PHONETICS AND THE DESCRIPTION OF LANGUAGE*

BJÖRN E.F. LINDBLOM

1. INTERPRETATIVE AND PREDICTIVE APPROACHES TO THE STUDY OF LANGUAGE FORM

In the scientific quest for insight, theory construction is an important tool. Involved in the evaluation of a certain theory is the notion of explanation. Explanatory goals have been formulated for many fields of inquiry including the study of language. For segmental phonology — with which we shall be concerned in the present study — concrete proposals as to how such goals might be attained have been explored notably by Chomsky and Halle (1968).

What should be meant by an explanatory theory of phonology? It would be unwise to try to give a comprehensive answer that most investigators concerned would find acceptable. Nevertheless, adopting a sufficiently general point of view it seems clear that such a theory would deal with 'classical' problems of phonology such as SOUND CHANGE (Passy 1890, Martinet 1968) and with the principles underlying for example, SYSTEMS (vowel and consonant inventories) (Jakobson 1968) and CONTRASTS (Jakobson *et al.* 1952, Ladefoged, 1967a, Wang, 1967, Chomsky and Halle 1968, Chapter 7), SYLLABLE STRUCTURES (phonotactic patterns of segments) (Greenberg 1965, Sigurd 1968), or the origins of 'rules', 'rule ordering', 'natural classes', and 'features'.

Is such a phonology at all possible at the present stage? An optimistic answer may be questioned on a number of grounds. Linguists might express doubts because of the enormous task of accounting for the wealth of facts involved, or more importantly, because of limitations of present research strategies. Let us briefly examine the latter motivation, with particular regard to the role that phonetics plays within the influential framework of recent generative phonology (Chomsky and Halle 1968)

* Reworked version of oral paper entitled "Numerical Models in the Study of Speech Production and Speech Perception: Some Phonological Implications" which was presented at the Congress.

The work reviewed in this article is largely a result of teamwork. I would like to thank several members of our research group for their contributions: Si Felicetti, Karin Rapp, Johan Liljencrants, Stefan Pauli, and Johan Sundberg.

The research was supported by the National Institutes of Health under Research Grant NS 04003-08.

and ask whether optimal use is made of it with respect to the long-term goals of phonology.

As a point of departure, take the following statement by Chomsky (1968:12):

> It seems to me that the most hopeful approach today is to describe the phenomena of language and of mental activity as accurately as possible, to try to develop an abstract theoretical apparatus that will as far as possible account for these phenomena and reveal the principles of their organization and functioning, without attempting, for the present, to relate the postulated mental structures and processes to any physiological mechanisms or to interpret mental function in terms of "physical cause"...

The research strategy recommended can be compared with that on which most of *The Sound Pattern of English* (Chomsky and Halle 1968) appears to be based. We find a close agreement. The implications of this philosophy for the status of phonetics in the study of language are evidently that, if used at all, phonetically relevant aspects of physics and physiology are to be applied in an interpretative fashion rather as part of the predictive theoretical machinery of linguistics. Accepting Chomsky's and other linguists' predilection for abstract models, we are led to leave phonetics relatively unexploited in the area of language description and to conceive of it as a science dealing primarily with speech, that is, the physical manifestations of the symbolic devices postulated by the linguist and the relation between these in the behavior of the speaker and listener. This task is of course neither a minor nor an unimportant one. Attempts to construct models of speech production or formulate phonological descriptions with pronunciation specified in terms of an auditorily testable acoustic output (Ladefoged 1967a:58) might provide the linguist with important feedback as to how the grammar — as an account of an idealized speaker's linguistic competence — should be set up (Tatham 1969, Mattingly 1970). Interest in this area concerns the relation between sound and an underlying form of utterance representation. The hope has sometimes been expressed that this level of representation will turn out to be isomorphic with the elements of some level of linguistic representation. In so far as it does, or does not, — an open question at the moment — it will consequently provide evidence that could be used to confirm or refute linguistic hypotheses (Lindblom 1968). The function of the system of phonetic rules set up to relate symbols and sound can be characterized as interpretative: they assign physical shape to underlying utterance representations, or the converse, recover the latter from the acoustic speech signal. The long-term relevance of this work for the psychology of speech production and speech perception remains beyond a doubt. Linguists may admit that eventually, research in this area — which from their point of view concerns the lowest-level, most superficial part of the grammar — may prove to be of importance also to them but after all, they are primarily interested in the structure of language, not in speech and cannot be expected to patiently abide the moment when models of speech production and speech perception reach such perfection that their 'linguistic' relevance begins to shine out in a conspicuously compelling manner. It would not be incorrect to claim that among linguists in general, a lack of interest in phonetics is evident. As ironically noted by Ladefoged (1967a:57): "Much to the discomfort of some phoneticians (and some linguists), phonetics is not a science that linguistics must presuppose". Chomsky and Halle (1968:293) also discuss this topic:

> Many structural linguists have felt that phonetics has very little to offer them and have therefore assigned to it a secondary, peripheral role...As an illustration of this lack of interest in phonetics we may cite the numerous articles on phonological subjects that have appeared in the last thirty years in journals such as the *International Journal of American Linguistics* in which information concerning the phonetic properties of the phonemes of a language is often restricted to a simple listing of alphabetic symbols.

In their own opinion "there can be no question about the relevance of phonetics to the study of language" since "phonetics is concerned with grammatically determined aspects of the signal" (1968:294). Note, however, that "the relevance of phonetics" refers to the relevance of an impressionistic, perceptual, and introspective phonetics. Thus, in spite of Chomsky and Halle's explicit claim to the contrary, we must conclude that phonetic theory plays its traditionally marginal role also in generative phonology for, as Chomsky and Halle make quite clear, beyond the "grammatical" data of introspective phonetics, phonetic facts become irrelevant to the linguist (cf. e.g., discussion of the reality of phonetic representation, 1968:24-26).

According to our own analysis, the reason for the peripheral role that practically all schools of linguistic thought (including generative grammar) have bestowed upon phonetics in the study of language, is deeply rooted in one of the cornerstones of linguistic methodology: the PRIMACY OF LINGUISTIC FORM over the variables of language use and its substantive bases. From a historical point of view the importance and unquestioned validity of this belief — traceable back to de Saussure and still entertained by generative grammarians (cf. quotation above) — is understandable. For it is undeniable that the elimination of speech and language use from the immediate concerns of the linguist once created a productive freedom — freedom from too much phonetic detail. Once the 'linguistically relevant' aspects had been distilled from an infinitely variable speech behavior and postulated as primitives in the theory, sound and phonetic substance could be ignored. Thus it is in complete harmony with this fundamental reasoning that it has been proposed that the study of competence must precede the study of performance for progress to be made in linguistics (cf. Chomsky 1964).[1] It is as a consequence of the primacy of linguistic form that the linguist generally conceives of phonetics as an interpretive, form-based science.[2] Provided that phonetics is viewed in this fashion, the conclusion

[1] "It seems natural to suppose that the study of actual linguistic performance can be seriously pursued only to the extent that we have a good understanding of the generative grammars that are acquired by the learner and put to use by the speaker or hearer. The classical Saussurian assumption of the logical priority of the study of *langue* (and the generative grammars that describe it) seems quite inescapable" (Chomsky 1964:52).

[2] For a particularly good illustration of this point of view, see Katz (1964). He writes: "Now it is clear that the linguist, though he claims that this theory describes a neurological mechanism, cannot immediately translate the theory into neurological terms, i.e. into talk about synapses, nerve

must be that it has a long difficult road to go before its contributions can begin to bear on the complex linguistic issues. On such a view, a lack of one-to-one correspondence between the physical facts and theories of the phonetician and the formal characterization of the linguist is therefore likely to be considered as the problem of the phonetician, not the linguist, also in the future.[3]

Having examined the role of phonetics in present-day linguistics, we can return to our initial question: is phonetic theory used in a maximally efficient manner for purposes of predicting and explaining how language is built? Our own answer, implicit in the preceding analysis, is negative. It is based on the conviction that

(1) the primacy of linguistic form must be questioned;

(2) once it has been rejected, an approach to the study of sound structure that makes more predictive than interpretative use of the knowledge embodied in phonetic theory should become possible;

(3) ultimately, a more comprehensive substance-based theory formalizing the phonetic as well as sociological mechanisms of language use might be developed to predict and explain the nature and historical development of "sound patterns".[4]

fibers, and such. But — and this is the crucial point in showing that the mentalist is not a psychophysical dualist — this failure to have a ready neurological translation means only that he cannot yet specify what kind of physical realization of his theoretical description is inside the speaker's head. Since linguistics and neurophysiology are independent fields, it does not matter for the linguist what kind of physical realization is there. For the purpose of linguistic investigation, it is immaterial whether the mechanism inside the speaker's head is in reality a network of electronic relays, a mechanical system of cardboard flip-flops and rubber bands, or, for that matter a group of homunculi industriously at work in a tiny office. All of these possibilities, and others, are on a par for the linguist as physical realizations of this mechanism, so long as each is isomorphic to the representation of linguistic structure given by the theory of the language. The critical distinction is, then, between an abstract, formal characterization of linguistic structure — the theory itself — and a physical system of some kind which instances this structure. Discovering what kind of a physical system in the human brain instantiates the representation of structure given by a linguistic theory is the task of the neurophysiologist".

[3] In his introductory article in the recent *Manual of Phonetics* (Malmberg 1968), Malmberg concludes his presentation of phonetics as follows: "A combination of a strictly structural approach on the form level with an auditorily based description on the substance level will be the best basis for a scientific analysis of the expression when manifested as sounds. This description has to start by the functional analysis, then it must establish in auditory terms the distinctions used for separating phonemic units, and finally, by means of appropriate instruments, find out which acoustic and physiological events correspond to these different units. The interplay between the different sets of phenomena will probably for a long time remain a basic problem in phonetic research" (1968: 15). This quotation is a good illustration of the belief in the "primacy of linguistic form" as well as the opinion that finding the physical and physiological correlates of phonemic units established by "functional analysis" is what must be assumed, in the absence of an explicitly described supplementary program, to be the major task of phonetics. Note also that the title of the article, "The Linguistic Basis of Phonetics", follows the beaten track whereas the logically equally possible alternative, "The phonetic basis of linguistics" would probably have been taken as indicative of an instance of bad judgment on the part of the author. This is all the more surprising since in the beginning of his studies the linguist might naively entertain the presupposition that some of the boundary conditions under which language develops as a biological, social, and historical product are indeed phonetic. Unfortunately, the serious student of language could not be blamed if upon reading Malmberg's introduction to the field, he concluded that phonetics is of only marginal relevance to phonological theory.

[4] A dissatisfaction with "overly formal" approaches to rules and features is also evident in Chomsky and Halle's own appraisal of the success that generative phonology has had in attaining the desired goals (see Chomsky and Halle 1968: 400-401, and Fromkin 1968, for some relevant remarks).

A predictive use of phonetics as suggested under (2) would aim at making explicit the complex of phonetic 'causes' that constitute part of the origins of phonological structure. Rather than ask questions about the phonetic correlates and psychological reality of phonological elements and processes (thereby directing efforts towards the discovery of phonetic and psychological evidence supporting the positing of such elements and processes), this approach starts at the other end. It starts out from hypotheses about preconditions for speech communication and its development and attempts to derive aspects of phonological structure as a consequence of these hypotheses. We shall devote the next section of our paper to sketching a fragmentary and tentative version of such a theory. About the framework outlined in (3), we shall have very little to say. In a way, this is an answer to one of the questions raised initially: is explanatory phonology possible? Evidently, we must conclude that much research remains to be done and it seems necessary, first of all, to investigate the possibilities of reevaluating the notion of "linguistically relevant phonetic fact" and of theories founded on the substantive bases of language use and acquisition. A beginning is to ask, why just postulate and interpret linguistic form, try to predict it.

