In Butterworth, B., Comrie, B. and Dahl, B. (eds.): Explanations of Linguistic Universals, Mouton: The Hague. 1983.

Self-organizing processes and the explanation of phonological universals

BJÖRN LINDBLOM, PETER MacNEILAGE, and MICHAEL STUDDERT-KENNEDY

Where do phonological universals such as segments and features come from and what general form would explanations of their origin take? In the present study we shall address these questions by trying to simulate their emergence with the aid of a *self-organizing model* of 'phonological structure'. This model was implemented in a series of computational experiments organized to select sequentially — in the presence of certain production-based and perception-based performance constraints — a subset of k phonetic signals from a larger inventory of n universally 'possible gestures'. Atlhough these gestures resembled stop-vowel syllables, their specification did not presuppose an analysis in terms of segments. Rather a possible gesture was defined as a *holistic transition* running between an arbitrary point in the universal phonetic space of 'possible closures' and a similarly arbitrary point in the universal space of 'possible vowels'. These gestalts are phonetically described as articulatory, acoustic, and auditory patterns. To select systems of k signals, an optimization criterion was applied. It was defined so as to produce paradigms achieving 'sufficient perceptual benefits at acceptable articulatory costs'.

Simulations of the present kind can be evaluated with at least the following two questions in mind. First, how well do the derived syllables predict observations on favored systems of stops and vowels in the languages of the world? Some preliminary results will be reviewed. Second, the possible syllable initiations ranged across the following places of articulation: [b d̪ d ɖ ɟ g ɢ]. The simulations favored [b d̪] and vowel-dependent alternations of [ɟ g]. Favored syllable end points were similar to 'peripheral' and 'front, rounded' vowel qualities. These predictions were thus found to show tendencies that are well known from published surveys of typological observations (Maddieson 1980; Crothers 1978; Nartey 1979).

However, our main concern in this paper is not with trying to predict phonetic systems — a project that deserves continual attention as our

understanding of performance constraints develops. Rather we shall concentrate on a second question which is preliminary to, and rather more fundamental than, the first. Is it the case that every derived syllable remains a gestalt pattern that cannot be fractionated into smaller parts occurring also in other syllables ('holistic coding')? Or do we find that every selected signal can indeed be reduced to subparts shared with other syllables of the subset ('phonemic coding')? In the simulated systems it is indeed possible to find 'minimal pairs', i.e. pairs of transitions that share the beginning or end portions. Hence we note that the findings do provide instances of 'phonemic' and 'segmental coding'. As a second result, we observe that it is possible to analyze the derived contrasts in terms of traditional distinctive feature dimensions (grave–acute, compact–diffuse, flat–plain). A third result is the 'rule' governing the distribution of palatal and velar allophone of the '/g/ phoneme'. We should draw special attention to the fact that segment, feature, and rule are not explicit constructs of the present theory. They are IMPLICIT properties of the phonetic signals.

In discussing these results we shall pay special attention to the conditions under which structuration into segments and features arise: the mechanism favoring phonemic coding (=the repeated contrastive use of a syllable outset or offset) requires that k — the 'lexicon' — be much greater than the number of available onsets or offsets. If the performance constraints severely limit the phonetic variation of onsets and offsets, and if k becomes sufficiently large relative to the phonetic repertoire, speakers can find a way of making their inventory of phonetic signals grow ONLY by invoking gesture onsets and offsets repeatedly and in new combinations (Studdert-Kennedy 1980; Studdert-Kennedy and Lane 1980).

How was this mechanism discovered in phylogeny? How is the child's discovery of it facilitated? We can obtain hints as to how those questions might be answered by viewing the behavior of the child and the strategy of our early ancestors in developing primitive sound–meaning correspondences as 'a random sampling of the universal phonetic space in the presence of performance constraints'. Doing so we shall be able to argue that segmental and featural structuration seems to be built into the phylogeny and ontogeny of speech as a statistical bias and arises IMPLICITLY in a 'self-organizing' manner. A few brief remarks on the implications of the present results for phonological theory will conclude the paper.

The segmental and featural structure of speech: a phonological universal that demands an explanation

All languages have phonologies that uniformly structure pronunciations in terms of sequences of segments (phonemes, allophones) and certain phonetic dimensions of contrast (distinctive features).

The segmental and featural structure of speech is a strong phonological universal, so strong that it is perhaps taken for granted and treated as a fact that need not be explained. For instance, in current phonological studies segments and features are posited as universal linguistic categories. They have the same status as noun, verb, noun phrase, subject, etc., have among syntactic categories and exemplify so-called substantive universals (Comrie 1981). They are axiomatically given. As primitives of the theory they do not require explanation.

As we consider the matter more closely, however, it appears rather remarkable that all languages should use the same method for coding meanings phonetically. Although by definition semantically irreducible, their smallest meaningful elements — the lexical and grammatical morphemes — can be further analyzed phonologically into a limited number of phonemes and allophones, i.e. vowels and consonants. These phonetic segments are not the ultimate units but are in turn drawn from a finite and restricted set of simultaneous constituents: distinctive features. Thus rather than use (the logically equally possible alternative of) gestalt coding — one holistic phonetic signal per morpheme — languages uniformly favor phonemic coding mapping meaning onto sound by forming combinatory patterns of the quantal phonetic attributes that we term segments and features.

Phonemic coding in this sense is a key property of spoken language contributing to distinguish it from animal communication systems (Hockett and Ascher 1964; Chafe 1970: 24–29). It is paralleled by the cheremic structure of sign languages (Stokoe 1969; Klima and Bellugi 1979; Bergman 1979). It represents a conquest that, during its speech development, every normal child makes so miraculously without conscious effort. In principle, it promises to keep the experimental phonetician out of unemployment for a long time, since it raises the so-far unresolved issues of phonetic invariance and segmentation (Lindblom 1982a). It falls at the heart of phonological theory, which must be said to be still in pursuit of a definitive, universally valid distinctive-feature framework.

Where do segments and features come from? Our position is that not only is that a reasonable question to ask. According to a widely accepted interpretation of the notion of explanation, it must be regarded as THE most fundamental question of theories aiming at providing truly explana-

tory accounts of sound patterns. In view of the theoretical and practical ramifications of the problem, even partial answers might be of considerable interest.

What would count as an explanation?[1]

In what direction do we begin our search for possible explanations? We shall consider two general approaches: one 'mentalistic', the other 'mechanistic'.

