain relational properties of the phonemes considered as purely abstract elements of a set—that is, as elements of a set algebra. Viewed thus, the phonemes appear as more or less abstracted entities, divorced from the physical observations that served as the starting point of description.

Although the analytic processes which intervene between phonetic extraction and phoneme crystallization may obscure, or even eliminate, the basis for any "straightforward" physical definition of the phonemes, it is nevertheless true that the phonological description of a language must enable one to recover the phonetic specification of an utterance from its phonemic representation. It seems reasonable, moreover, to suggest that the phonetic specification of an utterance must somehow be convertible to an audible signal which can be presented to an informant for identification. In other words, one may regard the symbols of a phonemic transcription as representing a sequence of instructions which should lead more or less directly to a token of the utterance type symbolized.

[12] More precisely, on the basis of these characteristics as perceived by the linguist.

From transcription to speech: practical difficulties

The view that the phonological description of an utterance ought to pass some test of its "generative adequacy" has been generally held by linguists; certainly they have regularly resorted to informal testing in order to make sure that each distinct utterance type of a corpus has a unique representation that can be "read back" as an acceptable token of that type. The linguist's demonstration that he can convert the phonemic representation into an utterance, while it may convince us of the plausibility of his description, does not constitute a rigorous test of that description. He is involved in two operations in performing such a test: (1) he must replace the phonemic symbols by phonetic ones, and (2) he must produce the utterance by adhering slavishly to the physical specifications which define the symbols of the phonetic transcription. The first requirement can be met without too much difficulty, but the second implies a degree of control in the production of speech sounds to which no human speaker can lay claim. In the absence of any way of guaranteeing that the phonetic transcription has been read back in strict accord with the symbols, it is not possible to talk in any precise way about the "accuracy" of a phonetic transcription.

The fact that there is no way of ensuring that an utterance has been correctly reconstituted from its phonetic transcription does not prevent linguists from claiming that certain phonological statements rest on some sort of commutation testing. In particular, the special variety of commutation test concerned with minimal pairs (which themselves are only established by such testing) has sometimes been appealed to as a basis for asserting that a given pair of segments are phonologically distinct. Although some of the discussions of commutation testing might suggest otherwise, such tests have been for the most part strictly "gedanken-experimental" in nature, simply because it is not possible for a speaker to control his speech production with the requisite accuracy.

In order to make commutation testing a workable experimental procedure for deciding whether two segments of phone length are linguistically the same or different, it is necessary somehow to get around the problem of controlling the speaker's production. Various evasive tactics have been attempted: mechanical "talking machines"[13] in great variety have been used to investigate speech, though not as serious contenders in phonological testing; of the more recent methods, one involves the manipulation of speech after it has been recorded, and the other resorts to electronic devices which can produce synthetic speech. As we shall see, only the methods that employ synthetic speech turn out to be feasible for commutation testing at the phonological level.

When magnetic tape recorders became generally available, many students of language hoped that phonologically relevant data might be obtained by cutting tape-recorded speech into segments, commuting and resplicing these into new sequences, and then getting informant identifications of the processed signal. This method of commutation testing has been well explored by now and has turned out to be of very limited utility.[14] This is so because of the very basic fact (the root of the "problem of segmentability") that it is impossible to cut the tape so as to isolate each of the phones composing the recorded utterance. When tape segments of about the size of the individual phones (cut to minimize overlap and with as many tape segments as there are phones) are reassembled in new combinations, the audible signal may still be perceived as speech of a sort, but it is speech so garbled that it may not even be identifiable as an utterance. Because of this, the commutation of segments by manipulating tape recordings is, in general, no more useful in testing phonological statements than is a procedure which requires that the linguist be capable of something like absolute control of his own vocal mechanism, though for a basically different reason.

Altogether, it appears that we must give up the hope of developing any rigorous method for testing phonological description that involves the experimental manipulation of either the speaker's behavior or his recorded

[13] A summary account of some very early talking machines is to be found in H. Dudley and T. H. Tarnoczy, "The Speaking Machine of Wolfgang von Kempelen," *Journal of the Acoustical Society of America* XXII (1950), 151–166. Another well-known account of these machines is that of R. Paget, *Human Speech* (London: Kegan Paul, Trench, Trubner, 1930), pp. 12–19.

[14] For an account of the serious difficulties that arise, see Cyril Harris, "The Building Blocks of Speech," *Journal of the Acoustical Society of America* XXV (1953), 962–969. Interesting studies by Wang and Peterson and by Peterson, Wang, and Sivertsen have shown that if one wishes to cut and resplice segments that approximate phonemes in size and yet recombine to produce intelligible speech, he must (1) have a very large inventory of segments (over 1,000 even when intonation differences are ignored) and (2) be satisfied with segments (called 'dyads') so cut that each one contains "parts of two phones with their mutual influence in the middle of the segment." (See W. Wang and G. Peterson, "Segment Inventory for Speech Synthesis," *Journal of the Acoustical Society of America* XXX (1958), 743–746; G. Petersen, W. Wang, and E. Sivertsen, "Segmentation Techniques in Speech Synthesis," *Journal of the Acoustical Society of America* XXX (1958), 739–742). The size of the segment inventory varies according to the size and type of the segment, as shown by E. Sivertsen, "Segment Inventories for Speech Synthesis," Part II, *Report No. 5, Speech Research Laboratory* (University of Michigan, Ann Arbor). In certain special cases it is possible to cut and recombine segments of approximately phoneme length, as has been done for example, in studies of the cues for speech perception by C. Schatz and K. S. Harris. (See C. Schatz, "The Role of Context in the Perception of Stops," *Language* XXX (1954), 47–56; K. S. Harris, "Cues for the Discrimination of American English Fricatives in Spoken Syllables," *Language and Speech* I (1958), 1–7.)

