Emergent Phonology
Author(s): Björn Lindblom
Proceedings of the 25th Annual Meeting of the Berkeley Linguistics Society (2000), pp. 195-209

Emergent phonology

BJÖRN LINDBLOM
University of Stockholm
University of Texas at Austin

0. Speech and phonology

As is well known the physical properties of speech are determined by a large number of factors. They vary depending on the language spoken, on the age, gender and identity of the speaker. They undergo stylistic modifications in innumerable ways owing to the interaction of physiological, cognitive, social and communicative factors.

Linguistics deals with this problem by assigning the study of the speech signal to phonetics. Phonology investigates postulated sound structure, that is, abstract entities and processes which are assumed to underlie speech behavior and which are by definition independent of performance and language use. In this way, the study of speech sounds is split into two: Phonology becomes digital, and phonetics analog.

> "... the fundamental contribution which Saussure made to the development of linguistics ..." was "... to focus the attention of the linguist on the system of regularities and relations which support the differences among signs, rather than on the details of individual sound and meaning in and of themselves. ... For Saussure, the detailed information accumulated by phoneticians is only of limited utility for the linguist, since he is primarily interested in the ways in which sound images differ, and thus does not need to know everything the phonetician can tell him. ... By this move, then, linguists could be emancipated from their growing obsession with phonetic detail." (Anderson 1985:41-42).

1. The child's problem

It is interesting to examine this traditional division of labor from the viewpoint of speech development. If children's phonetic input is indeed massively variable, how do they find phonology behind that variability? How do they discover the hidden structure of speech? We shall develop a tentative answer in the following steps.

Perceptual categories form as emergent consequences of accumulated phonetic experience (speech transforms complex but lawful). Motor development is guided by a physiological economy principle which helps the child spontaneously discover many patterns in the ambient language (motor bootstrapping). That heuristic presupposes that sound systems are (in part) adapted to be spoken. There is a great deal of evidence supporting this claim.

2. Avoiding circularity

To avoid circularity, it appears desirable to impose the following methodological conditions. We should not assume phonological structure to be pre-specified: Say no to nativism! Neither should it be postulated from the data to be explained: Say no to curve-fitting! The long-term goal is to deduce sound structure from the child's experience and minimal assumptions about 'initial knowledge'—technically speaking, as a *behavioral emergent.*

3. Role of perceptual experience

3.1 Infant-directed speech

Is infant-directed speech less variable? It is true that Baby Talk shows less complex sentences and simpler words? It contains repetitions and is often produced more slowly with greater emphasis on the most informative elements (Ferguson 1977). Accordingly, it exhibits adaptations, but does that mean that BT moves the 'hidden structure' of speech to the surface so that the child can pick up the phonology in the signal in a more direct way?

There is converging evidence that BT does not solve the 'invariance problem' for the child in that way. A recent cross-linguistic study illustrates the point (Kuhl et al 1997). It compares the formant frequencies of vowels in adult-to-adult conversations and infant-directed speech. The data indicate a tendency for BT vowels to be somewhat more peripheral in the vowel space, but by and large the variability is extensive and comparable to that of adult-to-adult speech. Similar observations (Fónagy 1983, Davis & Lindblom 1994, Sundberg 1998) indicate that BT in no way undoes the context-dependence of acoustic phonetic patterns. Typically it is prosodically lively and emotionally positive (Fernald 1984). Rather than make phonological units more transparent, emotive coloring and prosodic liveliness have the effect of increasing the complexity of the acoustic encoding (Lindblom et al 1992).