Following the first approach, we shall make use of an argument based on information theory that treats speech as an error-correcting code. We shall take note of the fact that, compared with holistic coding, phonemic coding is an extremely efficient method for mapping semantic information onto the signal medium (Mandelbrot 1954). Holistic coding tends to rapidly crowd the phonetic space and is therefore incompatible with communicative demands for sufficiently distinct signals. Would it be reasonable to attribute the following reasoning to one of our early ancestors (as he engages in thought processes taking place without the support of language, to be sure)? 'Since I have somehow developed all these concepts that I would like to communicate, and since the vocalizations that I produce by combining various distinct "features" are communicatively much superior to holistic signals, I will SIMPLY INVENT "phonemic coding" based on segments and features!' Presenting the argument in this rather extreme and admittedly caricatured form serves the purpose of questioning any explanatory attempt that involves elements of conscious intent on the part of the originators of the phonemic principle. For even if we assume that this early ancestor of ours possessed the intellectual powers to fathom and solve the sound–meaning coding problem in wordless thought — a questionable assumption in the first place — how was he to impose his brilliant idea on others? Our conclusion from pursuing such scenarios is that the segmental and featural structure of speech is much more likely to have arisen by accident than as a result of inspired thinking. We are inclined to speculate that our ancestors must as it were have stumbled collectively and blindly over phonemic coding, driven by forces that they did not have direct or conscious control over.

Bickerton (1981) discusses the phylogeny of human language in similar terms. Although he deals with human language as a whole, his remarks apply with equal force to the present topic. Quoting Lamendella (1976: 261) he states that 'it strains credulity to pretend that language as we know it suddenly sprang up intact as a cultural invention in the absence of extensive cognitive and communicative preadaptations'.

An ontogenetic perspective on the origin of segments and features lends further support to the choice of mechanistic over mentalistic explanations. The literature on the development of phonology (Yeni-Komshian et al. 1980) mentions a stage during which children appear to use words as unanalyzed wholes. Later a phonological fine structure emerges in the form of segmental and featural contrasts. Although it is clear that this transition needs to be investigated in greater detail, enough is known about this process to warrant the claim that children are never aware of having acquired phonemic coding. It appears to emerge in a completely automatic and implicit manner.

There is a certain class of explanations that would qualify as mechanistic and that would seem worth exploring in investigating the origins of linguistic form. What we have in mind are the accounts provided by the *theory of self-organizing systems*. This is a scientific paradigm which has recently arisen at the intersection of physics, chemistry, biology, and sociology and which aims at formulating the general laws that are believed to govern the spontaneous occurrence of order in nature and the evolutionary dynamics of such seemingly diverse phenomena as those encountered in physical, biological, and sociocultural systems (Jantsch 1981). The optimism of the proponents appears to rest on the assumption that, wherever there is interaction between subsystems, this interaction must obey certain principles that have considerable generality. In other words, THE INTERACTION OF SUBSYSTEMS — BE THEY SUBSYSTEMS OF MATTER OR INFORMATION OR BEHAVIOR — GIVES RISE TO STRUCTURATION (Haken 1981). The self-organizing framework encompasses concepts at first developed independently by various research groups, e.g. dissipative structures, synergetics, autopoiesis, hypercycles, catastrophe theory, and boot-strap models.

So far linguists do not seem to have discovered — or have not been tempted to explore — the explanatory potential of the new paradigm. The following presentation contains an attempt to model the emergence of phonological structure, i.e. segments and features, as structuration arising as a consequence of self-organization and interaction among subsystems. Before proceeding to the phonetic applications we shall give a more specific illustration of the concept of self-organization and the general form of explanation that it offers: termite nest building, an example brought to our attention by Michael Turvey (Bellugi and Studdert-Kennedy 1980: 91) and discussed in some detail also by Prigogine (1976).

Termites construct nests that are structured in terms of pillars and arches and that create a sort of 'air-conditioned' environment. The form of these nests appears to arise as a result of a simple local behavioral pattern which is followed by each individual insect: the pillars and arches

are formed by deposits of glutinous sand flavored with pheromone. Pheromone is a chemical substance that is used in communication within certain insect species. Animals respond to such stimuli after (tasting or) smelling them. Each termite appears to follow a path of increasing pheromone density and deposit when the density starts to decrease. Suppose the termites begin to build on a fairly flat surface. In the beginning the deposits are randomly distributed. A fairly uniform distribution of pheromone is produced. Somewhat later local peaks have begun to appear serving as stimuli for further deposits that gradually grow into pillars and walls by iteration of the same basic stimulus–response process. At points where several such peaks come close, stimulus conditions are particularly likely to generate responses. Deposits made near such maxima of stimulation tend to form arches. As the termites continue their local behavior in this manner, the elaborate structure of the nest gradually emerges.

What is the relevance of this account and numerous similar ones proposed to explain order and form in physics, biology, and sociology (Prigogine 1980; Bouligand 1980; Jantsch 1981)? In the context of linguistics for which distinctions such as form–substance, langue–parole, and competence–performance are among the methodological cornerstones, we should draw attention to the following points in particular:

(i) 'Structure' CAN be explained, i.e. deductively derived rather than axiomatically postulated.

(ii) The explanations, although offered for unrelated phenomena, share the property of avoiding the assumption of an explicit dichotomy into FORM–SUBSTANCE. In the formation of, for example, beer bubbles (Smith 1981), snow flakes, physiological systems, schools of fish (Partridge 1982), 'form' is implicit and inextricably interwoven with 'substance'.

(iii) In spite of the seemingly purposeful and intricate design of termite nests, it seems neither realistic nor scientifically parsimonious to attribute to these animals a 'mental blue print' for the finished product or assign 'psychological reality' to the 'nest' as an autonomous unit of termite cognition. The architecture of the nest is most simply interpreted as the indirect consequence of a local behavior that tends to be recursively performed in the presence of certain local stimulus conditions.

(iv) The interaction of microprocesses and subsystems is seen to give rise to patterns at macrolevels that can be highly complex. Accordingly we should not neglect to observe that the 'structure-causing' power of local blind processes can be considerable. Are

they sufficiently powerful to account also for aspects of linguistic structure? We suggest that at least their explanatory value should not be underestimated.

What would count as an explanation of the origin of phonological universals such as segments and features? The preceding digression on self-organizing systems provides us with some general guidelines.

(1) Acceptable explanations are to be found among accounts that derive segments and features deductively from independently motivated principles.

First, postulating segments and features as primitive universal categories of linguistic theory should be rejected, even if it could be done in such a way as to describe successfully all available typological data on possible segments and possible features. This is the traditional axiomatic approach (Jakobson et al. 1952; Chomsky and Halle 1968). It deliberately avoids addressing the issue of explaining where segments and features come from.

Second, a more satisfactory explanation is one for which the degree of 'independent motivation' is stronger. To exemplify, let us consider a hypothetical attempt to motivate the axiomatic postulation of segments and features with reference to their biological evolution. It could be argued that treating segments and features as primitives does not deliberately avoid 'addressing the issue of explanation' but simply assumes that these aspects of language structure are caused by genetic idiosyncracies unique to language users and therefore cannot be explained in terms of independently motivated nonlinguistic information. (For this sort of view of language biology see, for instance, Chomsky [1976].) While autonomous language-unique phenomena seem by no means biologically implausible, demonstrating their existence or proving their nonexistence requires a single research strategy: DERIVE LANGUAGE FROM NONLANGUAGE! Only when such attempts have been exhaustive are we entitled to conclude that we are probably dealing with properties that are unique to language and that should be regarded as major discontinuities or 'mutations'. Our position is that jumping to such conclusions in the case of segments and features is no doubt premature, since deriving phonology from nonphonology can hardly be said to be an exhaustively explored research paradigm.