production. The linguist's picture of language as made up of a limited number of elements which, like the printer's movable type, can be more or less freely combined, cannot at present be matched by any direct operations on speech itself.

Commutation testing and synthetic speech

A feasible method of commuting segments of phone length is to deal, not with speech itself, but with an approximation to speech—synthetic speech. In this connection we should point out first that there do now exist instrumental techniques for speech synthesis that allow the production of acoustic signals which the native speaker will identify as words and sentences of his language; these speech synthesizers also permit the manipulation of acoustic signals subject to very precise control, and in ways that make it possible to determine just which acoustic features contribute most to the identifiability of a signal as some particular utterance of the language. We can, for example, fix on some particular time interval within a signal, vary that portion of the signal in statable ways so as to generate a family of test stimuli in which the acoustic variation is perceived as a more or less gradual shift in phonetic quality, and then present these stimuli to native speakers for identification as utterances. In the event that the speakers differentiate among stimuli, their responses make it possible to point to the particular segments in the acoustic pattern that served a distinctive function.

When we resort to speech-making devices, we can not only synthesize reasonably satisfactory approximations to human speech; we can, indeed, generate synthetic productions that are in a certain way superior to human speech: it turns out that in producing audible signals acceptable as speech, one is not constrained to copy faithfully the patterns found in spectrograms; one can as well make do with signals whose spectrograms are drastically stripped-down versions of the spectrograms of humanly produced speech, and, in so doing, devise signals that have segmentability built into them. This is not to say that we can cut and resplice tape recordings of such synthetic speech—segmentability of this kind is likely to prove impossible of attainment, so long as we deal with language at the acoustic level—but that the information supplied to the synthesizer can be formulated as a set of rules for synthesis, and that these rules can be fed to the synthesizer in any desired order to produce an intelligible speech-like output. Such rules are, in a very real sense, the functional counterparts of conventional phonemes in that they serve as the freely commutable elements of a description of the language. They have, however, the important advantage of being explicitly convertible to speech. Moreover, in the process of conversion the rules for synthesis generate sub-rules which ensure that each phoneme of the synthesized utterance will be represented by its appropriate (allo)phone. Thus, the technique of speech synthesis makes it possible, and for the first time, to use commutation testing as a feasible way of determining the relations among phones, and hence a check on phonological description.

Rules for synthesis as a descriptive system

Rules for synthesis did not spring full-blown from a speech synthesizer, nor are they the result of a research enterprise directed specifically at producing them. Rather, they were formulated from data collected in a series of studies[15] aimed at finding the acoustic features which serve as cues for the perception of utterances.

When, after many experiments, we had most of the important acoustic cues in hand, we were able to produce speech that was synthetic in the strictest sense. That is, we were able to formulate data about the cues for phones as explicit instructions for converting a phonetic transcription into a schematized spectrogram, and then, by putting the spectrogram through a pattern playback, to produce intelligible speech. Moreover, and more importantly for this paper, it was not difficult to reorganize the data about the cues in such a way as to consolidate the sets of instructions for specific phones (i.e., the sub-rules) into more general rules for synthesis that can be applied directly to a *phonemic* transcription. Thus, by re-working the results of perceptual operations on more-or-less arbitrary acoustic stimuli, we obtained eventually certain sets of entities (sub-rules and rules) that correspond roughly to the two kinds of elements (phones and phonemes) used in phonological description.

The nature and structure of these rules for synthesis have been described in detail elsewhere,[16] together with an example of their application to a specific phonemic transcription. The example, in particular, exposes the internal structure of the rules. When the instructions which make up the rules are considered separately, it appears that they are, in general, not unique to a particular rule. Instead, the rules are susceptible to grouping on the basis of similarities in their component specifications (instructions). As might perhaps be expected, the groupings agree very closely with the usual phonetic classification of speech sounds in terms of place and manner

15 Many of these studies are summarized in A. M. Liberman, "Some Results of Research on Speech Perception," *Journal of the Acoustical Society of America* XXIX (1957), 117–123.

16 A. Liberman, F. Ingemann, L. Lisker, P. Delattre, and F. Cooper, "Minimal Rules for Synthesizing Speech," *Journal of the Acoustical Society of America* XXXI (1959), 1490–1499.

of articulation. Thus the rules for English /pbm/ share certain specifications which quite evidently relate to a subphonemic feature of labiality; the rules for /mnŋ/share specifications that can be paired with nasality (though not necessarily with nasalization) and, in general, subsets of the complete specifications for a rule show correspondences with the familiar dimensions of articulation.