2. PHONETIC TOOLS IN PHONOLOGY

1. *A point of departure: language form as shaped by acquisition and use.* — We shall assume then that the framework of which phonetic considerations form a part is that more comprehensive theory y, that is, a complete model of the styles of speech behavior characteristic of an entire speech community. Following Weinreich *et al.* (1968) we are careful to speak of STYLES of speech behavior convinced that it is correct to look upon language not as a homogeneous object but as a structure consisting of many coexistent systems. The variation associated with such 'orderly heterogeneity' falls only within limits set by the phonetic variables of language use. These variables provide raw materials that can be selected and incorporated as regular features of the grammar of the speech community for purposes of social differentiation. The need for such differentiation has been shown to arise according to various sociological mechanisms and criteria (age, sex, class, prestige...) (Labov 1965, Sommerfelt 1968) and has been proposed as a factor involved in linguistic change (Weinreich *et al.* 1968). In this perspective, one of the tasks of the linguist is to make explicit precisely how linguistic structure is a function of phonetic mechanisms and boundary conditions, and what the phonetic input to a model of the sociological mechanisms of speech communication can be.

In the past, explanatory phonetic or so-called functional principles have often been proposed in treatments of phonological subjects. They have not, however, been formalized and systematically explored in linguistic theories, perhaps chiefly because of the 'structuralist' bias of these theories (cf. above). We have in mind the principle of LEAST EFFORT and the principle of MAXIMAL PERCEPTUAL CONTRAST. Variants of these principles can be found in Passy's discussion of sound change (1890) as pointed out by Martinet (1955) who makes extensive use of them himself, cf. also Jespersen (1941, esp. pp. 30-31) and Wang (1968a). In Jakobson (1968) the former is rejected, whereas the latter is used to account for the development of phonological contrasts in children. Clearly these principles have proved viable although they may occasionally have fallen into disrepute because no serious attempts have been made to quantify them (cf. Jakobson 1968:21). The position of the present work is that renewed attempts should be made at analyzing these now admittedly vague notions and — in case they turn out be to accessible to rigorous theoretical treatment — assessing their explanatory power in relation to other factors. They are associated with the behavior of two language users: the speaker whose tongue tends to be lazy whenever possible and the listener who demands perceptual differentiation of speech stimuli. These perhaps often conflicting demands on speech communication appear not only in adult use but in particular during language acquisition as a result of maturational processes. Most authorities are agreed that acquisition proceeds towards gradually greater differentiation of various language skills. This is apparently the case with phonological contrasts (Jakobson 1968), syntactical patterns (MacNeil 1966) as well as with concept formation (Vygotsky 1962). The growth of motor and sensory processes appears to follow a similar pattern: from the general to the specific. According to the interpretation of Lenneberg (1967:324-325):

> With maturation, the neonate begins to organize the perceptually available stimuli surrounding him and also to organize the movements of his muscles. Sensory data become grouped into as yet undifferentiated, global classes of gross patterns, and these, subsequently, become differentiated into more specific patterns. Similarly, movements which at first involve the entire body become differentiated into finer motor patterns. Both the perceived patterns and the self-produced patterns of movements become organized or grouped in functional categories, and hierarchies of categories. Members of a particular category are functionally equivalent because they either elicit an identical response or they serve one and the same function within the over-all structure of a particular behavior pattern. It is these general principles of differentiation and categorization that appear in specialized form in verbal behavior. They influence the organization of perceived material as well as the organization of the motor output.

Similarly, Bolinger (1969) suggests that "genericness" as a linguistic phenomenon may be a special case of general, "all-purpose" perceptual and cognitive mechanisms. Along the same lines, Hardy (1970) states that "it has been established for some time that the auditory discrimination of speech events improves as a function of age (e.g., Templin 1957)" and that "stereognosis of the oral cavity has been shown to improve with age (McDonald and Aungst 1967)". Although in "Kindersprache" Jakobson explicitly denies the relevance of substantive explanations in terms of least physiological effort (1968:21) and sensory maturation (1968:22), and attributes the observed parallels between the sound systems of the languages of the world and those of children to internal, structural factors, his discussion is not devoid of substantive arguments. According to a rather free interpretation of Jakobson's material, it might be suggested that the motor, sensory and psychological inmaturity of the developing child constitutes a time-varying 'BOTTLENECK' through which language must pass. That the maturational constraints constituting this bottleneck should be a peculiarity in the history of language use, leaving no traces in its design, seems to be a possible point of view but, as convincingly demonstrated by Jakobson, an unlikely one. Rather, from an evolutionary point of view we should expect language structure to have been shaped by the "filter characteristics" of the acquisition process and to be subject to optimization with respects to the constraints of the development of language and speech communication in the child. It does not seem controversial to suggest that contrasts and structures that the child discovers, produces and uses easily and at an early stage, should stand a better chance of 'slipping through' and remaining in the language. Assume then that language is shaped by acquisition and use, and that there are various hypothetical boundary conditions that suggest themselves both with respect to developmental processes and adult speech communication. If language were a result of making 'optimal' use of the given resources (maximizing communicative efficiency, attaining 'maximum effect with minimal means' etc.), what would it be like, what structuring would the hypothetical principles entail, how close to human languages would such derived signal systems be, etc.? To sound realistic, such a research proposal presupposes that a quantitative model is available in terms of which the derivations of "structure" can be made. Below we shall discuss some preliminary steps in the development of such a model: the problems of finding an adequate representation of the speech mechanisms and the quantification of boundary conditions. We shall examine some consequences of two such conditions: ARTICULATORY EASE and PERCEPTUAL DIFFERENTIATION (cf. Jespersen 1941, Martinet 1955, 1968).

2. *Numerical models: the Mechanisms of Speech.* — In modeling the mechanisms of speech production and speech perception, investigators naturally look for shortcuts through the overwhelming amounts of physical and physiological data. In certain technological and commercial applications of speech research, extreme decisions are sometimes encountered: as long as the synthesizer or recognition device reaches a certain level of practical performance it is regarded as acceptable no matter how unsatisfactory the scheme happens to be as a description of human behavior. In the context of language study a boundary between relevant and irrelevant facts must be drawn in a different manner owing to the scientific character and psychological orientation of the research. The notion of 'linguistically relevant physiological and physical fact' should no doubt be interpreted rather generously. But at some point, depending on the current state of information, simplification and abstraction will set in. In work on models of speech production, efforts appear at present to deal with e.g., the problems associated with describing vocal tract geometry (Mermelstein, Maeda, and Fujimura 1970, Ladefoged *et al.* 1971), the interplay between the varia-

bles of glottal behavior, phonation and aerodynamic processes (Stevens 1971). Gradually the modeling of speech production mechanisms proceeds 'upstream', that is, by comprising stages of the speech signal at a successively smaller distance from higher neural centers. However, specificational frameworks based at articulatory, muscular, and neural levels cannot as yet be considered as part of established phonetic knowledge in spite of recent progress in these areas. For instance, the representation of vocal tract shape is one of the thorny problems being debated to-day. Finding the natural degrees of freedom of the vocal tract is difficult, partly because of the requirement that the specification be psychologically correct and linguistically revealing, partly because of the great complexities that the physics of the tongue and other components offer. Below we shall explore some of the linguistic implications of an articulatory model recently proposed. Much of the remaining discussion will be based on this model. It has been described in detail elsewhere (Lindblom and Sundberg 1971a), so we shall limit the presentation to some major points.

The model (henceforth the *LS* model) defines a procedure for deriving a set of formant frequencies from information on the state of the lip muscles, the position of the jaw, the position of the tongue tip, the shape and position on the tongue body and larynx height. The steps involved in the computations are shown in Figure 1.

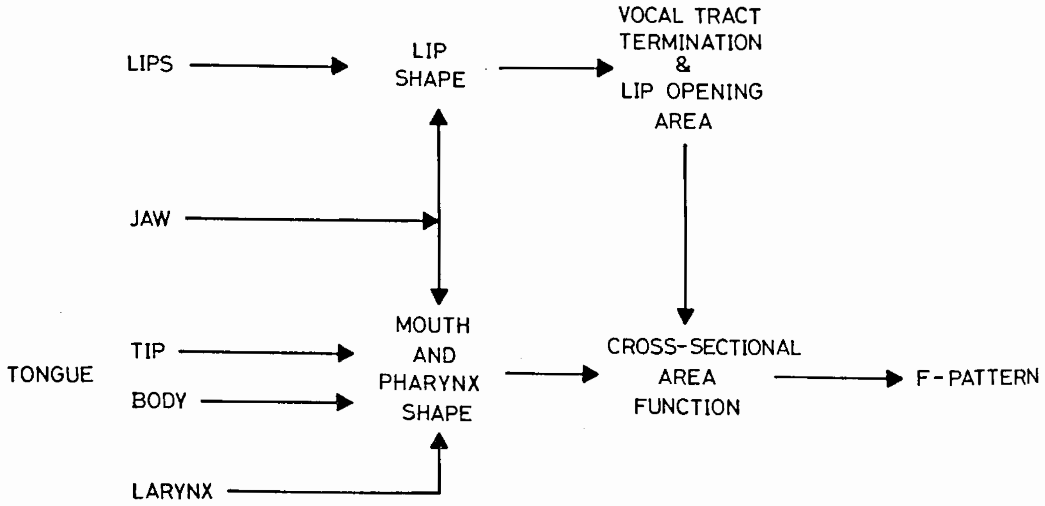

Fig. 1. Model relating articulatory parameters to vocal tract shape and area function, and to acoustic transfer function.

A comparison with the classical three-parameter models of Stevens and House (1955) and Fant (1960) reveals some similarities as well as some novel features. The *LS* model is similar to the previous models in that the PLACE of the tongue constriction and the DEGREE of this constriction can be controlled individually. However, the parameters of the *LS* model generate a jaw-based tongue contour whereas the three-parameter models use parameters more directly related to the

cross-sectional area function. The introduction of the jaw and the lips as separate parameters, and the particular definition of the tongue tip and tongue body parameters, are novel features, and will be illustrated and to some extent justified in the following discussion.

When faced with phonetic data e.g., a speech spectrogram or X-ray data on articulatory movement, an investigator typically asks: what features does the talker control and what aspects are secondary and predictable? These questions may lead him on to investigating the role of SERVO-MECHANISMS in speech production. This question has received a great deal of attention recently (Öhman 1967, MacNeilage 1969, Hardcastle 1970, Hardy 1970, Ringel 1970, Leanderson 1972). In seeking a balance between considering too many physical and physiological facts at great depth and gaining too little insight by choosing too superficial a description, it may be helpful, also in studies of speech production, to remember the often-quoted statement that the nervous systems thinks in terms of movements rather than in terms of muscles. In other words, events are in general output-oriented, directed towards the achievment of some goal external to the system. Examples of such activity can be found e.g., in the patterns of muscle recruitment that were observed by Ladefoged (1967b) during a fairly long single-breath passage of connected speech. The function of these muscular activity patterns seems to have been the maintenance of a relatively constant subglottal pressure. The reorganization of fundamental frequency regulating laryngeal muscle activity in various motor contexts reported by Ohala and Hirose (1969) is another case in point. Our own research group has recently done some work that appears to throw light on the topic of output-oriented mechanisms in speech production (Lindblom and Sundberg 1971b). A classical phonetic experiment was performed. Sustained vowels were recorded with the subject's mandible in a controlled position. For this purpose subjects held a small block (calibrated to produce 5, 15, 25 mm openings) between their teeth. Spectrographic measurements of formant frequencies were made at the moment of the first glottal pulse, a point in time at which there would not yet be any possibility of auditory feedback. The experiment showed that for all vowels and all externally controlled jaw positions, the formant patterns sampled at the onset of phonation approached the normal values rather closely. Since mandibular movement has been shown to have a considerable effect on formant frequencies (Lindblom and Sundberg 1971a), the independence of formant frequency values on jaw position can be taken as evidence of COMPENSATORY ARTICULATION. These results suggest the hypothesis that the central nervous system can make use of afferent information to control the spatial location of articulatory components. Phonation did not set in until an approximation of the target area function had been achieved. No diphthong-like approaches to formant targets were observed on the spectrograms. The experiment and an interpretation of it are summarized in Figure 2. According to this diagram three hypothetical components are attributed to the speech production system: a sensory component capable of providing information on the actual shape