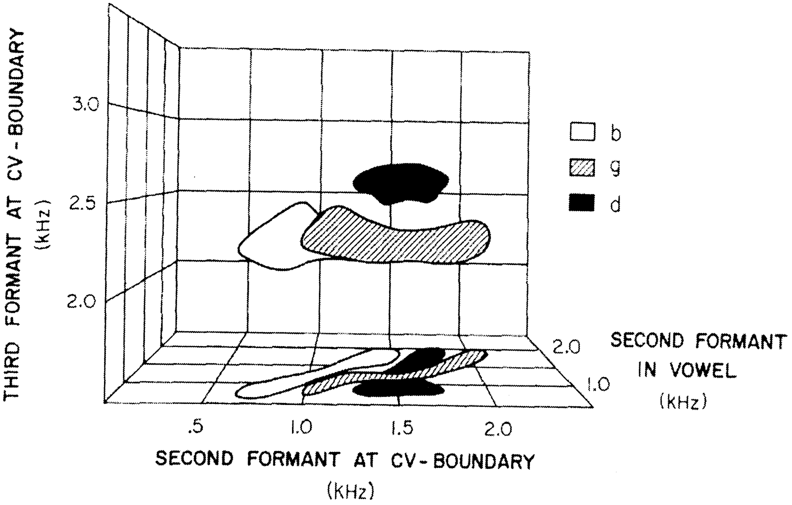
3.2 Perceptual categories as emergents of phonetic experience

While it is certainly true that speech signals vary greatly, it is equally clear that this variation is non-random. Articulatory dynamics and vocal tract acoustics shape phonetic variations in complex, but basically highly systematic ways. What role do such regularities play in speech development? Some results on speech perception by animals provide some clues to answering that question.

Japanese quails were trained to peck in response to syllables beginning with /d/, but to avoid pecking when hearing syllables with initial /b/ or /g/ (Kluender, Diehl & Killeen 1987). The items were produced by an American English speaker. They contained /i/, /u/, /æ/ or /a/ and exhibited the normal patterns of consonant-vowel coarticulation. At the end of training it was found that the birds had successfully learned to discriminate /d/- from /b/- & /g/-syllables. The same quails were then presented with a set of new stimuli containing the same consonants but different vowels: /ɪ ɛ ʊ ʌ eɪ oʊ oɪ ɚ/. The task was to peck in response to stimuli with /d/, and to avoid pecking when hearing other syllables. Again the birds performed successfully. They were able to generalize their experience of the first experiment to the new stimuli of the second.

It appears possible to explain the quail's behavior by hypothesizing an exemplar-based learning process. This account assumes that each individual stimulus (more precisely, each auditory pattern) gets stored in a holistic manner, say as a 'neuro-spectrogram'. As more patterns accumulate, the data sort themselves into clusters. The physical similarities and differences between the stimuli cause three distinct representations to form in auditory space.

(1) Acoustic separation of place in coarticulated stops.



A simplified version of the process is illustrated in (1) where $F2_{onset}$, $F3_{onset}$ and $F2_{vowel}$ are plotted for a set of CV sequences similar to those used in the quail study. As can be seen, the three categories emerge as three non-overlapping, three-dimensional 'clouds'. It does not seem unjustified to assume that, *at the very least,* the quail had access to the information portrayed in this diagram. If so, they would have had the possibility of associating the /d/ cloud with pecking and the other clouds with no pecking.

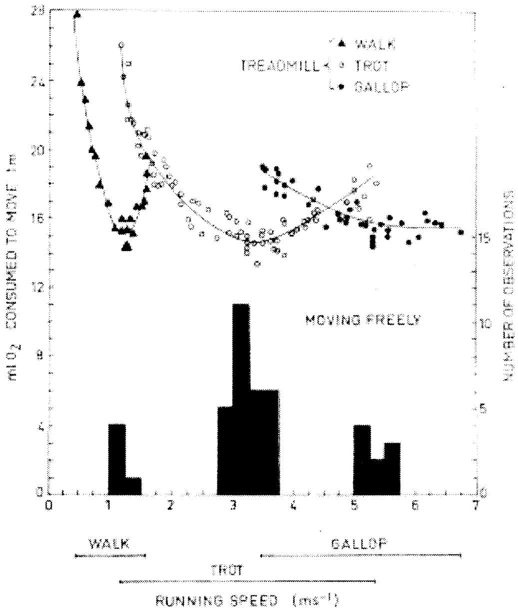
These facts and interpretations exemplify a paradigm that has been explored for some time by psychologists (Estes 1993). Why would a process of experience-driven category formation provide a realistic scenario also for speech? A comparison of exemplar-based accounts with more traditional approaches throws some light on this question. The differences stand out particularly clearly in the light of speech development.