The fact that the structure of the rules involves dimensions which correspond rather closely to the subphonemic dimensions of conventional phonology has practical utility as well as theoretical interest. It implies a substantial degree of interconvertibility (at the subphonemic level) between the articulatory and acoustic descriptions of phonological units. The practical consequence is that synthetic methods can be used to test some aspects, at least, of phonological descriptions of the usual kind. Thus, assumptions about the role of a particular feature (e.g., labiality or nasality) in establishing "phonetic similarity" for the native speaker can be evaluated experimentally: synthetic spectrograms can be constructed to include (or to omit) the specific feature in question while retaining all other features, and the sounds generated from these patterns can then be presented to native speakers for identification, as sounds of their language. Clearly, the method depends on the possibility of converting a statement about a subphonemic feature into explicit differences in schematized spectrograms. This is possible only to the degree that the dimensions of the rules-for-synthesis description correspond with the dimensions of the phonological description that is being tested; in practice, the correspondence is close but not perfect.

A different aspect of the internal structure of the rules appears when we ask, not about relationships among the rules, but about the process of converting a transcription into sound via schematized spectrograms and a synthesizer. We can then classify the individual instructions that comprise the rule according to their functions in synthesis. One group of instructions then relates quite clearly to the phonological unit *per se*; these "core" instructions specify such things as locus frequencies or formant frequencies that apply to the phonological unit in all of its variants. Another group of specifications is needed to provide continuity of pattern elements and thereby to match the continuous character of the articulatory movements; these "connectivity" instructions are, in fact, so simple that they appear only implicitly in the rules as now formulated, e.g., as part of the definitions of "explicit" and "implicit" loci. Other sets of specifications, which we may call "position modifiers," affect the application of the core instructions in ways that are determined by context, i.e., by the identities of neighboring rules in the particular sequence that is specified by the

transcription. It is, indeed, the function of these position modifiers, operating on the set of core instructions, to implement the requirement that rules for synthesis, like phonemes, should be freely commutable.

Of the three kinds of instructions which go into the rules for synthesis the core instructions may, as we said, be looked upon as acoustic specifications that correspond essentially to the phoneme. It is somewhat more difficult to show just how the other kinds of instructions may be related to aspects of the phonological description. The position modifiers in part can be connected with those phonetic differences which characterize the allophonic varieties within the phoneme; we may say that the position modifiers, together with the core instructions, provide the acoustic specifications for a "sub-rule" which is related to the allophone as the rule is to the phoneme. More strictly, the position modifier finds its analog in the description which accounts for subphonemic variability as the result of "slurring" between contiguous phonemes. As for the connectivity instructions, we find that their function in insuring pattern continuity is not matched, nor need it be, by that of any aspect of phonetic description, for this latter mode of description has no need to make explicit provision for a continuity of articulatory movement that the physical nature of the vocal tract enforces on the human speaker. The connectivity instructions are needed simply because we have chosen to use the acoustic mode of description in our examination of speech, and from this choice it follows that our synthesizer (or any synthesizer that uses our description of speech) must respond to an acoustic "input language." If, for example, we had chosen the articulatory mode of description, then we should have required a synthesizer of the vocal analog type, such as we will have occasion to talk about later in this paper; that kind of device might one day be designed with built-in constraints like those of the human articulatory mechanism, so that its input language would (like phonetic description) have no need of the particular kind of connectivity instructions required as part of an acoustic description. This somewhat detailed account of the anatomy of rules for synthesis might seem to imply excessive complexity; in fact, the rules are rather simple and not very numerous.[17] The reason is that the connectivity and positional instructions apply, in the main, to the relationships, not between individual rules (i.e., phonological units), but to those between classes of rules; for example, the connectivity instructions contained in the definition of implicit loci apply to *all* the stop consonants of English in every position; similarly, the position modifier for vowel length (in the example[18] cited) applies to *all* vowels. This happy situation is, to be

[17] See A. Liberman *et al.*, especially p. 1496.
[18] See A. Liberman *et al.*, especially pp. 1494–1495 and 1497–1498.

sure, no more than a reasonable consequence of the dimensional structure of the rules, described in earlier paragraphs.

It may be of interest to compare the phonemic representation of an utterance with the sequence of rules for that same utterance. The exact relation between the phonemes and the rules for synthesis cannot be determined in any *a priori* fashion, for we must suppose that the degree of correspondence will depend on the criteria by which each kind of element is set up and also, perhaps, on the language that is being described. There are, however, some fairly general questions that can be answered: what is involved in proceeding from phonemic transcription to acoustic signal? and is there a one-to-one correspondence between the "phonemes" of the two phonologies? When we examine the conversion of transcription into audible utterance, it appears that there are two steps to this process, whether it is accomplished by a human being or a synthesizer. In the first case, the human reader must in effect replace the phoneme symbol by one representing its appropriate allophone; in the second case, the synthesizer must apply those built-in provisions of the rule (especially the position modifiers) that alter the course of synthesis in order to take account of the specific context in which the rule is found. Hence, in a functional sense, if not in a formal one, both conversions of symbol into sound are two-stage operations, and there is a close parallel between them. The formal resemblance is even closer if, instead of replacing the phoneme by one of a set of discrete allophones, the reader applies certain "slurring" operations in rendering the phoneme symbol into speech.