The implication of our second point thus appears to be the following:

(2) Given two accounts, the more successful explanation is the one that more extensively traces the evolutionary roots of linguistic phenomena to preadaptations and extralinguistic factors.

Simulations of emerging phonetic structure

The following section summarizes the results of two sets of computational experiments described in greater detail elsewhere (Lindblom et al. forthcoming). There are certain assumptions that are common to all the simulations and represent independently given, *a priori* available information. For their mathematical definition see Lindblom et al. (forthcoming).

Size of 'lexicon' or signal inventory. The computations assign phonetic shape to k distinct hypothetical meanings ('lexical' elements). Simulations are initiated by specifying this number.

Universal phonetic signal space. A language-independent universal specification of the class of 'possible articulations' is presupposed. The description developed for the purpose of simulating phonetic structure is so far fragmentary and confined to vowels and voiced stops. 'Possible articulation' is defined with the aid of a numerical articulatory model. This model was constructed to accommodate, in a preliminary but physiologically realistic manner, both speech and nonspeech phenomena. It can be said to represent a set of hypotheses about A UNIVERSAL: the class of articulations (possible configurations as well as possible movements) available for the formation of phonetic systems (Lindblom and Sundberg 1971; Lindblom et al. 1974; Gay et al. 1981; Lindblom 1982a, 1982b).

Acoustic phonetic theory is applied to translate the 'articulatory space' — i.e. the articulations defined by the numerical framework — into an 'acoustic space', the corresponding acoustic events. With the aid of a model of auditory peripheral analysis (Schroeder et al. 1979), the latter can in turn be transformed into auditory spectra, or sequences of spectra, which define a 'perceptual space'. These mappings introduce knowledge about speech which is language-independent: the universal laws governing formant–cavity relations and the universal properties of human hearing.

The universal phonetic space is accordingly specific at three levels: articulation, acoustics, and perception. In the present application we shall examine a limited aspect of the universal phonetic space. We shall restrict our attention to articulations that involve transitions from a closed (stoplike) to an open (vowellike) state. In other words we shall deal with the events resembling CV syllables. Note, however, that the subspace defining 'possible stops' is a continuous space. So is the space characterizing vowellike elements. We define 'possible CV syllable' as any trajectory running between arbitrary points of the two subspaces. This implies that the CV space is in principle also a continuous one. It is made up of an

infinite number of trajectories. The main result of the present study will consist in demonstrating that in the presence of certain constraints a continuous space can become quantally structured. It is therefore necessary to emphasize that the basic inventory from which k phonetic signals will be chosen is *a priori* neither segmentally nor featurally structured. It is made up of HOLISTIC SIGNALS.

For computational reasons we chose to work with a specification of the CV space in terms of discrete points. The space of vowellike end points was quantized into 19 formant patterns roughly equidistant along roughly logarithmic F_1, F_2, and F_3 coordinates. The phonetic values of these patterns cover a large range of 'cardinal' qualities. They are approximately the following:

(3) [i y ü ʉ ɯ u e *o* ə ɣ o ɛ œ ʌ ɔ æ Œ ɑ a]

A three-dimensional display of the space is shown in Figure 1.

The continuum of place of articulation for the stop closures was divided up into seven points: [b ɖ d ɖ ɟ g ɢ]. A formant pattern — or