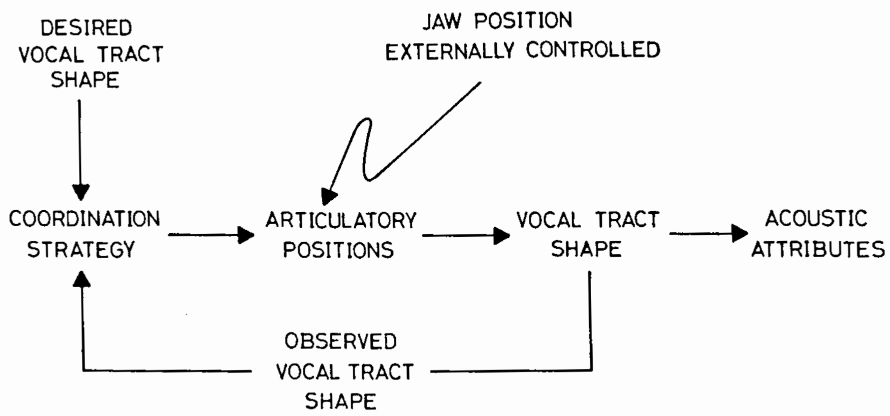

Fig. 2. Schematic representation of experiment on compensatory articulation.

of the vocal tract, a component providing a representation of the desired vocal tract shape and a strategy component that coordinates the articulatory muscles on the basis of these specifications. Furthermore, the results of the experiment can be taken to indicate that the sensory control is based neither on AUDITORY FEEDBACK nor on a GAMMA-LOOP coding in terms of invariant muscle fiber lengths. The latter inference is possible in view of the compensatory adjustment of articulatory position that is required to maintain the cavity shapes essentially constant in spite of an abnormal jaw position (cf. the 'superpalatalized' tongue shape needed to preserve the area function of an [i] pronounced with a 25 mm jaw opening constant, Figure 3). It is therefore likely that a spatial representation of the desired positions of the articulators is available and serves as reference in the nervous system, and that the receptors involved are primarily tactile and of a touch-pressure type. This interpretation is not incompatible with what is known at present about oral stereognosis (Hardy 1970, Ringel 1970). Naturally is does not exclude the possibility that in normal speech additional feedback mechanisms might be operative and if so, in flexible cooperation with the one suggested by the present experiment. It also remains to be established to what extent the present findings are representative of a larger number of talkers including, for instance, children and persons suffering from various speech disorders. The implication of the results obtained so far for articulatory models of speech production is that the notion of SPATIAL REFERENCE is a feature worth considering in future attempts to make such models more realistic from the point of view of motor control and neurophysiology (MacNeilage 1970, Nooteboom and Slis 1970).

3. *Some Boundary Conditions*: *Perceptual Differentiation and Articulatory Ease.*— We shall use the *LS* model to simulate some aspects of the development of speech. For expository reasons and for lack of information we shall discuss three stages separately but do not imply that in reality the development proceeds in terms of such discrete, non-overlapping phases. Figure 4 shows a schematic representation of the

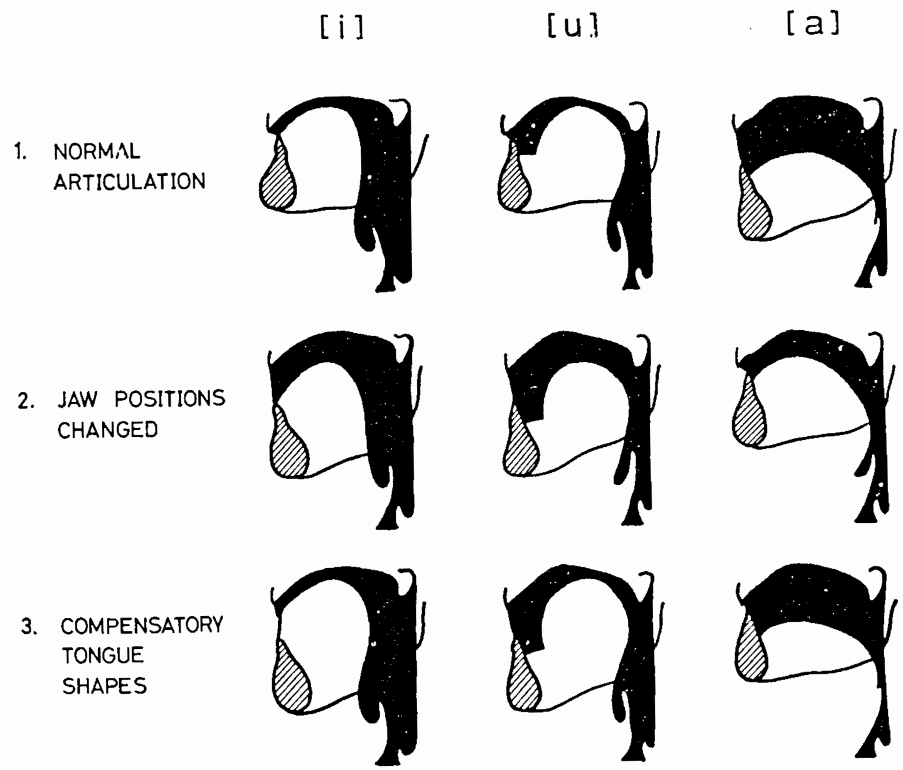

Fig. 3. Profiles of articulatory model for [i], [u], and [a]. The top line (1) shows normal articulations. The middle tracings (2) illustrate the modifications that the cavity shapes of these vowels undergo as the jaw positions are changed. The bottom row (3) shows compensatory tongue articulations necessary to restore the cavity shapes in (1) when jaw positions correspond to (2).

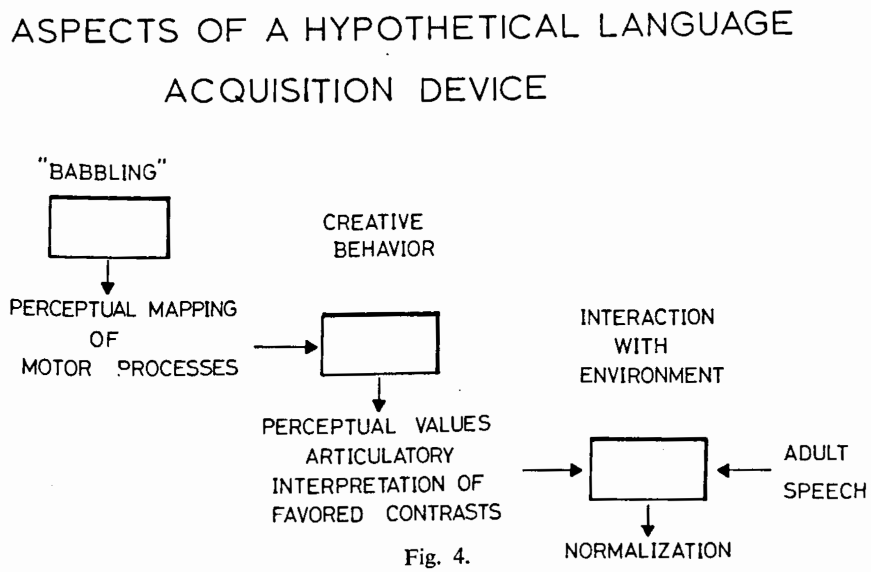

Fig. 4.

three aspects selected: babbling, 'creative' behavior, and interaction with the environment. The babbling period is the stage when the child is supposed to be "capable of producing all conceivable sounds" (Jakobson 1968:21). During this stage it might in principle arrive at a perceptual mapping of articulatory gestures by exploring its phonetic capabilities in a more or less random fashion. Both of these assumptions have not yet been examined in sufficient detail and merit further attention. If nevertheless, a random search is assumed to be characteristic of this period, some of its consequences can be studied with the aid of the *LS* model. In a two-dimensional representation, an F2-F1 diagram, Figure 5 shows the margins of the acoustic vowel space that an exploration of the articulatory possibilities of the model delineates. A periodic glottal sound source is assumed. Also shown is a formant trajectory

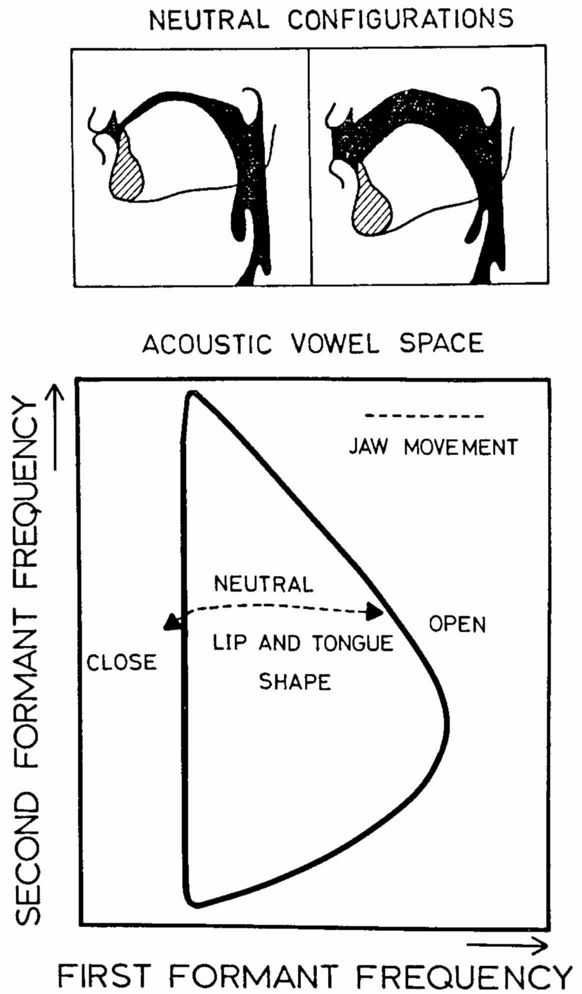

Fig. 5. The acoustic vowel space of *LS* model. The dashed line pertains to neutral lips and tongue shape, the arrows being associated with the model configurations at the top.

corresponding to a single 'uncoordinated' articulatory gesture that results when only the jaw is varied within its range (5-23 mm) while all other parameters are kept at constant neutral values. It is interesting to note that under such neutral conditions and when the jaw is lowered to 23 mm (upper right, Figure 5), a formant pattern of $F1 = 570$ Hz, $F2 = 1565$ Hz, and $F3 = 2500$ Hz is observed. The auditory value of a vowel sound with those formant frequencies is close to [æ] for an adult male talker. As the jaw is gradually raised, it changes into 'mid, central' and 'high, central' qualities. A certain similarity with so-called linear vowel systems (Trubetzkoy 1958) is apparent. Elevating the jaw still further with the lips and the tongue still in their neutral positions, produces a stop closure. For a jaw opening smaller than 4 mm the *LS* model will exhibit a 'double articulation'. The lips will close (Lindblom and Sundberg 1971a, Figure 8) and there will also be an apico-alveolar constriction (Figure 5, upper left). Depending on the state of the velum, opening and closing the jaw will result in noises that auditorily come close to [æ p æ p æ], [æ b̥ æ b̥ æ], or [æ m æ m æ], provided that glottal vibration and neutral lip and tongue shapes are assumed. As a result of spreading the lips and some slight perturbation of the neutral tongue shape, these sounds might easily change into [æ t æ t æ], [æ d̥ æ d̥ æ], or [æ n æ n æ]. It is tempting to compare these neutral articulations with Jakobson's remarks on the early phonological contrasts developed by children.

At the beginning of the first stage of language development, the acquisition of vowels is launched with a wide vowel, and, at the same time, the acquisition of consonants by a forward articulated stop. An *a* emerges as the first vowel, and generally a labial as the first consonant, in child language. The first consonantal opposition is that of nasal and oral stop (*e.g., mama-papa*), which is followed by the opposition of labials and dentals (e.g., *papa-tata* and *mama-nana*) (1968: 47-48).