The correspondence between conventional phonemes and rules for synthesis is, at least for American English, almost one-to-one—but not strictly so. The clearest case of a lack of such correspondence is presented by the velar stops: although the phoneme /k/ poses no special problem in a conventional phonology of American English, we require two rather distinct rules to synthesize a satisfactory [k] before all vowels (and likewise for [g]).[19] These two rules can, of course, be classed together on the basis of their non-contrastive distributions, but they will then constitute a new type of rule, essentially different in nature from the rules for the otherwise similar /p/ and /t/. The anomalous status of /k/ is presumably to be explained on the basis of some discontinuity, not at the articulatory level, but at the point of conversion of the articulatory gesture to acoustic output. It would appear that the acoustic description provides a poorer approximation to the ideal phoneme than does the articulatory characterization,[20] at least for /k, g/.

[19] See A. Liberman *et al.*, page 1496.

[20] Another example of the special problems that arise in synthesis from acoustic rules is given in the paper by A. Liberman, *et al.*, which we have already referred to. (See page 1494.) This is the case of the sequence /glu/ in initial position, where one finds that a simple concatention of the otherwise effective rules for juncture, /g/, /l/, and /u/ yields an audible signal that is significantly less intelligible than do the same rules in other combinations. Intelligibility is markedly improved when the rules corresponding to these phonemes are qualified by certain position modifiers, though these unfortunately reflect contextual constraints that are somewhat more severe than in the usual case, since they involve making changes in each rule which depend on a more restricted domain in which the position modifier operates.

The discussion of the rules for synthesis can be summed up rather briefly: even if one chooses to stay close to the acoustic stuff of speech, it is possible to generate a phonology that meets the major criteria for such a system, namely, parsimony of description and the employment of entities (rules) that are discrete, freely commutable, and explicitly convertible to sound. The rules, as phonological units, have the reassuring property of standing (almost) in a one-to-one correspondence with the phonemes of more familiar descriptions; moreover, their internal structure exhibits dimensions that map onto the articulatory dimensions of these phonemes. The correspondence at the subphonemic level provides a useful—and otherwise unavailable—method for conducting rigorous acceptance tests of the utterances specified by a conventional phonology.

The lack of an exact one-to-one correspondence and the occasional requirement for "position modifiers" points to a somewhat more remote connection between acoustic description and the ideal phoneme than between articulatory description and this goal. One must add, however, that neither description achieves the ideal, as attested by the lack of a direct, single-step process for converting the phonological units into speech. May there not be some third mode of description by which elements having the status of the phonemes and the rules might be mapped directly into the utterance?

Motor Commands: A Simpler Mode of Description?

We have seen that attempts to describe language in acoustic terms encounter an immediate difficulty due to overlapping manifestations of the successive phonemes; this is, of course, the well-known problem of segmentation that frustrates tape-cutting procedures, and that is so vividly portrayed in Hockett's Easter egg analogy.[21] Much of the foregoing discussion has been concerned, at base, with finding a solution that could be stated in acoustic terms.

[21] C. F. Hockett, *A Manual of Phonology* (Indiana University Publications in Anthropology and Linguistics, Memoir 11, 1955), pp. 210–211.

The particular acoustic description of a language as a set of rules for synthesis has the major virtue that segmentability has been regained: the units of description are, as indeed they must be, distinct and commutable; moreover, they correspond to the segmental phonemes, and so conform to the distributional and morphological constraints of the language. One may ask, however, whether this is a genuine gain or merely a contrived solution, with unresolved difficulties swept under the rules. It is clear that the rules must somehow provide a basis for generating speech sounds that will flow smoothly from phone to phone; we have seen that this can be done by the inclusion of connective and positional instructions within the set of rules appropriate to each phoneme, and that this is not an uneconomical procedure.

We have seen also that a close parallel exists between the sub-rules generated by the operation of the position modifiers in a rules-for-synthesis description and of allophones in the usual phonemic account of a language. Both are needed to implement actual tests of acceptance by a native listener. Both confess a less-than-ideal independence (commutability) of the units in which the description is couched. A quantitative, figure-of-merit comparison of the two kinds of descriptions would be difficult to make, but the two seem roughly comparable in the extent to which "allophonic variation" must be invoked.

Can we hope to find a simpler description—one in which the abstract units that meet our intuitive expectations about segmentability and commutability will also serve, without intermediate steps, to generate the sound sequences of the language? And where shall we look for such a description? If language exists anywhere in the simple, lawful form we seek, it must surely be within the brain of the user, safe for the moment from direct scrutiny. We know that the outward forms of language exhibit successive recordings (on the phonological level) as the message proceeds from oral to acoustic to aural mode. The inward versions should be the simpler the nearer they are to the neural center where the message is presumed to exist in pristine form, and so a description in terms of the innermost accessible modality should be a reasonable objective.