Exemplar models make minimal assumptions about 'initial conditions' and are therefore not guilty of 'resolving' issues such as the variability problem by postulating unknown, innate mechanisms. They make the most of the signal and its complex, but lawfully structured, variability before positing hypothetical decoding mechanisms whose operation is also yet to be specified.

As pointed out by Johnson & Mullenix (1997), traditional accounts assume representations (e.g., phoneme-sized units) to be simple (context-free invariants). The task of deriving such units from the speech signal calls for complex processes capable of extracting invariants. Mechanisms of this type have been proposed—e.g., the 'phonetic module' of the motor theory (Liberman & Mattingly 1985), the 'smart mechanisms' of direct realism (Fowler 1986) and the 'top-down' processes (reconstructive rules, inference making and hypothesis testing) of cognitively oriented approaches. How do these processes operate on-line in adult listeners, and how do they develop? It is fair to say that our understanding of these issues is currently far from complete.

Exemplar-based models adopt the opposite perspective, assuming representation to be complex and mapping to be simple. Categories form as emergent products of cumulative phonetic experience. The hypothesis is that signal statistics will go a long way towards establishing units. Right or wrong, these models commendably do address the mapping and the ontogeny problems. They say no both to nativisim and to mere descriptive curve-fitting.

(2) Movements shaped by a minimum-energy criterion.



4. Role of motor constraints

4.1 Movements are shaped by a minimum-energy criterion.

The claim that movements are shaped by a minimum-energy criterion rests solidly on a large body of physiological studies (McArdle, Katch & Katch 1996). A great deal of quantitative data is available on how much energy various species use during locomotor tasks.

This information is presented by plotting the amount of energy that the subject expends against traveling speed. An example is the data on horses walking, trotting and galloping, see upper part of (2) (adapted from Hoyt & Taylor 1981). When the energy used is expressed per unit distance traveled, the measurements tend to be U-shaped and have distinct minima. Significantly, these minima are found at speeds that subjects spontaneously adopt when moving freely and unconstrained by the speed of a laboratory treadmill (lower part). Experimental biologists interpret such findings to suggest that locomotion (human walking and running, bird flying, fish swimming, etc) is shaped by a criterion of 'minimum-energy expenditure'.

4.2 Is energetics relevant to speech?

The phonetician naturally asks: Are speech movements and whole body movements similarly organized?

The energy costs of speech movements are likely to be small in comparison with those of locomotion. Therefore our brains might say: 'Be my guest! Do anything you like'. On the other hand, being both a tinkerer and a miser, evolution tends to be parsimonious, which suggests that the same rules ought to apply for small and for big movements. Until measurements of speech energy costs are made, we should of course keep an open mind on these issues. However, several factors favor parsimony. First, energetics is needed to explain a number of general characteristics of spoken language (Lindblom 1983). Second, it appears highly plausible that energetics leads the child to those characteristics, in a manner analogous to how it shapes the locomotion of animals.

What aspects of the world's phonologies are shaped by energetics? Why would children need help in their articulatory search for those aspects?

4.3 Sound patterns as adaptations to motor constraints

Admittedly, 'articulatory ease' is a controversial topic in phonetics (Ladefoged 1981, Ohala 1981). However, by placing this issue in its broader biological context, we might be able to resolve it.

Whenever the issue of 'ease' arises, it is helpful to ask, '*Could it have been otherwise*?' As we ponder that question, it becomes easier to see that, in comparison with other activities, speech is a *pianissimo* phenomenon (Lindblom 1983). Languages make fastidious use of the gestural and acoustic possibilities in principle available for sound production (cf the DOF problem introduced below). This conclusion is reinforced by a number of observations. We should mention analyses of the development of speech production, the universality of the syllable and the typology of phonotactic structure. Further support comes from the facts of segmental dynamics and the phonetic contents of vowel and consonant inventories. Why is the transition from babbling to early phonetic forms continuous? Why do these forms show such a strong assimilatory organization (e.g. the 'co-occurrence patterns' reported by MacNeilage & Davis 1993)? Why do all languages have syllables? Why do favored syllable structures tend to be arranged phonotactically according to the so-called 'sonority hierarchy'? What is the origin of coarticulation and lenition phenomena such as assimilations, reductions and deletions? Why do phonetic segment inventories tend to be organized in implicational hierarchies (cf. Size Principle, Lindblom & Maddieson 1988)? The present claim is that all of these widespread aspects are not there by accident but constitute phonological adaptations to production constraints.