It seems as if the *LS* model could in principle be used as a basis for predicting the 'unmarked' character of these syllables and contrasts. However, why should jaw movement rather than tongue or lip movement be selected as the favored uncoordinated gesture? We leave this question open for the moment. Possibly perceptual contrast, acoustic stability, and articulatory non-specificity and reproducibility can be shown to play a role. Nevertheless, if it is assumed that the ability to optimize the use of the speech mechanisms with respect to various perceptual and articulatory conditions is an 'innate endowment', it follows that, in response to its communicative needs, the child might use this ability to facilitate the spontaneous discovery of elementary signals. Let us examine such CREATIVE behavior in somewhat more formalized terms. Suppose that two algorithms were devised: one for discovering an optimal set of vowel qualities for a system consisting of *n* vowels; another for discovering the optimal pronunciations of these vowels. Let the optimal vowel quality system be defined with respect to the maximization of perceptual contrast among the *n* vowels. Let optimal pronunciation be defined with respect to a criterion of minimal expenditure of articulatory energy (to be discussed later).

4. *'Perceptual Contrast'.* — Seeking a quantitative formulation of the problem

we must make certain assumptions about the acoustic space universally available for vowels, the perceptual representation of this space and the quantitative interpretation of contrast. The built-in articulatory constraints of the *LS* model delimit the range of vowel sounds that we should expect to be possible in human speech. An approximate representation of this space is shown in Figure 6. The left graph

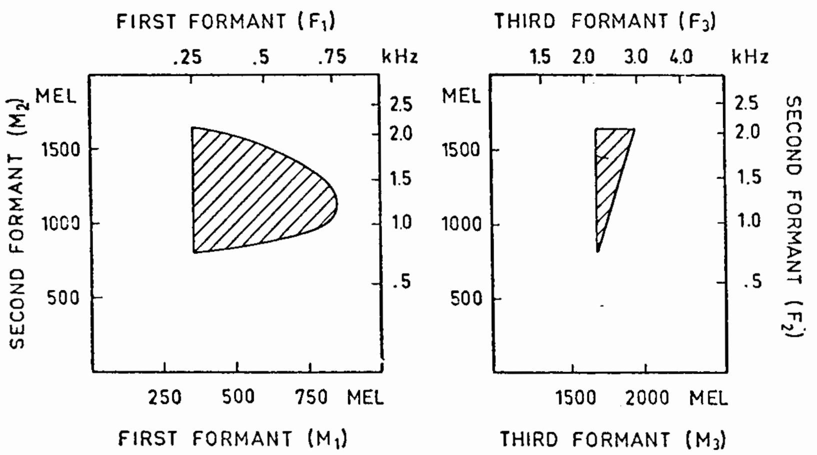

Fig. 6. Stylized version of acoustic space in three dimensions. This representation was used in the computation of 'perceptual contrast' for vowels.

depicts the space in terms of the first and second formants; the right diagram in terms of the third and the second formants. The axes have mel scales as well as the corresponding kHz values. By the transformation of the linear frequency scales into mel scales, the space assumes a shape that is more satisfactory from an auditory point of view. Let the calculations of contrast be carried out using this 'perceptual' representation. Make the degree of contrast between two arbitrary vowels depend on the linear distance in mel units between the points representing those vowels, the criterion used to maximize intervocalic distances, or perceptual contrast, being

$$\sum_{1}^{m} 1/r_i^2 \rightarrow \text{minimized} \qquad (1)$$

where r refers to the distance between the i:th pair of vowels and the number of pairs per system is $m = n(n-1)/2$ where n is equal to the number of vowels in the system. To make computations somewhat more tractable let r be interpreted as a two-dimensional distance within the hypotenuse plane of the space (Figure 6).

This procedure has in fact been implemented on a digital computer. For a description of the procedures, computer program, and some preliminary results, see Liljencrants and Lindblom (forthcoming). The output of the program is a set of formant frequency values. This algorithm can thus be regarded as an automatic procedure for finding a set of n vowel qualities (sounds) that have been optimized

with respect to a certain tentative definition of the concept of perceptual contrast. If some such procedure is built into our language acquisition device it will be possible for this device to discover perceptual values for vowel contrasts SPONTANEOUSLY without relying on an analysis of the speech stimuli of its environment. Figure 7

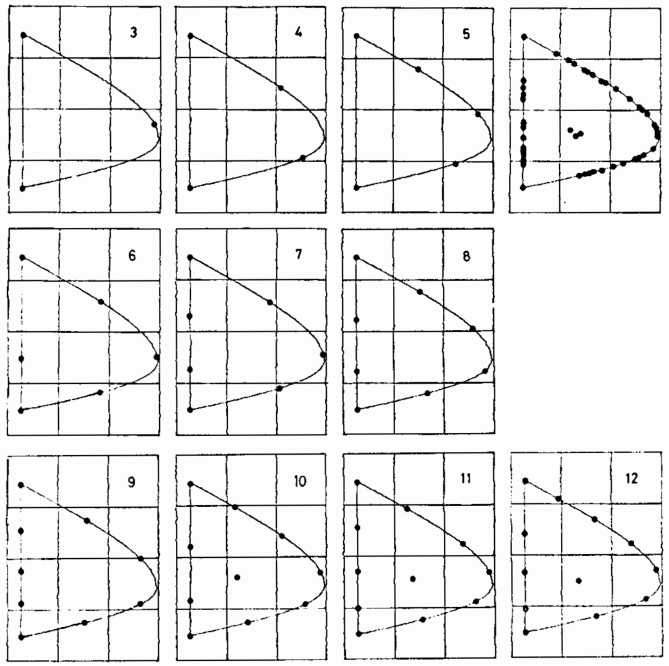

Fig. 7. Results of computations of formant frequencies for vowel systems.

shows the results of the preliminary calculations for the vowels of systems ranging in size from three through twelve vowels. The formant frequency data are presented in terms of linear frequency scales for $F2$ (ordinate) and $F1$ (abscissa). The horizontal and vertical lines correspond to divisions of $F1$ at every 200 Hz from 200 to 800 Hz and of $F2$ at every 500 Hz from 500 to 2500 Hz. In the study by Liljencrants and Lindblom (forthcoming), these results were evaluated in terms of a broad transcription. It was concluded that in the case of three-, four-, five-, six-vowel systems, the predictions came close to patterns of vowel qualities reported for various languages. For larger systems discrepancies were observed with respect to at least one of the qualities. In particular, it was found that a major deficiency of the model was its inability to generate an [ø]-like vowel and its predilection for "close, central" vowels such as [ɨ] and [ʉ]. The 'broad' interpretations of the systems predicted in that study are given in the Appendix.

5. '*Articulatory Ease*'. — Let us now consider the second algorithm dealing with the optimization of pronunciation with respect to the criterion of 'minimal expenditure of articulatory energy'. Suppose for simplicity that the procedure described

above has indicated three formant patterns corresponding perceptually to [i], [a], and [u]. To generate these patterns with the model, certain cavity shapes, or cross-sectional area functions, must be produced with an accuracy set by the limits of acoustic sensitivity. The fixed-mandible experiment teaches us that in principle these area functions can be made in great many ways by human subjects. Thus it is possible to produce an [i] by an abnormal lowering of the jaw and to compensate by raising the tongue more than usual, by 'superpalatalizing', as it were (cf. discussion above and Figure 3 on p. 73). Similarly for [u] and [a]. To accommodate the facts of compensatory articulation the *LS* model was extended by increasing the range of the neutralization parameter, that is, the parameter representing the deformation of the tongue shape from its neutral shape (Lindblom and Sundberg 1971a). With this extension 'supershapes' could also be generated. Such extreme shapes are illustrated at the bottom of Figure 3. Extreme compensatory tongue shapes for [i], [a], and [u] are shown also at the right-hand side of Figure 8. This figure shows more normal configurations to the left and the interplay between the tongue and jaw parameters necessary to have the model maintain area functions appropriate for [i], [a], and [u].

Characteristic of the supershapes is the antagonism between the tongue muscles

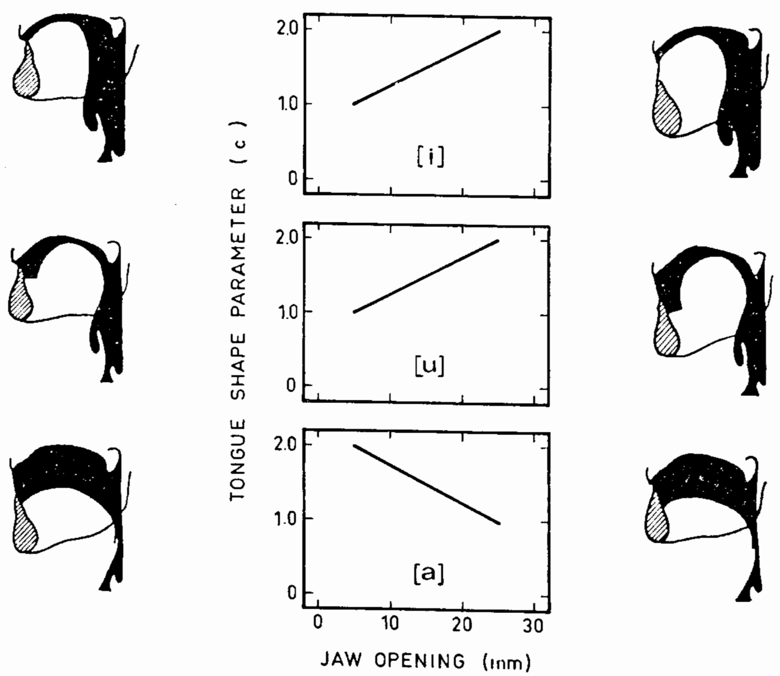

Fig. 8. Interplay between the degree of tongue shape deformation and jaw position as specified by *LS* model. An ordinate value of zero corresponds to neutral tongue shape.

and the jaw muscles. In a compensatory variant of [i] pronounced with wide jaws (top right, Figure 8), the depressor muscles of the mandible and the genioglossus assisted by various other muscles, such as the mylohyoid, oppose each other. Similarly in an 'open' [u], the labial musculature and the styloglossus, among other muscles, participate in overcoming the pull of the jaw muscles. And in a 'close' [a], activity exceeding what is normal in the hyoglossus and pharyngeal constrictors is likely to occur. It is obvious that these 'supershapes' are not physiologically optimal and we should expect the system to strive for a minimization of such antagonism. The information contained in Figure 8 helps us answer the following question: for any given formant pattern or area function, what is the optimal position of the mandible? Let us define optimal position as that entailing minimum physiological energy expenditure, or equivalently, minimum displacement of lips and tongue from their neutral positions. This task amounts to finding the jaw position associated with the minimum value of the ordinate or tongue deformation parameter. Figure 8 shows that this optimization procedure selects a close jaw for [i] and [u] and an open jaw for [a]. Consequently, we can conclude that the model in conjunction with the boundary conditions explains the origin of the open-close feature in vowels. This brings us to the first major point of our presentation. Note that the explanation that we propose does not rely on an a priori postulation and explicit mention of features as we know them from generative phonology. Instead the simple theory that we have explored predicts vowel contrasts both with respect to perceptual values and pronunciation. It does so on the basis of a numerical representation of the speech mechanisms and two boundary conditions. It is clear that it may not yet do so perfectly but still, there is a fair amount of success.

We summarize the computational experiments with our preliminary LAD in Figure 9. Again we want to emphasize that the discrete order in which the two algorithms have been applied must not be taken to imply that we attribute the same discrete order to human children. It seems highly likely that not only perceptual but also articulatory, physiological criteria may guide the child in its search for its first contrastive signals. Furthermore, as the general-to-specific development of motor skills and discriminatory ability progresses, it is clear that the number of contrasts that the child can manage will increase. This is reflected by incrementing n, the number of vowels per system, in the computations. However, the procedure used has no memory and makes no assumption about the utilization of past phonological conquests and experience. Clearly the simulations could be improved in a number of important ways. However, referring to the model as a LAD may be justified in spite of its present primitive character since, (1) the procedure outlined is automatic, which reflects the ability of human children to develop language creatively and not exclusively by analyzing the speech output of its environment; (2) the strategy is based on the principles of articulatory 'ease' and perceptual differentiation which, apart from their a priori plausibility, have been shown above to produce some reasonable results.