The dictum that "we speak to be heard in order to be understood" suggests that the search for this proximate modality should follow the message into the ear; this is the conclusion drawn by its authors,[22] reflecting their assumption that phonological simplicity is to be found at the level of auditory decoding. There are difficulties with this view, both in principle and in practice. One is a consequence of the unidirectional flow of speech, always from mouth to ear: the successive transformations by which the message is recoded from one modality to the next need not have inverses that will operate simply, or even uniquely, to restore the message to a prior form. Thus, in principle at least, it may not be possible to work the puzzle backward, from auditory attribute to acoustic stimulus; in practice also, the search for acoustic invariants to match the presumed auditory entities has proved largely unrewarding.

[22] R. Jakobson, C. G. Fant, M. Halle, *Preliminaries to Speech Analysis: the Distinctive Features and Their Correlates* (Acoustics Laboratory, Massachusetts Institute of Technology, Technical Report No. 13, January, 1952). See especially the concluding paragraphs of Section 1.3.

Moreover, the temporal priority of source over sensor, and the logical necessity for a motor encoding of the message if there is to be any sensory decoding, would seem to imply *two* centers of equal status. Such a model might not appear unreasonable for two-way communication, where speaker and hearer are different persons, but it is hardly a parsimonious one for that most attentive of all listeners, the speaker himself. All that he needs (in functional terms) is a single center served by a link between sensory and motor areas.[23] When he converses with his fellows, their speech, arriving at this same center, need only be interpreted as is his own. Viewed in this way, the speech process operates as a closed loop in which the message circulates through successive recodings, one of which corresponds in discreteness of units and regularity of structure to our perception of the message, or, more precisely, to the phonological constituents of the message.

Little can, or need, be said about the location(s) of this center, but we can usefully enquire into its internal structure: are the perceptual units arranged on the dimensional framework of an auditory space, or of a gestural space? The answer will strongly affect our choice of where to look for simplicity of phonological description.

The evidence is largely inferential but it points, in the main, to phoneme representation in motor terms. This would, of course, be a reasonable inference from the body of linguistic experience to the effect that articulatory shape provides a better basis for phonetic description than does sound impression. There is persuasive evidence, too, from experimental data about the perception of speech sounds: the acoustic cues that distinguish the phonemes fall into well-patterned arrays along articulatory dimensions, but do not arrange themselves so simply along the familiar psycho-acoustic dimensions. In certain crucial cases the perceptual evidence is more than merely persuasive: thus, a comparatively small change in articulatory shape can sometimes give rise to a rather large

[23] The distressing effects of delayed feedback would seem also to support this view.

and discontinuous change in acoustic pattern, thereby forcing the perception to reveal an affiliation with articulation rather than with sound.[24]

Description in terms of motor commands

If phoneme representation is, indeed, in motor terms, then investigation should be directed, not toward audition, but to the earliest accessible events in production. These may well be the electrical activities and contractions of the various muscles used in speaking. One can infer quite directly from these data the corresponding neural commands, so that it should be possible to derive a phonological description in terms of *motor command patterns*, i.e., in terms of the temporal–spatial patterns of activation of the muscles, or groups of muscles, responsible for generating and articulating speech.

Will this description differ, in any significant sense, from one in terms of the configurations and movements of the articulatory tract, i.e., from that of conventional articulatory phonetics? Both deal with the physiological events of speech production, so that similarities are certainly to be expected. Differences, if any, would arise in generating the overt gesture from the motor commands. In some instances, the relationship between the activation of a muscle and the resulting shape or movement of the articulators is simple, direct, and independent of neighboring events, but this is clearly the exception. Usually, the encoding of motor commands into shapes and movements of the tract is a complex transformation, one that could not be computed, even in principle, without taking account of interactions over stretches of the order of syllabic length; there is, in short, an unavoidable smearing of the phoneme boundaries, and a corresponding loss in segmentability. All these complications are to be attributed, presumably, to the interactions and constraints inherent in the mechanism of the vocal tract. The motor commands operate ahead of this part of the system, and so escape distortion by it. The difference between the two descriptions is, therefore, a significant one, and we should expect important gains in simplicity if we can arrive at a description in terms of motor command patterns.

Experimental work aimed at a systematic description of this kind is only well begun, but it has already confirmed that electromyographic and other physiological techniques can provide useful data; further, an early study on bilabial stop and nasal consonants has indicated a gratifying degree of independence among physiological features that are distinctive for these

[24] A review of the experimental data that support this conclusion is given in Liberman (see footnote 15).

sounds.[25] Simplicities of this kind, if they are found to hold generally, would point to a subphonemic structure characterized by dimensions that are not only mutually independent but also identified with observable parts and functions of the speech mechanism, a significant step, we think, toward a substantive model.

The requirement for testing.

Specification of the phonological units of language in terms of motor command patterns might then, if early hopes prove not false, meet the two criteria of parsimony of description and compatibility of the implied model with physiological fact. But more is necessary, as we have already insisted in connection with more conventional phonemic descriptions. The newly defined units must meet the commutation test so that, when permuted into combinations found in the language, they will fully prescribe the generation of acoustic sequences acceptable to the native speakers of that language.

Commutability is, of course, closely tied to the mutual independence of the sets of commands that comprise the individual phonological units. Thus, experiments of the kind already reported[26] offer presumptive evidence. The physiological model, too, leads one to expect no great difficulty in composing gestural sequences of syllabic length (or longer) by assembling sets of commands in the required temporal order.