4.4 The 'degrees of freedom' problem in motor control

The preceding remarks introduce an issue known as the *degrees-of-freedom* (DOF) problem. Motor systems offer their users an extremely rich set of possibilities for executing a given task. In principle, there is an infinite number of trajectories that a movement from one point to another could take. Solving the DOF problem means finding a unique movement in a very large search space.

Like other motor mechanisms, speech production offers talkers countless possibilities for any given task. Articulatory modeling (Lindblom & Sundberg 1971, Maeda 1991) has shown that there is a continuous trade-off between jaw opening and tongue raising in producing a given vowel, e.g., an /i/. The normal way of making this sound is to raise the jaw and adopt a moderately palatal tongue shape. However, experiments have demonstrated that, when speakers are asked to produce a normal-sounding /i/ with an atypically large jaw opening maintained by a 'bite-block' (Lindblom et al 1979), their output does not approach a predicted /ɛ/-like quality. In fact, subjects are able to match the normal quality and formant pattern of the vowel quite closely, a result that clearly indicates a compensatory mode of articulation. X-ray data (Gay et al 1981) have confirmed this interpretation showing that, for bite-block /i/:s, subjects compensate by raising the tongue higher than normal into a super-palatal position. Consequently the DOF problem definitely also applies to speech.

4.5 Solving the DOF problem.

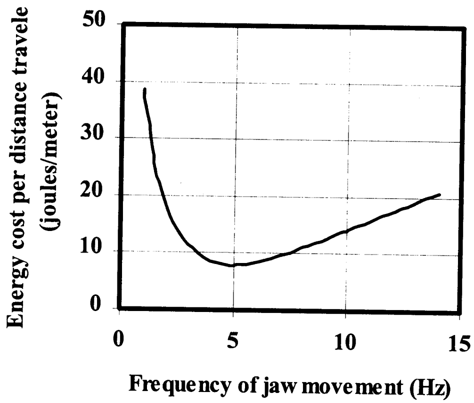
Some recent computational research on human walking (Anderson & Pandy 1999, Pandy in press) offers an elegant solution to the DOF problem. These investigators constructed a 3-D model of a musculo-skeletal system with 54-muscles and 23 degrees-of-freedom. The model and human subjects walked at a forward velocity of 81 m/min. Predicted displacements of anatomical structures were quantitatively similar to experimental observations. Muscle coordination patterns were consistent with EMG data from the subjects. Metabolic energy was expended at a rate comparable to that for human subjects. These predictions were made when a performance criterion of *least metabolic cost* was used.

Clearly, with so many muscles and mechanical dimensions, this type of model has a significant DOF problem. However, the success of the simulations implies that the optimization criterion drastically reduces the search space and makes it possible for the algorithm to identify unique and optimal movement trajectories for each subtask.

4.6 Estimating energy costs for jaw movement

A simplified model of the mandible was used to make some estimates of energy costs for speech (Lindblom et al 1999). The jaw was represented by a system defined by its mass (m), damping (b) and elasticity (k). The mass was equal to 250 g. A resonance frequency of 5 Hz was assumed.

(3) Energy costs of simulated jaw movement



The figure in (3) shows an estimate of energy consumption calculated for sinusoidal jaw movement of a 10 mm amplitude. We note that, despite the model's simplicity, the curve is similar to the results of (2) and those reported in the literature. It is U-shaped and has a distinct minimum.