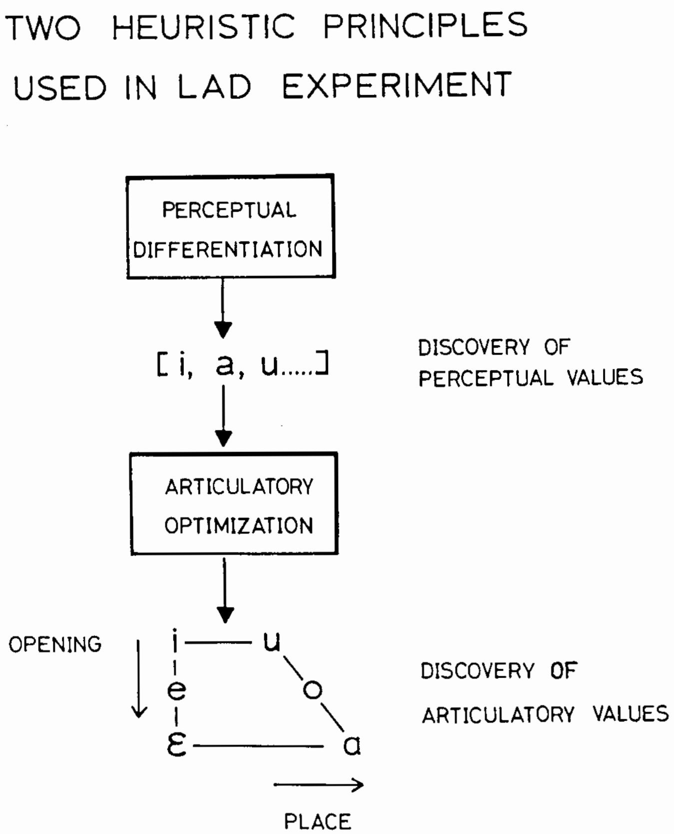

Fig. 9.

7. *Normalization.* — It may of course be argued that the criterion of maximal perceptual differentiation should not be attributed to the process of individual language acquisition but should be regarded as a factor of long-term linguistic evolution which has emerged in speech communities as a safe-guard against 'noise' and various factors degrading the distinctiveness of the speech signal. This view would imply that the principle has no psychological reality. The observations of the course of the sensory and motor maturation of the child seem to deprive this claim of some of its force. Also the interaction of the child with its adult environment is worth considering at this point. The material presented by Peterson and Barney (1952) and Fant (1966) can be used to draw attention to the well-known fact that the absolute formant frequency values differ for adult male speakers and children. Their vowel spaces overlap partially. Since in fact they overlap, it would in principle be possible for the child to match the absolute values of its formant frequencies to those of the grown-up speakers for adult vowel formant patterns that the space of the child encloses. So far no such partial adjustment has been reported. It is customary to note that the child arranges its vowel sounds in the acoustic space in such a way

that on a RELATIVE physical basis the vowel systems of children and adult speakers become similar. Its utilization of the space available for vowels is thus similar to the output of the algorithm for the maximization of perceptual distance which would tend to push vowels towards the periphery of the space. The physical variability associated with speaker-dependent differences in vocal tract size offers a problem of NORMALIZATION with which the child learns to cope. We know that he does cope with it since vowel qualities are categorized and 'equated' in spite of physical differences. The manner in which this process is pictured will differ depending on whether the child is credited with an ability to discover basic vowel contrasts spontaneously or is assumed to rely only on an analysis of what he hears. Suppose that there is such creative behavior and that the child is capable of developing a system of basic vowel contrasts on his own. The process of normalization should then be facilitated since, in addition to involving judgments of perceptual similarities between individual pairs of vowels, it can be based on discovering similarities between two systems and the 'function' that the elements have within the systems. As has been pointed out many times 'sameness' of vowel quality refers also to the phonological function of the segments. At present, the background of this normalization is not well understood. However, assumptions about creative behavior and perceptual differentiation may open up some research possibilities.

8. *Timing.* — Needless to say there is a great deal more to the development of phonological contrasts than the aspects considered so far. For instance, in spite of its relative simplicity, the pronunciation of isolated vowels offers many problems that the motor system must solve. One of these is illustrated in Figure 10: the coordination of phonatory and articulatory activity. Take a monophthongal vowel of a certain language and ask a native, mature speaker to pronounce it in isolation. It is a rather remarkable fact about this type of utterance that, in most cases, the vowel has its intended perceptual value from the onset of phonation and that formant frequencies will be close to their target values at this point. This seemingly trivial observation can be used to infer that phonation and articulation are fairly well synchronized, at least voicing does not begin till the articulators are very near their goals. Should we try to model this ability, it would become immediately clear that such synchronization presupposes a control system with some degree of sophistication. It seems highly likely that such a mechanism is available to the child only at a later stage of motor and sensory maturation. Whatever the exact information on this aspect of speech development, we shall assume that like many other speech production skills, the coordination of phonatory and articulatory gestures develops from a general, random, into a more specific pattern. The phonetic consequences of varying the relative timing of these activities may be studied with the *LS* model.

9. *Neutralization.* — Figure 10 shows schematically how all articulatory parameters leave their neutral positions (in synchrony) and gradually approach their target values appropriate for the vowel in question. Phonation is shown to begin and to end in three different ways relative to the articulatory activity: voicing is

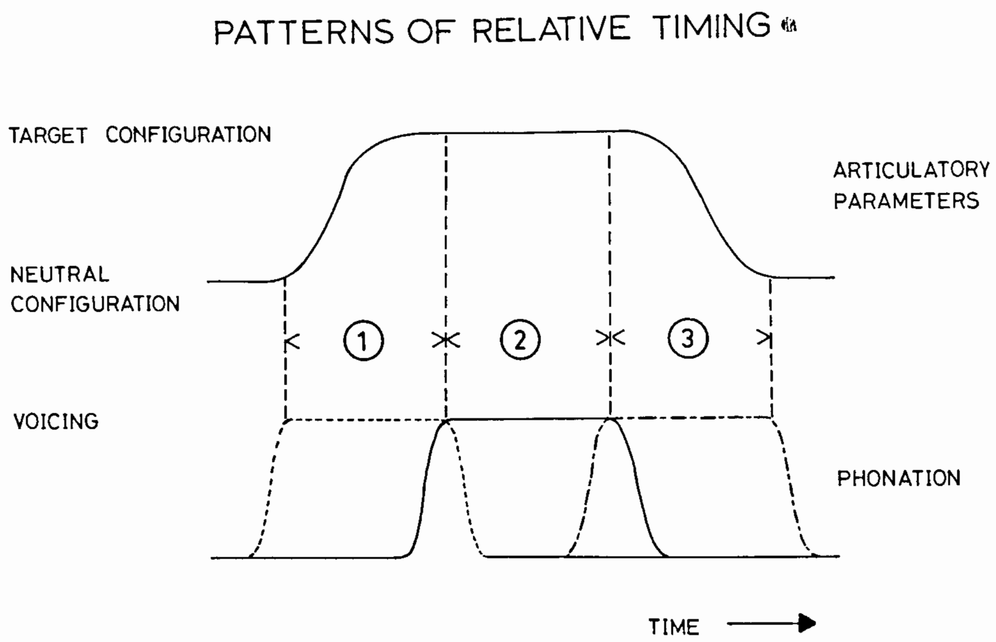

Fig. 10. Upper curve refers to an articulatory movement (lip and tongue) from a neutral configuration to a given vowel target and then back to neutral shape. The lower curve shows three different time locations for a segment of voicing.

present during (1) the articulatory on-glide, (2) steady-state, or (3) off-glide. The formant shifts associated with conditions (1) and (3) as well as the steady-state values corresponding approximately to the qualities of long Swedish vowels can be inferred from Figure 11. It demonstrates the acoustic consequences of assigning neutral shape to lips and tongue for all target configurations. In this diagram the jaw parameter is kept unchanged for each vowel. The arrows indicate the direction of the neutralization which, when complete, changes the vowel formant frequencies into patterns located along the dashed line. As can be seen, "back" vowels such as [u], [o], and [ɑ] are changed into [u, ʉ]-, [ø, œ, ə]-, and [æ]-like qualities whereas the 'front' vowels, as they approach the same continuum, undergo a slight 'raising'. Incomplete neutralization thus changes [i] into a more [e]-like sound. Condition (3) of Figure 10 has the acoustic correlates indicated by the arrows of Figure 11. Reading the arrows 'backwards' corresponds to condition (1).

10. *Diphthongization.* — There are many phonetic topics that usually involve considerations of reduction and neutralization. Let us examine one of them in some detail: DIPHTHONGIZATION. First of all, condition (3) will generate 'centering' and 'opening' diphthongs with schwa-like off-glides differing in 'opening'. Condition (1) will generate the mirror images of these diphthongs. It is of considerable interest to note that the latter sounds approximate Scanian diphthongs rather closely. The

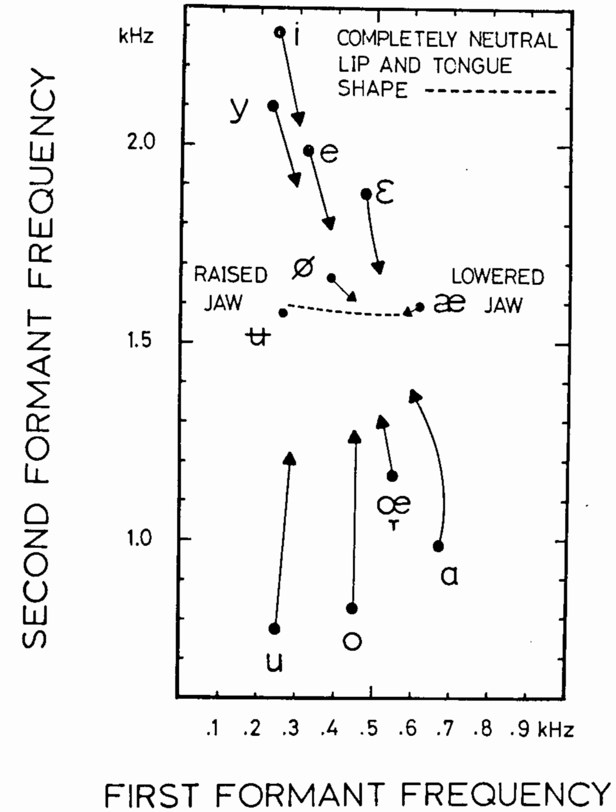

Fig. 11. Filled data points refer to model-generated vowels approaching the qualities of long Swedish vowels. The arrows indicate the direction of formant shifts induced by moving lips and tongue towards neutral positions.

vowels of this dialect of Swedish are described in auditory terms by Bruce (1970) as follows:

/i:/ → [ei]	/y:/ → [øy]	/u:/ → [eu]
/e:/ → [ɛe]	/ʉ:/ → [øʉ]	/o:/ → [ɛo]
/ɛ:/ → [æɛ]	/ø:/ → [œø]	/ɑ:/ → [æɑ]

Vowels synthesized according to condition (1) are perceptually similar to natural Scanian diphthongs and the acoustic basis of this impression can be somewhat clarified by comparing Figure 11 with Bruce's data (Figure 12). In fact these diagrams can be taken as partial confirmation of Bruce's own interpretation of this phenomenon (Bruce 1970:9-10). It must be regarded as particularly strong support of this interpretation that the neutralization hypothesis predicts that a vowel such as [æ] will not undergo radical change. Although this vowel is not dealt with by Bruce it does not seem to diphthongize in the varieties of Scanian familiar to the present author.