There is, however, a serious difficulty in carrying out the decisive test of generating acceptable utterances directly from the commands; it lies in the present lack of a speech synthesizer that can operate with motor command patterns as its input instructions. There do exist synthesizers that copy electrically the geometry of the vocal tract[27] and one, at least, that can change the configuration of its tract at normal speech rates.[28] These devices

[25] G. Lysaught, R. Rosov, and K. Harris, "Electromyography as a Speech Research Technique with an Application to Labial Stops," *Journal of the Acoustical Society of America* XXXIII (1961), 842 (Abstract).

[26] Lysaught *et al.* In addition, MacNeilage has shown that although the acoustic representation of English /f/ varies with its context, the muscle potential patterns are similar for all positions. (P. MacNeilage, "Electromyographic and Acoustical Study of the Production of Certain Final Clusters," a paper presented at the Sixty-Third Meeting of the Acoustical Society of America, New York, May 23–26, 1962.)

[27] H. K. Dunn, "The Calculation of Vowel Resonances and an Electrical Vocal Tract," *Journal of the Acoustical Society of America* XXII (1950), 740–753; G. Fant "Modern Instruments and Methods for Acoustic Studies of Speech" in *Proceedings of the Eighth International Congress of Linguists* (Oslo: Oslo University Press, 1958), pp. 349–351; K. N. Stevens, S. Kasowski, and C. G. Fant, "An Electrical Analog of the Vocal Tract," *Journal of the Acoustical Society of America* XXVII (1953), 734–742.

[28] G. Rosen, "A Dynamic Analog Speech Synthesizer," *Journal of the Acoustical Society of America* XXX (1958), 201–209; M. Hecker, "Studies of Nasal Consonants

operate, however, on the basis of the *shape* of the vocal tract, not the pattern of motor commands that *generate* that shape; accordingly, we must expect the control signals to be more complex and less segmentable than they would otherwise need to be.

It may prove feasible, as a practical measure, to enlist the aid of a computer in converting (from stored tabulations of empirical data) the desired sequence of motor command patterns into tract configurations, and thence (by synthesizer) into acoustic sequences. But this can serve as a test of the description itself only to the extent that the stored tabulation can be so written as to allow a distinction between entries needed to offset shortcomings of the description *per se*, and those required to account for the operation of the flesh-and-blood transducing mechanism. The alternative procedure, direct conversion of motor command patterns into speech sounds, would certainly be preferable; it does, however, require construction of a synthesizer that copies in its mechanism the interactions and constraints of the vocal tract. This seems now a remote prospect but, whatever the difficulties, some method of synthesis must be found if a rigorous appraisal is to be made of the description of phonological units in terms of motor command patterns.

One measure—we think an important one—of the merit of any phonological description is the degree to which acceptable sound sequences can be generated *directly* and *solely* from the independent, freely-commutable unit descriptions. There should, in the ideal case, be no need for intermediate stages in the generative process in order to accommodate varying contexts and, accordingly, there should be no role for intermediate entities (e.g., the allophones of conventional descriptions) or corrective procedures (e.g., the position modifiers in a rules-for-synthesis description). By what margin will a motor-command description miss this ideal?

Other Results of Experimental Studies

So far we have considered experimental procedures that can be applied to language in order to establish an accurate and economical description, at least on the level of phonology. That much is surely relevant in a direct way to linguistics and to the problems with which the linguist is commonly concerned. It is in order now to point out that essentially the same experimental procedures can be expected to yield, in addition, a great deal of information which ought to be of considerable interest to the linguist, whether or not he chooses to incorporate it into his system. We have in

with an Articulatory Speech Synthesizer," *Journal of the Acoustical Society of America* XXXIV (1962), 179–187.

mind, as an example, the light that can be shed on phonemes as perceptual units with differing psychological properties, especially in regard to the categorical versus continuous nature of the perception. Differences of this kind may make phonemes, or classes of phonemes, differentially efficient as vehicles of information, and thus determine, at least in part, their linguistic roles.

Categorical vs. continuous perception: experiments with voiced stops

Consider, first, the voiced stops, and imagine experiments in which we synthesize these sounds from spectrographic patterns that include the acoustic features appropriate to the class. Having selected an aspect of the pattern that is sufficient to cue the perceived distinctions within the class, we vary this feature in small, progressive steps, and then listen to the result. We hear stops, of course—first /b/, then /d/, then /g/. But more important for the purpose of this discussion is the fact that we do not hear a gradual change from one phoneme to the next, corresponding to the gradual change along the acoustic continuum. Rather, we hear the first several patterns as /b/, then very abruptly the perception changes to /d/; as we move further along the acoustic continuum we continue to hear /d/ until the perception again changes very abruptly, this time to /g/. Thus, continuous variations in the acoustic stimulus are perceived discontinuously or, in this case, categorically.