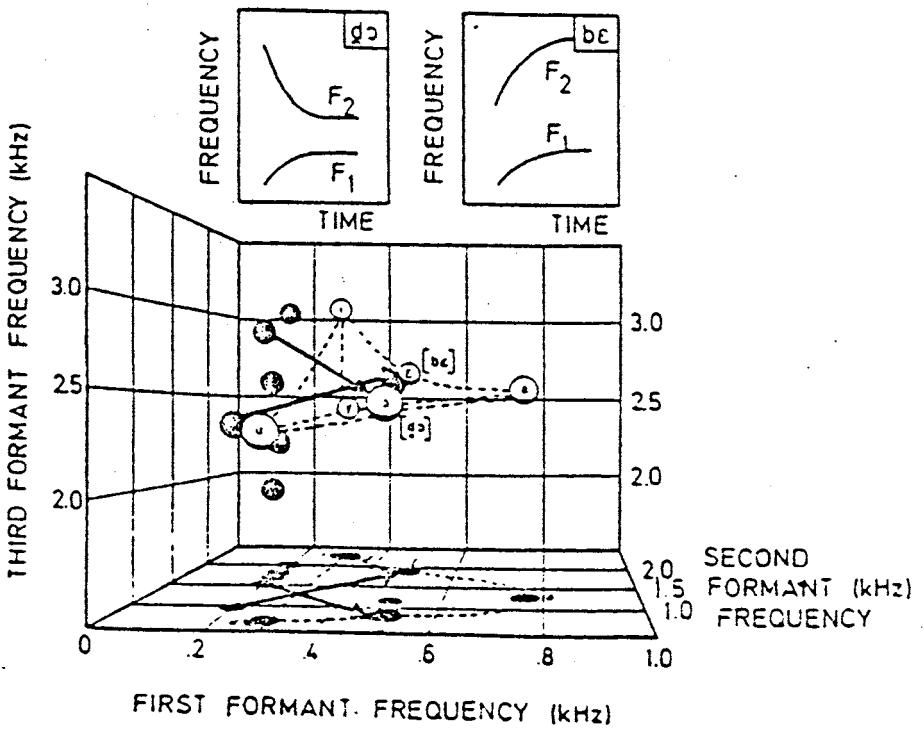

Figure 1. *Three dimensional display of the CV space*

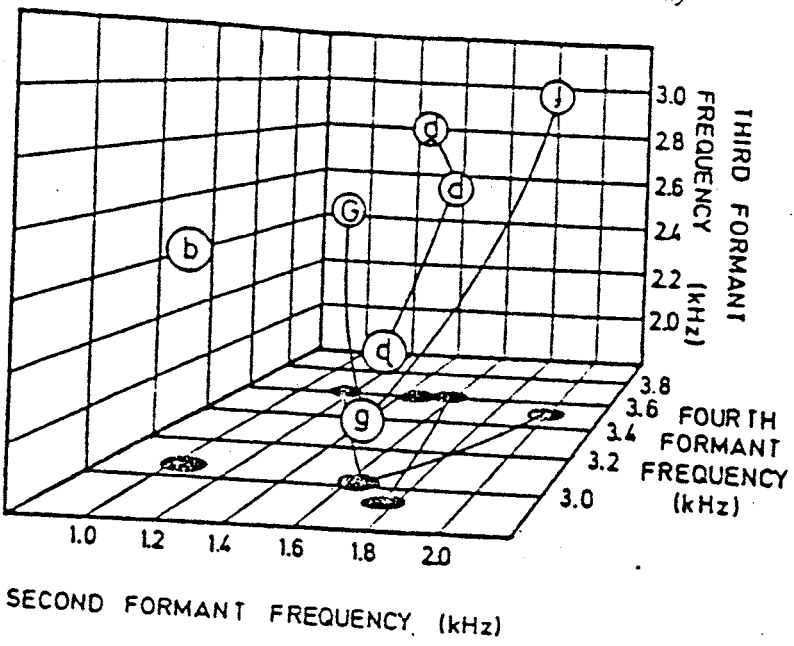

Figure 2. *Points of articulation in the CV space*

'locus pattern' — was assigned to each closure location. The locus pattern is known to be a major perceptual cue for place of articulation. Other correlates such as the burst generated at the moment of plosion were not used. An attempt to illustrate the manner in which these points of articulation sample the space of 'possible locus patterns' is made in Figure 2, which shows contours for possible apical and possible dorsal loci in three dimensions. A single point describes the labial closure. All values refer to articulations in which the occlusion is coarticulated with the near-neutral tongue shape of an [a]. They were determined on the basis of information available in the literature (Klatt and Stevens 1969; Fant 1973; Stevens and Blumstein 1975; Lindblom et al. 1974).

Returning to Figure 1, we can now present an acoustic definition of 'possible CV event' which can be pictured as a straight line coursing between a 'possible locus' and a 'possible vowel'. Two such trajectories are shown in Figure 1, together with their frequency vs. time interpretations (inserts). For our present purposes we define 'possible CV syllable' as any trajectory running between an arbitrary but possible locus and an arbitrary but possible vowel in the space defined by the first four formants. The total inventory of syllables contains $7 \times 19 = 133$ patterns. In principle it is infinite.

It is very important to note that the quantization just described is in no way incompatible with our claiming that the CV space is a continuous one made up of infinitely many holistic signals. The sampling is sufficiently fine-meshed to allow us to pursue the argument to be developed below. The adoption of a space specification in terms of discrete points was made solely for computational convenience and does not introduce — through the back door — the phenomena that we are about to study the emergence of.

Phonetic constraints. The present work departs from the assumption that phonetic signals tend to evolve so as to facilitate both their production and perception. We have explored the following hypothetical performance conditions:

Talker-based conditions:
1. sensory discriminability;
2. preference for 'less extreme' articulation.

Listener-based conditions:
3. perceptual distance;
4. perceptual salience.

The sensory discriminability constraint says that, everything else being equal, targets that are more distant in the sensory space of 'possible articulation' tend to be favored over targets more closely located. The purpose of this condition is to contribute toward making configurations sufficiently discriminable at the level of memory retrieval and at the level where sensory feedback information is monitored. It makes it more difficult to generate the subclass of [ɖ d d̪] than a set having a labial, an apical, and a dorsal consonant, e.g. [b d g]. The preference for 'less extreme' articulation introduces a ranking of both static configurations and movements. To illustrate the effect on static articulations, let us consider [ɖ d d̪], which represents a progression toward greater elevation and retraction of the tongue tip. The stimulations treat [d] as deviating more from a neutral articulation. With respect to movement, the present experiments arrange all gestures ('stop-vowel' transitions) along a dimension of extent of movement. Thus [ji] and [gu] exhibit shorter movements than say [ju] and [gi] and receive their numerical 'penalty points' accordingly.

The perceptual distance of two arbitrary CV transitions is a dimension used to rank-order all possible pairs of CV events. The measure adopted is developed from experimental data on the perception of static formant patterns (Carlson and Granström 1979; Bladon and Lindblom 1981; Lindblom forthcoming) as well as from information on the auditory

representation of dynamic events (Delgutte 1980; Klatt 1979; Lacerda n.d.). To take a few examples, the present definition makes [ɖV] and [dV] more confusable than, say, [bV] and [gV] for identical vowels, and treats [Ci]- and [Ca]- syllables as more different than, for instance, [Ce]/[Cɛ]- or [Cɔ]/[Cœ]- pairs.

Perceptual salience is a characteristic of individual transitions. It is defined in terms of the extent of the CV trajectory, i.e. the distance between the initial and final auditory spectra. This dimension tends to sometimes counterbalance the effect of the 'extent of movement' score. For instance, [gi] is regarded as more salient than [gu] (cf. above). The perceptual value of a given pair of CV events is a joint function of the perceptual distance measure and the salience score.

Two of the performance conditions make statements about individual CV events (preference for 'less extreme' articulation, perceptual salience). The other two describe properties of pairs of CVs (sensory discriminability, perceptual distance).

Having mentioned all these constraints, we can proceed to the rules governing their interaction. For any arbitrary pair of CVs we stipulate that the value of incorporating that particular pair into the 'lexicon' or signal inventory be computed as a ratio and assigned the dimension of PERCEPTUAL EFFECT PER ARTICULATORY COST.

Let us take an example. Suppose we inspect the simulations at a point where a choice between [Ci] and [Cy] is being evaluated. Articulatory costs are found to be comparable but the [Ci] syllable tends to augment the perceptual score somewhat more and is therefore the preferred candidate in the subsequent calculations. Or the program faces a choice between, say, [ju] and [gu], or between [gi] and [ji]. In the former case [ju] scores higher in terms of salience but gets penalized for being the more extensive gesture, as does [gi], which is a velar-palatal movement of considerable perceptual salience. Attention should be drawn to the fact that after a series of trial-and-error runs we adjusted the weights determining the balance between perceptual effect and articulatory cost in such a way that transitions are generally more articulatorily 'expensive' than they are perceptually 'valuable'. Thus there is a built-in bias toward selecting [ji] and [gu] rather than [ju] and [gi].

The articulatory-cost parameter is furthermore useful for filtering out 'nonspeech' phenomena such as the compensatory articulations observed in so-called bite-block speech (Gay et al. 1981). Syllables such as [Ci] can be generated in several ways by the model (and by human speakers, it seems). The normal natural way is to keep the mandible raised but variants for which the jaw is abnormally low are in fact also compatible with acceptable acoustic results. Why build a model that obviously

overgenerates and describes situations that are of only marginal interest? Our answer is that they are NOT of marginal interest since, according to our standpoint, phonetic theory should explain the universally valid fact that speech-sound inventories form highly restricted and similar subsets of all possible speech and nonspeech gestures and vocalizations (Pike 1943). Although the phonetic diversity of the world's languages is impressive (Ladefoged 1982), it is nevertheless true that they fastidiously underexploit the full range of possibilities (Lindblom 1982b). In the case of normal and bite-block vowels, a by-product of this work is that it suggests an explanation of where the open–close feature comes from in vowels (Lindblom and Sundberg 1971).