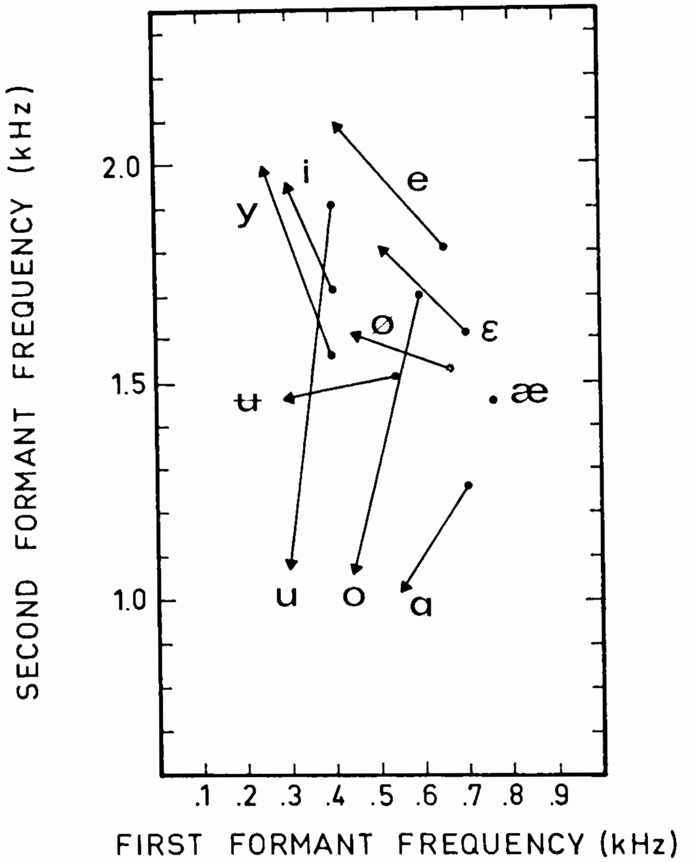

Fig. 12. Direction of formant shifts for Scanian diphthongs (after Bruce). Note the qualitative similarity of these trajectories and those of Figure 10.

Note the 'front' bias of the qualities associated with the neutral vowel continuum. Neutralization will tend to 'front' back vowels whereas front vowels undergoing the same modification will not become correspondingly 'back'. This asymmetry brings to mind certain observations on the asymmetry of phonological processes. According to Labov (1965), the fronting of back vowels is a more common historical change than the backing of front vowels. I-umlaut seems to be a case in point, *A*- and *U*-umlauts involving primarily lowering and rounding (Haugen 1969).

A full exploration of the notion of neutralization as defined in terms of the present model falls outside the scope of this presentation. We can conclude that the phonetic variation derived here as a result of neutralization and coordination of phonation and articulation constitutes examples of 'unmarked' processes. The model provides a basis for saying why certain alternations and diphthongs are more natural than others. These results, although preliminary, suggest new directions in which a formalism for capturing "naturalness" in phonology might be sought, and appear to recommend the incorporation of a neutralization dimension in models of speech production.

3. SOME IMPLICATIONS FOR PHONOLOGICAL THEORY

It has been argued that, so far, linguists have left phonetics relatively unexploited in the area of language description. An analysis of this circumstance appears to indicate that a fundamental assumption of linguistic methodology is one of the important factors behind this state of affairs: the primacy of linguistic form over the variables of language use. As a productive alternative point of departure, a 'substance-based' approach is proposed. It differs from 'form-based' methods in that linguistic form is not postulated but derived as a consequence of the structuring that substantive conditions impose on the speech signals. In such a framework, 'psychological realities' are not introduced to justify abstract structures and processes once these have been postulated (cf. generative grammar and its relation to early versions of "psycholinguistics" [Bever 1968][5]). Rather, phonetic and psychological realities form the starting point. This vantage point brings out the necessity to find a formalism suitable for the description of such conditions. An adequate numerical representation of the speech mechanisms must be devised and various phonetic and communicative boundary conditions must be quantified. On the preceding pages we have attempted to show that a "substance-based" approach is indeed possible and shows promise of producing results of great relevance in the study of language. We can summarize the results presented so far in the following fashion.

In introducing the *LS* model we demonstrated that it can be used to define the acoustic space universally available for vowels. Since it is an articulatory model it also delimits a hypothetical class of possible vowel articulations. Consequently, the manner in which a model of this sort is organized has a number of consequences that bear on linguistic universals and other topics. Setting all parameters except the jaw to zero or 'neutral' values, we find that the vowel qualities generated resemble those of 'linear vowel systems', the most open vowel being [æ]-like. Under such neutral conditions, jaw movement and vocal fold vibration will sound like [æpæpæ] or [æb̥æb̥æ]. The 'unmarked' character of these utterances is evident (cf. Jakobson 1968).

We then used the acoustic space of the *LS* model to develop an algorithm that could be used to predict the phonetic structure of vowel systems. It is of linguistic interest to compare the output of this algorithm with the phonetic structure of vowel systems for various languages[6] and with the gradual mastery of vowel contrasts in

[5] After this paper had been written the present author came across a few articles dealing precisely with the form-substance problem at issue here. Examples are Kiparsky (forthcoming) which discusses 'formal' versus 'functional' explanations in phonology and the paper by Bever and Langendoen (1971) on the role of perceptual constraints in the historical development of English relative clauses. Many of the points raised in the present paper appear to be in agreement with the arguments and evidence put forward in favor of physical models in phonology by Ohala in a recent report (1971).

[6] Such a comparison is presented in Liljencrants and Lindblom (forthcoming). No claim can be made to the effect that perceptual differentiation is the ONLY determinant involved in shaping a phonological structure of such complexity as a vowel (or consonant) system. A fully satisfactory account will no doubt necessitate consideration of phonetic, social, as well as historical factors.

the developing child. The 'creative' strategy of this algorithm draws attention to the possibility of similar 'spontaneous' behavior during human acquisition and to the 'psychological reality' of the principle on which the simulation is based, maximization of perceptual contrast. It does not seem unmotivated to regard this principle as a manifestation of a very general maturational process whereby the development of behavior proceeds towards greater differentiation, from the general to the specific. A further application of the *LS* model is illustrated in an attempt to quantify the notion of articulatory ease. It is shown that this condition accounts for the origin of the open-close feature in vowels.

1. *Linguistic Relevance of Phonetic Facts.* — The results obtained indicate clearly that the facts that must be considered 'LINGUISTICALLY RELEVANT' go far beyond the data of '*Ohrenphonetik*' and the introspective phonetics that generative as well as other schools of linguistics have so far been content to make use of. If it is the task of a theory of language to account for the origin of phonological features — clearly, it must be — seemingly trivial, unimportant facts must be given formal status in that theory. Take the open-close dimension discussed above. To predict the universal occurrence of this articulatory characteristic, the anatomical path along which mandibular movement takes place in speech comes to the foreground. Lowering involves not only a vertical component but also a displacement in a dorsal direction. As a result, the position of the mandible that would be optimal for achieving a pharyngeal constriction as for [a] would be an open one since that position would require minimal deformation of the neutral tongue shape. Clearly, the formalism necessary to explain why the openness of [a, ɑ, æ] is a universal cannot be obtained by an extension of the frameworks suggested so far in phonological theory since a numerical representation of the vocal tract and the notion of articulatory effort are necessarily involved. Similarly, the acoustic properties of the universal vowel space cannot be correctly derived unless the theory contains a representation of the natural degrees of freedom of articulation and their acoustic consequences. There are thus numerous, strong indications that it is necessary to make linguistic statements a function of phonetic theory. Physical phonetics is indeed a science that linguistics must presuppose. Any attempt to avoid facing this claim and comfortably ignore phonetic substance must reduce the power of any theory of language, either as a social phenomenon *(langue)* or as an individual, mental, derivative representation of this phenomenon (competence).

2. *The Non-Orthogonality of Phonological Features.* — The circumstance that the vowel space tends to be horse-shoe shaped, that is, narrower at the [+low] than at the [+high] end, is reflected in the number of vowel qualities that languages can use contrastively at these points (e.g., [i, y, ʉ, ɯ, u] as compared with [ɑ, a, æ]). This well-known fact exemplified the '2^n' problem inherent in any distinctive feature framework using *n* binary feature dimensions. Such a system is in principle adequate for distinguishing '2^n' segments, but is far too rich. Since feature dimensions are interdependent — for example, [+lateral] presupposes [+coronal] and [+nasal] rules out the possibility of a radical pharyngeal constriction, etc. — such systems must be supplemented with conventions that reduce their capacity. The function of these conventions is to increase the precision with which the overly rich binary system describes phonetic realities. A more natural way of reflecting the asymmetries inherent in the 'phonetic capabilities of man' is suggested above: the incorporation of a quantitative model of these capabilities in the theory.

3. *Are Phonological Features Binary or n-Ary?* — The data on Scanian diphthongs brings up this issue. Bruce considers both a binary and an *n*-ary treatment. Essentially it is proposed that diphthongization is brought about by a rule inserting a glide segment before the vowel. In the binary solution which is based on Wang's feature framework (1968b), the inserted element is [+palatal, $-\beta$high, αmid], the vowel being [α high, β mid]. In the *n*-ary solution the glide is [+palatal, (n+1) high], the vowel being [n high]. The desired phonetic shapes can be generated with either solution, provided that labialization is also considered and lexical representations are chosen appropriately. The *n*-ary treatment appears to capture a generalization that the binary solution fails to reflect. However, neither version accounts for the occurrence of [+palatal] in the rule. From the point of view of the formalism, [—pal] might also have been expected. The experiment with the *LS* model suggests an entirely different approach to the description of these phenomena. There seem to be two major questions. On the one hand, what is the phonetic basis of the process? Why this particular pattern of phonetic variation? On the other, how does the process find its way into the language? The first question was dealt with previously when it was demonstrated that a single unitary process may be one of the chief factors underlying the asymmetric data: neutralization and a delayed approach to target. The second question concerns the mechanisms according to which a speech community creates its styles of speech behavior from various continua of phonetic variation. This is probably to a large extent a sociolinguistic problem. On this view, the description of a certain phonological process such as the Scanian diphthongization rule should not be justified with reference to its psychological reality and representation in the mind of an individual speaker. Rather, it should be evaluated with respect to how well it elucidates the psycho-phonetic and social origins of the process. The description of the ability of mature, individual speakers to relate sound and meaning, their grammar and competence, is of course different from describing and explaining the phenomenon of language (= '*la langue*') as a complex product of phonetic, psychological, grammatical, social, and historical factors. For discussion relevant to the assumption that individual speakers "operate as if they were, in a certain sense, internal reconstruction devices which reconstruct nonphonetic underlying forms and the processes which lead from such forms to an eventual phonetic output" (Chafe 1970:37), see also Ladefoged (1971).

4. *The Correct Set of Features.* — During the past decades many distinctive feature frameworks (Jakobson *et al.* 1952, Chomsky and Halle 1968, Wang 1968b) have been proposed and their interpretation in terms of phonetic facts has been widely

debated (Ladefoged 1964 and 1967a, Lieberman 1970, Fant 1971). How many dimensions should we use? What is their physical basis? In general, observations of segments that are contrastive in a certain language lead to the identification of a feature dimension. Given a phonological contrast, the phonetic interpretation of this feature can begin. As is well known, there are many problems connected with such attempts. Is the identified dimension a new feature or should it be equated with an old one? Examples of such inquiries are the features of tense-lax, advanced tongue root and covered for vowels (Halle and Stevens 1969), or fortis-lenis, tense-lax, voiced-voiceless for consonants (Kim 1965). Another point which was brought up by Wang (1968) concerning vowel features is the number of heights that the framework should be capable of accommodating. Wang's proposal to use two binary dimensions [high] and [mid] is preferable to using [compact], [diffuse] or [high] and [low] since four distinctive heights can be distinguished and the Procrustean conventions [+low] → [—high], [+high] → [—low] need not be used to reduce the overcapacity of the system (Chomsky and Halle 1968:405).

Note that in the material presented above the computations presuppose no EXPLICIT mention of features such as rounded, open, close, front, back, etc. Neither do they necessitate any assumptions about number of 'heights'. These labels can be introduced to describe the articulatory output of the LAD but have no formal status in the theory. In other words, they are IMPLICIT consequences of the constraints and optimization criteria that form part of the theory. Similarly, the phonetic analysis of the Scanian data shows clearly that an account of the 'naturalness' of this particular type of diphthongization can be given that presupposes neither an explicit mention of features nor any of the other formal devices used in generative rules.