It has been possible in several studies[29] to convert the impressionistic evidence described above to more nearly precise terms. Thus, in the case of /b, d, g/ we have measured discrimination along an acoustic continuum (specifically the extent of second-formant transition) by arranging the stimuli into ABX triads and asking the listeners to decide whether X was identical with A or with B. To locate the phoneme boundaries on the same continuum, we presented the patterns with instructions to identify each

[29] See A. M. Liberman, K. S. Harris, H. S. Hoffman, and B. C. Griffith, "The Discrimination of Speech Sounds within and across Phoneme Boundaries," *Journal of Experimental Psychology* LIV (1957), 358–368; B. C. Griffith, "A Study of the Relation between Phoneme Labeling and Discriminability in the Perception of Synthetic Stop Consonants," unpublished Ph.D. dissertation University of Connecticut, 1958; A. M. Liberman, K. S. Harris, J. A. Kinney, and H. Lane, "Some Cues for the Distinction between Voiced and Voiceless Stops in Initial Position," *Journal of Experimental Psychology* LXI (1961), 379–388; A. M. Liberman, K. S. Harris, P. D. Eimas, L. Lisker, and J. Bastian, "An Effect of Learning on Speech Perception: the Discrimination of Durations of Silence with and without Phonemic Significance," *Language and Speech* IV (1961), 175–195; J. Bastian, P. Eimas, and A. Liberman, "Identification and Discrimination of a Phonemic Contrast Induced by Silent Interval," *Journal of the Acoustical Society of America* XXXIII (1961), 842 (Abstract).

one as /b/, /d/, or /g/. It was found that, with acoustic differences between stimuli held constant, discrimination was considerably better across phoneme boundaries than it was within the phoneme categories.

These increases in discrimination at the phoneme boundaries are a reflection of the essentially categorical perception of the stops. We wanted, however, to determine the degree to which the perception is categorical, and for that purpose we carried out the following procedure. First, we made the extreme assumption that the listener can *discriminate* the speech sounds only as well as he can *identify* them as phonemes—that is, that he can hear no intra-phonemic variations. Given the data on how the listener labeled the sounds as phonemes, we can then calculate what the level of discrimination would be if the extreme assumption of categorical (i.e., phonemic) perception were true. Discrimination functions that were derived on this basis were found to fit rather closely those that had been obtained in the experiment.[30]

Categorical perception: an effect of learning?

We will want in a short while to describe the very different perceptual properties of certain other phonemes. Before doing that, however, we should make several points about the categorical perception of the stops. Thus, it should be said that from the standpoint of standard psychophysics the results obtained with the stops are most unusual. Stimuli that lie on a single continuum are ordinarily perceived continuously—that is, the discrimination function is typically monotonic (without peaks) and the observer can discriminate many times more stimuli than he can identify in absolute terms.

The perception of the acoustic continuum on which the stops lie, is, indeed, so exceptional as to make us suspect that it has been much influenced by learning. To find out whether or not this is so, we have in several studies[31] measured the discriminability of an acoustic variable that cues the phonemic distinction, and then measured the discriminability of the same variable in a non-speech context. The point here, of course, is that we take the results obtained with the non-speech controls as an approximation to the discriminability that the speech stimuli would have had without

[30] The goodness of fit has varied somewhat from one experiment to another depending on certain procedural details and also, perhaps, on the nature of the distinction being investigated (e.g., whether /b, d, g/ or /d, t/). The closest approximation to categorical perception has so far been found with /d, t/. (See footnote 29, Liberman, Harris, Kinney, and Lane.)

[31] See Liberman, Harris, Kinney, and Lane; also Liberman, Harris, Eimas, Lisker, and Bastian.

benefit of linguistic experience. In all these control experiments we have found no increase in discrimination as stimulus values that would have corresponded to phoneme boundaries; thus, it appears that the discrimination peaks found with stimuli that are heard as speech sounds are an effect of learning. It has been found, further, that the discrimination of the control stimuli is generally poorer than that of the speech, from which we conclude that the effect of linguistic experience must have been to *increase* discrimination across the phoneme boundary; there was no evidence of a learned reduction in discrimination within a phoneme category.

Given that there is an effect of learning, one asks then more precisely: what is learned? As we implied in another connection earlier in this paper, we believe that a speaker (and listener) learns to connect speech sounds with their appropriate articulations, and that the sensory feedback from these movements (or, more likely, the corresponding neurological processes) come to mediate between the acoustic stimulus and its perception. It would follow, then, that acoustic stimuli which vary continuously along, say, the /b, d, g/ dimension are perceived discontinuously (i.e., categorically) because the relevant articulations are essentially discontinuous; thus, for speakers of English, there is no articulation between /b/ and /d/ and, consequently, no intermediate perception.

Non-categorical perception: experiments with vowels

When phoneme distinctions are produced by articulations that move continuously from one position to another, we should expect continuous perception, and that is, in fact, what we find. Thus, in one study[32] synthetic vowels were varied through a range large enough to include /ɪ/, /ɛ/, /æ/, and this range was divided into the same number of steps that had been used in the earlier study of /b, d, g/. In contrast to the stops, there were with the vowels no peaks at phoneme boundaries, and, as in the case of simple non-speech stimuli, like tones that vary in loudness or duration, discrimination of the vowel stimuli was very much better than identification.