We summarize the effect of the phonetic constraints on the space of 'possible CV events' by constructing a triangular matrix whose cells represent all possible pairs within the set of 133 syllables ($133 \times 132/2 = 8778$). Each cell now contains a number that can be compared with any other cell entry and that is calibrated in terms of 'perceptual value per articulatory cost'. For the i^{th} row and the j^{th} column, this number is designated as C_{ij}. Our aim is now to assign phonetic shape to a minilexicon consisting of k 'lexical elements' with distinct meanings. We stipulate that the parameter to be optimized for this system be the following:

$$(4) \quad \sum_{i=2}^{k} \sum_{j=1}^{i-1} 1/(C_{ij})^2 \rightarrow \text{threshold value}$$

Stated verbally, this condition says that systems of CV syllables tend to be selected in such a way that they achieve

(5) sufficient perceptual differences at acceptable articulatory costs.

For a minimal value of (4), a single optimal system will be obtained. As the criterion is relaxed, several 'sufficiently good' solutions will be generated. The results to be described below were obtained by using a sequential procedure for selecting k out of n (with $k = 24$ and $n = 133$; why k was set at this value will become clear later on): (1) to initiate, choose any CV syllable; (2) pick the next syllable so that eq. (4) yields a minimal value; (3) keep adding syllables until a system of k syllables has been obtained. For each new item apply eq. (4). Retain the candidate that minimizes the criterion.

A computer program implementing the present framework was run iteratively 133 times with each possible CV syllable serving as the initial item once. Disregarding the redundant first CVs and pooling all 133 sets, we obtain the distribution of CVs indicated by dark cells in Figure 3.

Although every syllable was used once as the initial item, the 133 runs

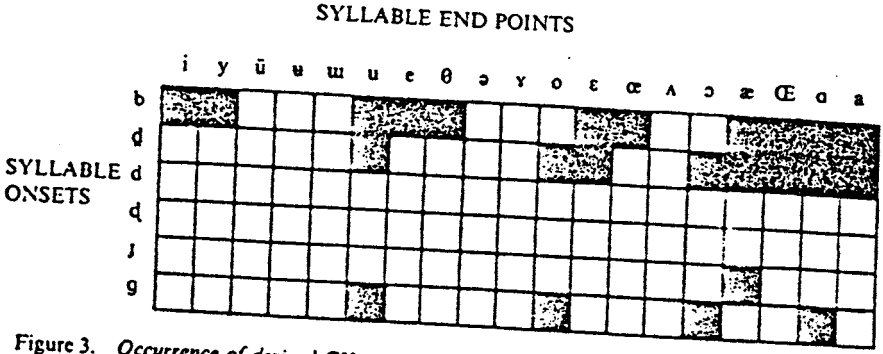

Figure 3. *Occurrence of derived CV combinations*

rapidly converged upon a common set. This observation is reflected in the frequencies of occurrence. In fact, all CVs except [bu, ɖe, bø, bɑ] occurred 100% of the time. From this we conclude that the computations were rather insensitivie to how the recursive search was initiated.

Figure 4 is a plot of these results. The ordinate represents percent occurrence. We see that most vertical lines have equal length (= 100%), whereas the syllables just mentioned show lower values.

A total of 3192 (24 × 133) syllables was generated. The most favored vowel turned out to be [æ]. It occurred 346 times. Figure 5 shows the distribution of vowels in these 3192 syllables plotted on a quasi-acoustic vowel chart. For a quantitative estimate of frequency of occurrence, the size of the vertical lines should be related to that of [œ].

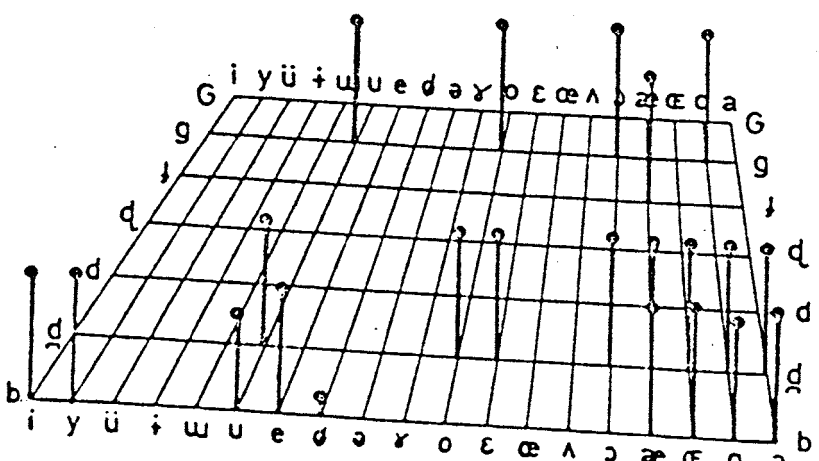

Figure 4. *Frequencies of occurrence of CV combinations*

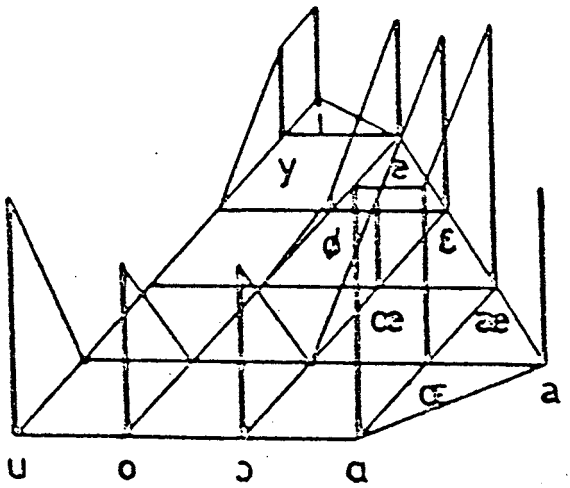

Figure 5. *Distribution of vowels*

A sequential search for 48 syllables was undertaken to examine in particular the distribution of vowels for [j] and [g]. Figure 6 shows the result.

The vowels are paired with the consonants in a complementary manner, [j] combining with 'front' vowels, [g] with 'back' vowels. Only one exception occurs: [u], for which there is a 'phonemic' contrast. However, those syllables appeared fairly late — only as the 46th and the 47th items, and for k roughly equal to $n/3$, that is after about one-third of the total inventory had been used up.

Before we proceed to evaluate these results, let us try to shed some further light on how the phonemic structuring of lexical items differs from a holistic coding. Consider a minilexicon containing 12 words, all in the form of CV syllables. The total number of possible CV trajectories is very large. Suppose we systematically search for sets of 12 CV sequences that are optimal with respect to a given criterion. We can *a priori* envision the two extreme outcomes of such a hypothetical search.