5. *Where Do Features and Rules Come From?* — The present results, although preliminary, appear to indicate how the problems associated with the non-orthogonality and binary and/or *n*-ary character of phonological features might be overcome. Consequently, the question arises concerning the true origin of both features and rules in generative phonology. To what extent are the (binary or *n*-ary) explicit features that we use in phonological description isomorphic with elements that the human mind abstracts (much like the linguist) and weaves into its command of a given sound pattern? To what extent are explicit features descriptive devices that, up to a point, provide a convenient way of recording facts about phonological contrast but which characterize the psychological and phonetic basis of contrast so superficially that a number of pseudo-problems arise? In so far as the latter interpretation is correct, it follows that it would not be worthwhile to look for a set of explicit features, binary or multivalued, such that measures could be defined on this set that would reflect hierarchical relations among segments and the naturalness of classes and rules. Rather, to capture naturalness and hierarchies, it would be necessary to go considerably more deeply into the physics and psychology of contrast than we have done so far with shorthand notational devices such as the explicit features of generative phonology. Similarly, the well-known question arises to what extent

phonological rules are psychologically genuine generalizations characteristic of the competence of the individual speaker or to what extent they are regularities in the medium of spoken language (cf. Ladefoged 1971b). In so far as the latter is the case, our present rule notation provides little insight into the origins of the processes in question.

Further research will be necessary to clarify these issues. However, the material presented in this article demonstrates that they are real issues and reinforces the remarks made initially about the necessity to abandon the interpretative role of physical phonetics and to make it part of the predictive theoretical machinery of linguistics.

Department of Speech Communication
Royal Institute of Technology (KTH), Stockholm, and
Department of Phonetics
University of Stockholm

APPENDIX

Phonetic values of predicted vowels according to Liljencrants and Lindblom [forthcoming].

i u a	i ü/ʉ ɨ u ε ɔ a
i u ε a	i ü/ʉ ɨ u ε ɔ æ a/ɑ
i u ε ɑ æ/a	i ü ʉ ɨ u ε ɔ æ ɑ
i ʉ u ε ɔ a	i ü ʉ ɨ u e o/ɔ ε/æ a ɑ

i ü ʉ ɨ u
e
ə ɔ
ɛ/æ
a ɑ

i ü ʉ ɨ u
e
ə ɔ
ɛ
æ a ɑ

REFERENCES

Bever, T.G.
1968 "A Survey of Some Recent Work in Psycho-Linguistics", mimeographed.
Bever, T.G., and D.T. Langendoen
1971 "A Dynamic Model of the Evolution of Language", *Linguistic Inquiry* 2:433-464.
Bolinger, D.
1969 "Genericness: A 'Linguistic' Universal?", *Linguistics* 53:5-9.
Bruce, G.
1970 "Diphthongization in the Malmö Dialect", *Working Papers* 3:1-9 (Phonetics Laboratory, Lund University).
Chafe, W.L.
1970 *Meaning and the Structure of Language* (Chicago and London, The University of Chicago Press).
Chomsky, N.
1964 "Current Issues in Linguistic Theory", in *The Structure of Language*, J.A. Fodor and J.J. Katz, eds. (New York, Prentice-Hall), pp. 50-118.
1968 *Language and Mind* (New York, Hartcourt, Brace and World).
Chomsky, N. and M. Halle
1968 *The Sound Pattern of English* (New York, Harper & Row).
Fant, G.
1960 *Acoustic Theory of Speech Production* (The Hague, Mouton and Co.) (2nd edition, 1970).
1966 "A Note on Vocal Tract Size Factors and Non-uniform F-Pattern Scalings", *STL-QPSR* 1:22-30 (Stockholm).
1971 "Notes on the Swedish Vowel System", in *Form and Substance (Phonetic and Linguistic Papers Presented to Eli Fischer-Jørgensen*, L.L. Hammerich, R. Jakobson, and E. Zwirner, eds. Copenhagen, Akademisk Forlag), pp. 259-268.
Fromkin, V.
1968 "The What and Why of Intrinsicalness", *Working Papers in Phonetics* 10:156-168 (UCLA).
Greenberg, J.H.
1965 "Some Generalizations Concerning Initial and Final Consonant Sequences", *Linguistics* 18:5-34.
Halle, M. and K.N. Stevens
1969 "On the Feature 'Advanced Tongue Root'", *Quarterly Progress No.* 94:209-215 (Research Laboratory of Electronics, Massachusetts Institute of Technology).
Hardcastle, W.
1970 "The Role of Tactile and Proprioceptive Feedback in Speech Production", *Work in Progress* 4:100-111 (Department of Linguistics, University of Edinburgh).
Hardy, J.C.
1970 "Development of Neuromuscular Systems Underlying Speech Production", Paper presented

at the Conference on 'Speech and the Dento-Facial Complex', New Orleans, February 1970.
Haugen, E.
1969 "Phonemic Indeterminacy and Scandinavian Umlaut", *Folia Linguistica* III:107-119.
Jakobson, R.
1968 *Child Language, Aphasia, and Phonological Universals* (The Hague, Mouton & Co.).
Jakobson, R.. G. Fant, and M. Halle
1952 "Preliminaries to Speech Analysis: The Distinctive Features and Their Correlates", *Technical Report No.* 13 (Acoustics Laboratory, Massachusetts Institute of Technology).
Jespersen, O.
1941 *Efficiency in Linguistic Change* (Copenhagen, Munksgaard).
Katz, J.J.
1964 "Mentalism in Linguistics", *Language* 40:124-137.
Kim, C.
1965 "On the Autonomy of the Tensity Feature in Stop Classification", *Word* 21:339-359.
Kiparsky, P.
forthcoming "Explanations in Phonology", to appear in *Goals of Linguistic Theory*, S. Peters, ed.
Labov, W.
1965 "On the Mechanism of Linguistic Change", *Georgetown University Monographs on Languages and Linguistics* 18:91-114.
Labov, W., and B. Wald
1969 "Some General Principles of Vowel Shifting", Paper presented at the 44th Meeting of the Linguistic Society of America.
Ladefoged, P.
1964 *A Phonetic Study of West African Languages* (Cambridge, Cambridge University Press).
1967a "Linguistic Phonetics", *Working Papers in Phonetics* 6 (UCLA).
1967b *Three Areas in Experimental Phonetics* (London, Oxford University Press).
Ladefoged, P., J.L. deClerk, and R. Harshman
1971 "Factor Analyses of Tongue Shapes", *Working Papers in Phonetics* 19 (UCLA).
Ladefoged, P.
1971 "The Limits of Phonology", in *Form and Substance (Phonetic and Linguistic Papers Presented to Eli Fischer-Jørgensen*, L.L. Hammerich, R. Jakobson, and E. Zwirner, eds. (Copenhagen, Akademisk Forlag).
Leanderson, R. and A. Persson
in press "The Effect of Trigeminal Nerve Block on the Articulatory EMG Activity of Facial Muscles", *Acta Otolaryngologica*.
Lenneberg, E.
1967 *Biological Foundations of Language* (New York, The Wiley and Sons).
Lieberman, P.
1970 "Towards a Unified Phonetic Theory", *Linguistic Inquiry* I:307-322.
Liljencrants, J. and B. Lindblom
forthcoming "Numerical Simulation of Vowel Quality Systems: The Role of Perceptual Contrast", accepted for publication in *Language*.
Lindblom, B.
1968 "On the Production and Recognition of Vowels", doctoral thesis (University of Lund).
Lindblom, B. and J. Sundberg
1971 "Acoustical Consequences of Lip, Tongue, Jaw and Larynx Movement", *Journal of the Acoustical Society of America* 50:1166-1179.
forthcoming "Neurophysiological Representation of Speech Sounds", in *Proceedings of the XVth World Congress of Logopedics and Phoniatrics*, Buenos Aires, Argentina, 1971.
MacNeil, D.
1966 "Developmental Psycholinguistics", *The Genesis of Language* F. Smith and G.A. Miller, eds. (Cambridge, Mass., The M.I.T. Press).
MacNeilage, P.F.
1970 "Motor Control of Serial Ordering of Speech", *Psychological Review* 77:182-196.

MacNeilage, P., R. Krones, and R. Hanson
1970 "Closed Loop Control of the Initiation of Jaw Movement for Speech", *Journal of the Acoustical Society of America* 47:104(A).
McDonald, E.T. and L.F. Aungst
1967 "Studies in Oral Sensorimotor Function", in *Symposium on Oral Sensation and Perception*, J.F. Bosma, ed. (Springfield, Ill., C.C. Thomas), pp. 202-222.
Malmberg, B.
1968 "The Linguistic Basis of Phonetics", in *Manual of Phonetics*, B. Malmberg, ed. (Amsterdam, North-Holland Publ. Co.), pp. 1-16.
Martinet, A.
1955 *Économie des changements phonétiques* (Berne, A. Francke).
1968 "Phonetics and Linguistic Evolution", in *Manual of Phonetics*, B. Malmberg, ed. (Amsterdam, North-Holland Publ., Co.), pp. 464-487.
Mattingly, I.
in press. "Speech Synthesis for Phonetic and Phonological Models", in *Current Trends in Linguistics, Vol. XII*, T.A. Sebeok, ed. (The Hague, Mouton & Co.). (Also in *Status Report on Speech Research, Haskins Laboratories, SR*-23:117-149. (1970).
Mermelstein, P., S. Maeda, and O. Fujimura
1970 "Description of Tongue and Lip Movement in a Jaw-Based Coordinate System", Paper U1 presented at the 80th Meeting of the Acoustical Society of America.
Nooteboom, S.G. and I. Slis
1970 "A Note on the Degree of Opening and the Duration of Vowels in Normal and 'Pipe' Speech", *IPO Annual Progress Report* 5:55-58.
Öhman, S.E.G.
1967 "Peripheral Motor Commands in Labial Articulation", *STL-QPSR* 4:30-67 (KTH, Stockholm).
Ohala, J.
1971 "The Role of Physiological and Acoustic Models in Explaining the Direction of Sound Change", *Monthly Internal Memorandum*, (May): 57-74 (Phonology Laboratory, Project on Linguistic Analysis, University of California, Berkeley, Calif.).
Ohala, J. and H. Hirose
1969 "The Function of the Sterno-Hyoid Muscle in Speech", *Reports of the* 1969 *Autumn Meeting of the Acoustical Society of Japan* (Tokyo), pp. 359-360.
Passy, P.
1890 *Études sur les changements phonétiques et leurs caractères généraux* (Paris).
Ringel, R.L.
1970 "Oral Sensation and Perception: A Selective Review", Paper presented at the 'Conference on Speech and the Dento-Facial Complex', New Orleans, February 1970.
Sigurd, B.
1968 "Phonotactic Aspects of the Linguistic Expression", in *Manual of Phonetics*, B. Malmberg, ed. (Amsterdam, North-Holland Publ. Co.), pp. 450-463.
Sommerfelt, A.
1968 "Phonetics and Sociology", in *Manual of Phonetics*, B. Malmberg, ed. (Amsterdam, North-Holland Publ. Co.), pp. 488-501.
Stevens, K.N.
1971 "Airflow and Turbulence Noise for Fricative and Stop Consonants: Static Considerations", *Journal of the Acoustical Society of America* 50:1180-1192.
Stevens, K.N. and A.S. House
1955 "Development of a Quantitative Description of Vowel Articulation", *Journal of the Acoustic Society of America* 27:484-493.
Tatham, M.A.A.
1969 "On the Relationship Between Experimental Phonetics and Phonology", *Occasional Papers* 5:20-26 (University of Essex, Language Centre, Colchester).
Templin, M.C.
1957 *Certain Language Skills in Children* (Minneapolis, Minn., University of Minnesota Press).

Trubetzkoy, N.S.
1958 *Grundzüge der Phonologie* (Göttingen, Vandenhoeck and Ruprecht).
Vygotsky, L.
1962 *Thought and Language* (Cambridge, Mass., The M.I.T. Press).
Wang, W.S.-Y.
1967 "Phonological Features of Tone", *International Journal of American Linguistics* 33:93-105.
1968a "The Basis of Speech", *Project on Linguistic Analysis Reports (second series)* 4 (University of California, Berkeley).
1968b "Vowel Features, Paired Variables, and the English Vowel Shift", *Language* 44:695-708.
Weinreich, U., W. Labov, and M.I. Herzog
1968 "Empirical Foundations for a Theory of Language Change", in *Directions for Historical Linguistics*, W.P. Lehmann and Y. Kalkiel, eds. (Austin, University of Texas Press).