It should be noted, too, that the vowels proved to be more discriminable than the stops—that is, many more intra-phonemic differences could be heard. They were inferior to the stops in that the phoneme labels were

[32] See D. B. Fry, A. S. Abramson, P. D. Eimas, and A. M. Liberman, "The Identification and Discrimination of Synthetic Vowels," (in preparation). For similar results with phonemic tones and phonemic vowel duration (both in Thai) see A. S. Abramson, "Identification and Discrimination of Phonemic Tones," *Journal of the Acoustical Society of America* XXXIII (1961), 842 (Abstract); A. S. Abramson, and J. Bastian, "Identification and Discrimination of Phonemic Vowel Duration," (in preparation).

applied with less consistency. Indeed, the identification of the synthetic vowels was in many cases determined almost completely by the context (of the other vowels) in which the sound was presented.[33]

Some speculations about perceptual properties, distinctiveness, and linguistics

Now what has all this to do with linguistics and the uses of experiment in language description? The point is that experimental procedures not very different from those described earlier in this paper have yielded results which point to important perceptual differences between phoneme classes. An appreciation of these differences may not be necessary for a linguistic description which meets minimum logical requirements, but the differences are none the less interesting or basic.

The obvious implication of these differences is that, other things equal, the stops will be more distinctive than the vowels—that is, they will be more quickly and easily identified on an absolute basis. In the case of the stops, the most acutely perceived differences occur at the phoneme boundaries, which is precisely where they should be for the absolute identification of the acoustic stimuli. Such an arrangement provides an effective basis on which incoming speech sounds can be sorted rapidly and unequivocally into the appropriate "phoneme bins." The result is that in perceiving /d/, for example, the listener is not reduced to judging the particular stimulus as more or less like /d/ than the last sound he heard; rather, he hears it simply as /d/. The perception of the vowels, on the other hand, is more like that of pitch or loudness or any one of many simple stimulus continua. These stimuli are all highly discriminable, but they tend in perception to shade into each other, and it is much easier to judge the stimuli relationally than to identify them absolutely.

The more distinctive a phoneme, the more efficient it will be as a carrier of information, or so we would think. If we can assume that people are motivated to communicate with some reasonable amount of efficiency, we may suppose that the function and fate of a phoneme would be determined, at least in part, by its distinctiveness. In addition, and quite apart from any

[33] The result is probably related to the assumption by Joos (see M. Joos, *Acoustic Phonetics*, Supplement to *Language* XXIV (1948)), that vowels are perceived in terms of a frame of reference, and also to the finding by Ladefoged and Broadbent (see "Information Conveyed by Vowels," *Journal of the Acoustical Society of America* XXIX (1957) 98–104) that the identification of synthetic vowels is influenced by the acoustic characteristics of the carrier sentence. We suspect that the continuous (as opposed to categorical) nature of vowel perception lies behind these examples of the tendency to perceive vowels relationally.

assumption about human motives, it can be argued that a reasonable level of distinctiveness is not merely desirable, but rather necessary if language is to take the form that all human languages do. It is obvious enough that a phonemic system requires by its very nature that the phonemes be identifiable absolutely. (And apparently the nature of man's mental processes requires that all languages be phonemic.) Somewhat less obvious, perhaps, is the psychological necessity that most such identifications be made quickly: if the rate of flow of phonemic information is not above some rather high minimum, the organization of the phoneme units into morphemes, words, and sentences becomes psychologically impossible. These considerations mean that some reasonable fraction of the phoneme elements must be highly distinctive. By measuring the perceptual properties of phonemes in many languages, we should be able to define the limits within which distinctiveness may vary, and in addition, perhaps, uncover some interesting equivalences and constancies.

Other needs and circumstances will, of course, cause elements of rather low distinctiveness to become part of the language, but we should then expect that the relative distinctiveness would be important as a determinant of the linguistic roles that the various phonemes come to play. In this connection we should expect, for example, that the information load borne by a given phoneme or class of phonemes would be determined in significant measure by its distinctiveness. Thus, we might find, for each phoneme, a relationship between distinctiveness and the number of minimal pairings.

In a rather different area, we should suppose that the greatest amount of historical, idiolectal, and dialectal variation would occur among the relatively less distinctive elements. It is possible, too, that not only the amount, but also the kind of variation would be correlated with distinctiveness. Those changes which do occur might well be quantal (i.e., phonemic) or continuous (i.e., merely phonetic) depending on the degree to which the phoneme has the property of categorical perception (and production) we have here discussed.

One other aspect of the experimental findings summarized here has some relevance to the linguist's problems and ought, therefore, to be commented on. We refer to the result which indicates that the interesting perceptual properties of the stops do not inhere in some innately given relation between the sound and its perception, but are, rather, a result of learning. Surely it is not news to the linguist to hear that one's perception of language is influenced by his linguistic experience. The possibly novel point is that it is quite feasible, as we saw earlier, to measure the nature and extent of such influences. These procedures might well yield data that would be of interest to the linguist in connection with the effects of contact between languages, second-language learning, the development of language in children, and like areas; essentially the same procedures might even be applied to the calibration of the linguist as a measuring instrument.

Haskins Laboratories
New York City

University of Pennsylvania
Philadelphia, Pennsylvania

University of Connecticut
Storrs, Connecticut