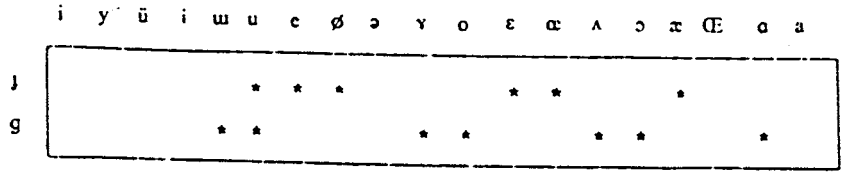

Figure 6. *Tendency toward complementary distribution of stops*

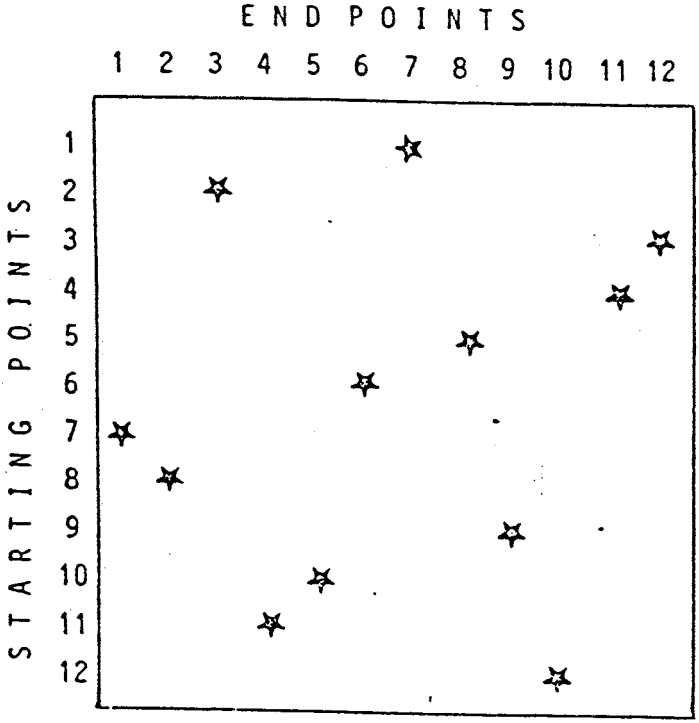

Figure 7. *All words phonetically distinct*

In the first case we find that all the 12 CV words are phonetically distinct with respect to both their beginnings and their ends, as shown in Figure 7. The other extreme is the case where the words are generated by forming all possible combinations of three starting-points and four end points (or conversely), as shown in Figure 8.

Whereas Figure 7 contains no minimal pairs, Figure 8 exhibits the

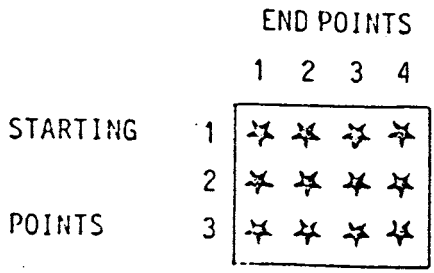

Figure 8. *All possible combinations of points used*

maximum number. The solution in Figure 7 exemplifies the holistic coding: every CV transition is a gestalt that cannot be fractionated into smaller parts also occurring in other CV sequences. The solution in Figure 8, on the other hand, is combinatorially maximally compact. It illustrates the phonemic principle that every CV event can indeed be reduced to subparts shared also with other CV syllables. Patterns resembling Figure 8 will be used as evidence for phonemelike units and those more similar to Figure 7 as support for a gestalt code.

Were we to perform a 'phonemic analysis' of the syllables of Figure 3, we would find three consonant 'phonemes', /b d g/, and the following vowel 'phonemes': /i y u e *o* o ɛ œ ɔ æ Œ ɑ a/. These 'segments' occur in at least one minimal-pair contrast, and since the CVs by definition are 'lexical items', that is, they mean different things, classifying them as 'phonemes' would seem justified according to most analysis procedures.

In terms of our digression on holistic and phonemic coding (cf. Figures 7 and 8), we can say that Figure 3 differs from holistic coding in that individual CV-transitions can indeed 'be fractionated into smaller parts also occurring in other CV sequences'. But does this mean that there is evidence of phonemelike contrasts? This question arises since Figure 3 differs also from Figure 8, which exemplifies the maximally tight case of phonemic patterning. Natural languages, however, do show 'accidental gaps', i.e. cases that are compatible with phonotactic rules but are left lexically unexploited. Therefore the gaps of Figure 3 do not in themselves raise serious objections. In fact, to form 24 words, 3 Cs and 13 Vs are used. The efficiency of the coding could be expressed as $24/(3 \times 13)$ or 62%.

However, the following criticism may be more to the point: the computational experiment is organized to produce sets of 24 CVs (semiarbitrary number). Since this value makes us run out of both Cs — of which there are only seven — and Vs — of which there are 19 — expansion of the vocabulary beyond seven, or 19, words cannot take place unless some starting points and some end points are used more than once. Does not this show that the quasi-phonemic combinations are inevitable and are consequently forced upon us as an artefact of the design of the experiment?

The answer is no. Let us first deal with 'running out of onsets/offsets' as a numerical artefact. To test this possibility we ask the computer to generate no more than seven syllables. In that case we could clearly run out of neither starting points nor end points, and a result like Figure 7 seems theoretically possible. Let us follow a particular run syllable by syllable. We begin by specifying the first item, say [ba]. Given conditions identical to those of the experiment of Figure 3, that choice produces the following sequence:

(6) [ɖu, bæ, bɛ, ɖa, gu, dæ, bi...].

Apparently that set does indeed also exhibit 'minimal pairs' in terms of both Cs and Vs. Therefore we can conclude that the simulated 'phonemic constraints' are not artefacts. Nevertheless, could they be a result of 'running out of onsets/offsets'? They could. According to our interpretation, that is in fact precisely the nature of the mechanism underlying phonemic coding in the simulations. A mechanism that favors and necessitates the repeated use of the onsets and offsets of phonetic trajectories requires that k be much greater than the number of available onsets or offsets. Query: what determines the size of that inventory? Answer: the severity of the performance constraints. The reason why example (7) does not contain [ɖi, ju, gø...] is thus that conditions such as 'preference for less extreme articulation' and 'perceptual discriminability' introduce such high penalty scores for combinations of these syllables so as to put them out of play for a long time. Figure 9 shows the gradual emergence of a near-optimal sequence of syllables in a schematic way.

We can think of the spiral as symbolizing the step-by-step expansion of the phonetic space, the radius representing a production-based performance dimension with neutral and low-cost configurations/movements at the origin and increasingly nonneutral and high-cost articulations centrifugally.