DISCUSSION

LADEFOGED (Los Angeles)

As often before, Dr. Lindblom, in an excellent paper, has shown us the way in which experimental phonetic studies may contribute to our understanding of phonological problems. I am particularly impressed by techniques he has suggested for setting up vowel systems in which the individual vowels are perceptually as far apart as possible. I would like to propose alternative criteria that can be used in establishing perceptual distances, but in doing so, I am building on the foundations he has laid.

Lindblom bases his estimates of perceptual similarity on the relations between the pitches of the formants of each vowel. There is no doubt that the pitch of the first formant is the single most important feature of perceptual vowel quality, probably because about 80% of the total energy of most vowels is in the region of the first formant. But the pitches of the second and third formants are not in themselves appropriate dimensions for assessing the perception of vowels.

Recent experiments at UCLA (by Terbeek, Harshman, Lindau and myself) have indicated that both in producing a given set of vowels, and in assessing the perceptual similarity between vowels, the second most important feature is the distance between the first and second formants. I would imagine that Lindblom's model for predicting vowel qualities would be even more successful if it represented each vowel just in terms of two dimensions: (1) the first formant, and (2) the distance between the effective second and first formants.

Phoneticians usually describe vowels in terms of the highest point of the tongue, and in terms of the degree of lip rounding. The two measures we have been discussing correlate well with two of the traditional terms, in that formant one is strongly correlated with so-called vowel height, and the effective formant two minus formant one is correlated with the traditional front-back dimension. But there is no uniform acoustic correlate for the third traditional dimension, the degree of lip rounding. It is also interesting that Lindblom notes that his acoustic model fails to generate front rounded vowels. Nevertheless the degree of lip rounding is undoubtedly an important attribute of vowels, playing a major role in many phonological rules. Since it has no uniform auditory or acoustic correlates, it is at least possible that it is simply an articulatory property. It does not seem at all unlikely to me that listeners and speakers (and languages) organize two aspects of vowel quality in auditory/acoustic terms and have a third feature, quite separate from the other two, which is organized in articulatory terms.

TATHAM (Colchester)

Dr. Lindblom has reached the conclusion that insufficient exploitation is made of phonetic theory in explaining or predicting the sound patterns of language. He lays, if my reading is correct, at least part of the blame at the feet of phonetics. It is interesting that Ohala, in his paper yesterday afternoon, laid the blame at the feet of phonology whilst essentially making the same claims about the correct use of phonetics in providing not the interpretive rôle that it plays almost exclusively, but a more explanatory rôle.

There can be no doubt that within the theory itself phonetics HAS explanatory ambitions and there can be no doubt either that the theory itself these days stands on a firm foundation of data collected and assembled one way or another FOR THE PURPOSES OF CONSTRUCTING THAT THEORY. However, there are those who still think that being 'scientific' is all that is necessary — where 'being scientific' means practising a rather crude and narrow policy of empiricism of which Lindblom is obviously not guilty. There is surely more to science than that. There is, of course, every reason for the inclusion of abstractions in theories — indeed that is usually what building theories entails, and there is also every reason for the remarks of Chomsky, which Lindblom quotes, proposing the formalizing of competence before performance.

Let me return to the three convinctions Lindblom has as a basis for his conclusion that phonetic theory is insufficiently exploited as an explanatory construct impinging on phonology:

(1) the primacy of linguistic form must be questioned;

(2) once it has been rejected, an approach to the study of sound structure that makes more predictive than interpretive use of the knowledge embodied in phonetic theory should become possible;

(3) ultimately, a more comprehensive substance-based theory formalizing the phonetic as well as sociological mechanisms of language use might be developed to predict and explain the nature and historical development of 'sound patterns'.

I do not know whether outright rejection of the primacy of linguistic form is what is required — maybe a partial rejection of some of the more *ad hoc* founded areas, but I do not think these are as many as might seem. It may be within the very nature of a formal approach committed to being as little fragmentary as possible that apparently ill-founded notions turn up.

To conclude, let me say that a more substance-based theory of phonology would undoubtedly result from Lindblom's proposals. What I think we must systematically avoid is the danger of basing phonology — which IS not phonetics — on the DATA of phonetics — believing erroneously, I assert, that somehow that data is more real or more respectable in some sense. What I think needs underlining is that if phonology is to draw more on phonetics, or if phonetics is to be more conscious of its predictive rather than interpretive power, then it is not the simple DATA of phonetics which is to be offered to phonology, but data via the THEORY of phonetics which must be constructed in an appropriate way.

Finally, let me underline Peter Ladefoged's remark on the importance of this paper by saying that I think we shall come to look back on Dr. Lindblom's address as a landmark in the development of our discipline.

WAJSKOP (Bruxelles)

Le rapport présenté par le Dr. B. Lindblom constitue une contribution fondamentale pour les progrès de la phonétique. Le développement de notre discipline dépend dans une très large mesure, comme l'a souligné le Prof. D.B. Fry, d'études approfondies sur le fonctionnement réel de la production et de la perception de la parole. Le point de vue défendu par le rapporteur peut paraître hardi, il a le mérite de s'appuyer sur une infrastructure expérimentale et théorique indéniable.

(1) Si je me rallie aux critiques générales énoncées par B. Lindblom vis-à-vis d'une linguistique dont le caractère de plus en plus abstrait l'éloigne chaque jour davantage des données recueillies par l'expérimentateur (cf. Fant 1969); si on peut s'accorder avec lui à propos de sa critique concernant le statut marginal accordé à notre discipline, le danger est cependant grand de fonder une linguistique sur une base purement phonétique qui ne pourrait traiter ni expliquer les faits de syntaxe, de sémantique et de sociologie. On risque d'assister à une glissement vers un 'phonétisme', qui, bien que plus sévèrement articulé, ne serait pas très éloigné de la phonétique atomiste du XIXe siècle à moins que le propos du rapporteur ne conduise à créer un modèle linguistique prédictif qui présente les mêmes avantages que le modèle phonétique qui nous a été soumis.

(2) La puissance prédictive du micro-modèle proposé paraît des plus efficaces sur le plan articulatoire si l'on en juge par les résultats de l'auteur et par les communications présentées entre autres par MM. Lisker, Ladefoged et Jonasson. Le modèle permettrait une taxonomie des systèmes vocaliques indépendante du principe binariste qui s'est révélé incapable de résoudre un certain nombre de difficultés (cf. "French Phonemic Pattern", Jakobson, Lotz et Malmberg 1961).

(3) Par contre, la liaison proposée entre modèle de la production et modèle de la perception me paraît schématique. Si le principe de contraste perceptif a été effectivement négligé, il est ici surtout présenté comme un axiome. On peut cependant ajouter que le phénomène de la localisation spatiale des composantes articulatoires mis en relief grâce aux expériences sur la compensation s'ajusterait de manière adéquate au phénomène du cadre vocalique interne. Les données recueillies par l'équipe du S.T.L. (Stockholm) indiquent dès à présent que les voyelles suédoises

ne sont pas réparties arbitrairement dans un espace physique mais qu'elles sont organisées en séquences linéaires avec une nette tendance à un espacement équidistant dans un espace perceptif. Cette conclusion est parfaitement compatible avec le principe du contraste maximum.

PADDOCK (Wolfville, N.S.)
I would like to support Prof. Ladefoged's suggestion that Dr. Lindblom should investigate the possibility of using the difference between $F2$ and $F1$ as the basis for one of his primary auditory or perceptual dimensions for vowels. Eli Fischer-Jørgensen's claim (1968) that this difference seems to be the basis of a predominant dimension of vowel quality is supported by my own experimental work with Russian sharping (= palatalization) and Arabic flatting (= 'emphasis') at University College London.

In particular, my spectrographic analysis of Egyptian Colloquial Arabic revealed that $F1$ and $F2$ approach each other closely at least once during the duration of any 'emphatic' syllable in that dialect. Native speakers identified synthetic CV syllables as 'emphatic' whenever the difference between $F1$ and $F2$ was reduced below a value of about 400 mels on a pitch scale.

Note that I refer here to synthetic vowels in which $F2$ is not usually at the same frequency as in the corresponding natural vowel but is adjusted (upward) so as to produce a sensation in the listener which is roughly equivalent to the sensation which he probably derives from $F2$ plus some higher formant or formants in many natural vowels. Ladefoged's $F2$ minus $F1$ formula is therefore only a crude approximation to the frequency which corresponds to this sensation of higher pitch in a vowel. However, Fant (1959) has proposed a formula for calculating this 'effective' frequency.

Additional evidence for this postulated 'difference' dimension comes from Dr. Fourcin's paper which was presented here this morning. He showed that the pitch heard in noise containing spectral peaks depends on the spacing of such peaks rather than on their absolute frequencies.

Further discussion of these questions is found in Paddock (1970).

REFERENCES

Fant, G.
1959 "Acoustic Analysis and Synthesis of Speech with Applications to Swedish", *Ericsson Technics* 15:3-108.
Fischer-Jørgensen, E.
1968 "Perceptual Dimensions of Vowels", *Zeitschrift für Phonetik* 21:94-98.
Paddock, H.J.
1970 *Lingua* 25:142-151.

FROMKIN (Los Angeles)
Dr. Lindblom's paper is an extremely important contribution to the development of both linguistic theory and phonetic theory. What is important is not the primacy

of linguistic forms or phonetic facts but the interaction between the two, which relationship is revealed by this paper. Linguists have, for example, posited that the least marked vowels are *i*, *u*, *a*. Lindblom has provided a possible explanation for this fact. Similarly, one can find many African languages in which reduplication is a very productive grammatical process. Often one finds that the reduplication vowel is /i/ or /u/, agreeing in backness and rounding with the stem vowel but not in height. Linguists have pointed to this fact in support of the notion that /i/ and /u/ are less marked than the mid vowels, but again, Lindblom provides an explanation for the occurrence of these vowels to the exclusion of others.

In the paper two rules are provided, both of which account for the particular glides in one Swedish dialect set of diphthongs. Dr. Lindblom correctly points out that neither the rule using binary features nor the rule using n-ary features is descriptively adequate given the particular set of features now used in generative phonology. He suggests that this model can more adequately handle this phenomenon in an explanatory fashion. Unfortunately the way it would be handled was not presented and it would be helpful if Dr. Lindblom would show the formalisms involved so that the hypothesis can be tested by application to phonological systems.

LINDBLOM
I am well aware of the need to take a closer look at the perceptual space for vowels than we have been able to do so far. Both the remarks of Dr. Ladefoged and Dr. Paddock are very much to the point here. As pointed out by Paddock, it is necessary to consider not only the first two formant frequencies but also the third formant. In fact, an 'effective' $F2$ based on $F2$ and $F3$ was used in the present calculations. I would like to draw attention to some recent work by Fant and his collaborators on this question (see the *STL/QPSR* reports). The difference measure mentioned by Ladefoged represents an interesting suggestion that should definitely be looked into.

I certainly agree with Dr. Tatham that the "rejection of the primacy of linguistic form" must not be interpreted to mean that formal notions and theory should be rejected. On the contrary, my point is simply that models of language should not be developed along lines incompatible with a truly and maximally predictive use of phonetics. This remark may serve as a comment also on Dr. Wajskop's warning that the proposed approach might be carried too far. I find myself in agreement also with Dr. Fromkin who draws attention to the important question of how phonological rules should be formalized. Before we have a satisfactory answer, it will of course be necessary to explore further the possibilities of numerical models of speech production and speech perception. On the other hand, the work I reported on today DOES suggest a formalism or notation which means that hypotheses formulated in terms of this notation can in fact be tested against phonological data also at this early stage. [Reference: R. Carlson, B. Granström, and G. Fant, "Some Studies Concerning Perception of Isolated Vowels", *STL-QPSR* 2-3: 19-35 (1970).