Following the contour, we encounter a sequence of syllables arranged according to the rank order which is uniquely determined in the simula-

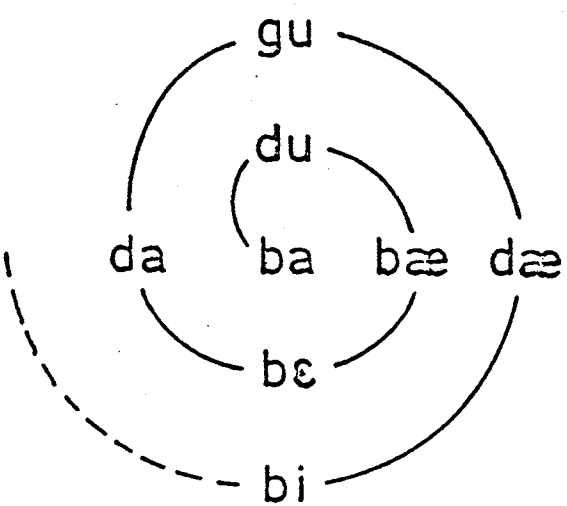

Figure 9. *Optimal sequence of syllables*

tions by applying eq. (4) and by specifying the first syllable. Locally within the spiral the order is influenced by perceptual criteria. At a certain distance away from the origin — i.e. at a certain level of articulatory cost — several syllables may become eligible since they are equally 'expensive'. There the rank order is determined by the perceptual distance and salience criterion. The spiral simplifies somewhat, but it may help us to visualize more clearly the gradual unfolding of the phonetic signals according to an articulatory–perceptual cost–benefit criterion.

The origin of segments and features: an explanation based on the concept of self-organization

Returning to our previous discussion of the phylogeny and ontogeny of the segmental structure of speech, we are now in a position to approach more closely the goals that were sketched initially: model the emergence of phonological structure, in particular the pattern of segments and features, as a self-organizing system! In what sense can the present simulations be said to be compatible with such objectives?

Our results indicate the following:

(i) Random sampling of the possibilities offered by the universal phonetic space should make all such possibilities equally probable. However, in the presence of certain constraints, nonuniform preferences for certain syllables over others arise ('quantal structuration'). The constraints in question are the performance constraints, but it is also important to realize that the very construction of a 'system', or signal inventory, plays a role. The notion of 'system' implies that certain paradigmatic relations among the elements of that system must hold. We have expressed that condition in terms of eq. (4). When those relations are present, structuration occurs. In other words, implicit form emerges and the causes of such pattern formation are indirect. They are both the performance constraints and the 'system'. We conclude that our simulations resemble the characterization of self-organizing systems presented above (Haken 1981): where there is interaction among 'subsystems' — in this case CV gestalts — there is structuration.

(ii) If once more we picture our early ancestors on the point of discovering phonemic coding, a somewhat more specific scenario now appears possible. Let us assume for instance that their joint effort to define the phonetic shapes of a growing set of concepts can be described as 'A RANDOM SAMPLING OF THE UNIVERSAL PHONETIC

SPACE IN THE PRESENCE OF PERFORMANCE CONSTRAINTS'. Our preceding reasoning should apply, which means that quantal structuration is built into the phylogeny of speech as a statistical bias. When several individuals find that their random samplings sometimes converge and similar signals are favored, a situation that might be conducive to socially conventionalized 'naming' appears to be at hand.

A few speculations on the ontogeny of phonemic coding also seem justified. A possible description of the child might be, 'a partly random, partly stimulus-controlled sampler of the universal phonetic space in the presence of performance constraints'. To what extent are the phoneme and the feature EXPLICITLY PRESENT in the speech signals that the child experiences? There seems to be no quantitative empirical measurements that could help us answer that question right away and fully satisfactorily. Venturing an informed guess, and guided by three or four decades of acoustic phonetic research, we nevertheless suggest that phonemic segments and features are NOT explicitly present in the input to the child. Although often slightly overarticulated and characterized by a maximization of cues, baby talk presents the acoustic phonetician with the same central issues as adult speech: those of segmentation and invariance. If this turns out to be correct, the present theory still appears to offer some hope for accounting for how the child, at least implicitly, 'discovers the phoneme'.

(iii) When we referred to 'structuration' above, we had the following results in mind: Figure 3 exhibits 'minimal pairs'. Hence it provides evidence of phonemelike or segmental coding. Second, when we examine the pattern that the dark cells of Figure 3 form, we find that the 'consonants' have LABIAL, DENTAL, and PALATAL/VELAR places of articulation and that the 'vowels' contrast along dimensions such as OPEN–CLOSE, FRONT–BACK and ROUNDED–UNROUNDED. On the whole, the predicted syllables bear some, if not strong, resemblance to natural sets of syllables. A third result is the rather realistic allophonic variation of the '/g/ phoneme' shown in Figure 4. Note that FEATURE, PHONEME, and ALLOPHONIC RULE are present only as IMPLICIT properties of the behavior. They do not have the status of explicit constructs in the present theory. They are derived rather than axiomatically postulated as 'substantive universals'. They are in fact imputed to our simulated speaker the moment we subject the derived syllables to a 'conventional linguistic analysis'. In a sense, our description is a phonology without 'features', 'segments', and 'rules'. The analogy with the

termite story should be obvious. It appears that, like the explanation of the nest building, our account also manages to dispense with 'mental blueprints' — in the present case for feature, segment, and rule.

Although the possibility of 'mental blueprints' for explicit, autonomous phonological constructs should remain viable, these results raise a number of serious implications for phonological theory. They could be taken to suggest that current overly formal approaches to phonology — e.g. Chomsky and Halle (1968) and its offshoots — underestimate the role of performance constraints in the formation of sound patterns as well as neglecting the exploitation of the explanatory power of the concept of self-organization. A consequence of such underestimation and neglect might be that current approaches severely limit their possibilities to formulate theoretically a deeper understanding of the dynamics of sound patterns in historical and ontogenetic development as well as in adult on-line performance. In that context it does not seem unmotivated to regard the present demonstrations of implicit phonological structure as particularly provocative. To what extent are current theoretical descriptions of phonology analogous to descriptions of termite nest building that invoke mental blueprints and are therefore wrong? In other words, are 'feature, segment, and rule' — as defined axiomatically rather than deductively in current phonological theories — artefacts due to choosing a formalism that sacrifices theoretical depth for descriptive efficiency? Or are they truly psychologically explicit and autonomous realities? Stating that issue has been a major goal of this paper.

Stockholm University
University of Texas, Austin
CUNY and Haskins Laboratories

Notes

1. An obvious source of inspiration for the present work is Stevens (1972). While relevant to the present theme, it will not be reviewed in this context. For an extensive discussion of the quantal theory proposed by Stevens and its relationship to the research reported here, see Lindblom et al. (forthcoming).

References

Bellugi, U. and Studdert-Kennedy, M. (eds.) (1980). *Sign Language and Spoken Language: Biological Constraints on Linguistic Form*. Dahlem Workshop. Weinheim and New York: Chemie.

Bergman, B. (1979). *Signed Swedish.* National Swedish Board of Education. Stockholm: Liber.

Bickerton, D. (1981). *Roots of Language.* Ann Arbor: Karoma.

Bladon, R. A. W. and Lindblom, B. (1981). Modeling and judgements of vowel quality differences. *Journal of the Acoustic Society of America* 69(5): 1414–1422.

Bouligand, Y. (1980). *La morphogenèse de la biologie aux mathématiques.* Paris: Maloine.

Carlson, R. and Granström, B. (1979). Model predictions of vowel dissimilarity. *STL-QPSR* 3–4, 84–104.

Chafe, W. L. (1970). *Meaning and Structure of Language.* Chicago and London: University of Chicago Press.

Chomsky, N. (1976). On the nature of language. In S. R. Harnad, H. D. Steklis, and J. Lancaster (eds), *Origins and Evolution of Language and Speech*, 46–57. Annals of the New York Academy of Science, vol. 280. New York.

—and Halle, M. (1968). *The Sound Pattern of English.* New York: Harper and Row.

Comrie, B. (1981). *Language Universals and Linguistic Typology.* Oxford: Blackwell.

Crothers, J. (1978). Typology and universals of vowel systems. In J. H. Greenberg (ed.), *Universals of Human Language* 2, 93–152.

Delgutte, B. (1980). Representation of speech-like sounds in the discharge patterns of auditory-nerve fibers. *Journal of the Acoustic Society of America* 68(3), 843–857.

Fant, G. (1973). *Speech Sounds and Features.* Cambridge, Mass.: MIT Press.

Gay, T., Lindblom, B., and Lubker, J. (1981). Production of bite-block vowels: acoustic equivalence by selective compensation. *Journal of the Acoustic Society of America* 69(3), 802–810.

Grillner, S., Lindblom, B., Lubker, J., and Persson, A. (1982). *Speech Motor Control.* London: Pergamon.

Haken, (1981). Synergetics: is self-organization governed by universal principles? In E. Jantsch (ed.), *The Evolutionary Vision: Toward a Unifying Paradigm of Physical, Biological and Sociocultural Evolution.* Boulder: Westview.

Harnad, S. R., Steklis, H. D., and Lancaster, J. (1976). *Origins and Evolution of Language and Speech.* Annals of the New York Academy of Science, vol. 280, New York.

Hockett, C. F. and Ascher, R. (1964). The human revolution. *Current Anthropology* 5, 135–147.

Jakobson, R., Fant, G., and Halle, M. (1952). *Preliminaries to Speech Analysis.* Cambridge, Mass.: MIT Press.

Jantsch, E. (1981). *The Evolutionary Vision: Toward a Unifying Paradigm of Physical, Biological and Sociocultural Evolution.* Boulder: Westview.

Klatt, D. H. (1979). Speech perception: a model of acoustic-phonetic analysis and lexical access. *Journal of Phonetics* 7, 279–312.

—and Stevens, K. N. (1969). Paryngeal consonants. *Quarterly Progress Report, MIT Research Laboratory of Electronics* 93, 207–219.

Klima, E. and Bellugi, U. (1979). *The Signs of Language.* Cambridge, Mass.: Harvard University Press.

Lacerda, F. (n.d.). Unpublished thesis, Stockholm University.

Ladefoged, P. (1982). Cross-linguistic studies of speech production. In P. F. MacNeilage (ed.), *The Production of Speech.* Heidelberg: Springer.

Lamendella, J. (1976). Relations between the ontogeny and phylogeny of language: a neo-recapitulationist view. In S. R. Harnad, H. D. Steklis, and J. Lancaster (eds), *Origins and Evolution of Languages and Speech.* Annals of the New York Academy of Science, vol. 280. New York.

Lindblom, B. (1982a). The interdisciplinary challenge of speech motor control. In S.

Grillner, B. Lindblom, J. Lubker, and A. Persson, (eds), *Speech Motor Control*. London: Pergamon.

—(1982b). Economy of speech gestures. In P. F. MacNeilage, (ed.), *The Production of Speech*. Heidelberg: Springer.

—(forthcoming). Phonetic universals in vowel systems. In J. J. Ohala, (ed.), *Experimental Phonology*. New York: Academic Press.

—, MacNeilage, P., and Studdert-Kennedy, M. (forthcoming). *The Biological Bases of Spoken Language*. San Francisco: Academic Press.

—, Pauli, S., and Sundberg, J. (1974). Modeling coarticulation in apical stops. In G. Fant, (ed.), *Proceedings of the Speech Communication Seminar, Speech Communication* 2. Stockholm: Almqvist and Wiksell.

—and Sundberg, J. (1971). Acoustic consequences of lip, tongue, jaw and larynx movement. *Journal of the Acoustic Society of America* 50, 1166–1179.

MacNeilage, P. F. (ed.) (1982). *The Production of Speech*. Heidelberg: Springer.

Maddieson, I. (1980). Phonological generalization derived from the UCLA Phonological Segment Inventory Database. WPP 50: 57–68, UCLA.

Mandelbrot, B. (1954). Structure formelle des langues et communication. *Word* 10, 1–27.

Nartey, J. N. A. (1979). A study in phonemic universals — especially concerning fricatives and stops. WPP 46, UCLA.

Partridge, B. L. (1982). The structure and function of fish of schools. *Scientific American* 246(6), 90–99.

Pike, K. L. (1943). *Phonetics*. Ann Arbor: University of Michigan Press.

Prigogine, I. (1976). Order through fluctuation: self-organization and social system. In E. Jantsch and C. Waddington (eds.), *Evolution and Consciousness*, 93–126. Reading, Mass.: Addison-Wesley.

—(1980). *From Being to Becoming*. San Fransisco: Freeman.

Schroeder, M. R., Atal, B. S., and Hall, J. L. (1979). Objective measure of certain speech signal degradations based on masking properties of human auditory perception. In B. Lindblom and S. Öhman (eds), *Frontiers of Speech Communication Research*, 217–229. London: Academic Press.

Smith, C. S. (1981). *A Search for Structure: Selected Essays on Science, Art and History*. Cambridge, Mass.: MIT Press.

Stevens, K. N. (1972). The quantal nature of speech: evidence from articulatory-acoustic data. In P. B. Denes and E. E. David, Jr. (eds.), *Human Communication: A Unified View*, New York: McGraw-Hill. 51–66.

—and Blumstein, S. E. (1975). Quantal aspects of consonant production and perception: a study of retroflex stop consonants. *Journal of Phonetics* 3, 215–233.

Stokoe, W. C. (1969). *Sign Language Structure*. Studies in Linguistics: Occasional Papers 8. Buffalo: University of Buffalo Press. (Reprinted 1978, Silver Spring, Md.: Linstock.)

Studdert-Kennedy, M. (1980). The beginnings of speech. In G. B. Barlow, K. Immelmann, M. Main, and L. Petrinovich, (eds), *Behavioral Development: The Bielefield Interdisciplinary Project*. New York: Cambridge University Press.

— and Lane, H. (1980). Clues from the differences between signed and spoken language. In U. Bellugi and M. Studdert-Kennedy (eds), *Signed and Spoken Language: Biological Constraints on Linguistic Form*, 29–40. Dahlem Konferenzem 1980. Weinheim: Chemie.

Yeni-Komshian, G. H., Kavanagh, J. F., and Ferguson, C. A. (1980). *Child Phonology*, vol. 1. New York: Academic Press.