4.7 Articulatory boot-strapping: 'Easy-way-sounds-OK'.

Given the information in (2) and (3), let us consider the following scenario: Suppose that young children vocalizing behave like subjects walking and running in preferring energetically low-cost movements. Further assume that, for vegetative reasons, the jaw and the area around the mouth opening are particularly salient regions of the vocal tract (Lindblom & Lubker 1985) and are therefore likely to be explored early on.

What would the articulatory and acoustic characteristics of opening and closing the jaw at minimum energy cost be like? An approximate answer is given by the minimum value of the U-shaped curve in (3). It corresponds to an open-close alternation near the jaw's resonance frequency. Combining this movement with phonation would produce a quasi-syllabic acoustic output resembling [bababa]. In other words, least effort applied to the jaw would produce an utterance not unlike *canonical babbling*.

Let us take this thinking one step further. The low-energy articulatory search (start *pianissimo*!) unfolds only a fragment of the child's phonetic space. It narrows down the range of alternative movements thereby helping the child spontaneously discover many articulatory patterns used by the ambient phonology.

We note that for this strategy to work the following must be true: (1) The DOF problem for speech is solved in the same way as it is solved for non-speech movements. This implies a strong statistical bias in favor of low-cost motor patterns. (2) Many aspects of the world's phonologies are indeed low-cost motor patterns (absence of vegetative sounds and mouth sounds, feature composition of phonetic segments, e.g., /i/ a universally 'close' vowel, syllabic & phonotactic organization, coarticulation and lenition, system-dependence of phonetic dimensions in segment inventories). (3) Theoretically, the world's phonologies could have developed less optimal motor patterns for cultural reasons, but have done so only to a limited extent. The reason is, we suggest, that sound patterns are adapted for phonetic development. Low-cost motor patterns are retained so as to accommodate the child's energy-efficient search by providing ambient reinforcement of the child's efforts. The phrase *'easy-way-sounds-OK'* captures the nature of this boot-strapping. Ambient confirmation establishes perceptuo-motor links between the emergent perceptual categories and the fortuitously discovered articulations (Studdert-Kennedy 1987, in press).

5. Role of memory constraints for the origin of combinatorial structure

5.1 Phonological units: formal idiosyncracies or adaptive emergents?

The preceding remarks provide a sketch of how phonology might get shaped by developmental phonetics. The discussion has concentrated on *substantive* aspects. Do behavioral constraints also play a role in the child's acquisition of *formal* universals of sound structure, e.g. featural and phonemic coding? We believe that it does indeed.

To argue this point in an instructive way, a simple algorithm was devised to automatically break down holistic patterns into smaller elements and then re-use those elements. The re-use implies a combinatorial organization.

The point is that the derived units are emergent consequences of system growth and that they do not have to be pre-specified. We suggest that this mechanism is formally similar to what goes on in learning phonology. We interpret the holistic patterns as Gestalt motor scores. The segmentation into smaller elements defines 'articulatory gestures'. Gestural re-use is promoted by the fact that memory storage is associated with a biochemical cost. This cost is hypothesized to derive from the energy metabolism of memory formation

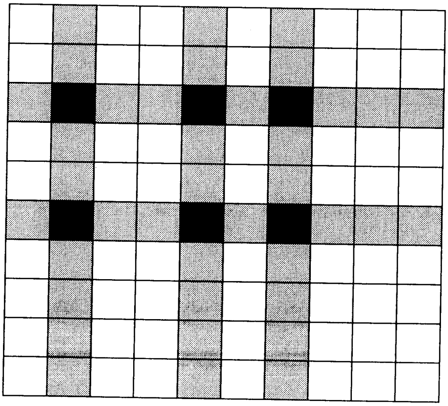
(Gonzales-Lima 1992) and is an increasing function of the novelty of the stored materials. Since novelty is expensive, holistic coding is disfavored whereas parts of these wholes can get identified and re-used.

This account of emerging structure is completely non-teleological, but once combinatorial coding comes into existence it exhibits several functional advantages. For instance, it uses memory in a manner that makes it possible to develop large vocabularies. We suggest that being able to code up to 100,000 words or more (Miller 1977) in a phonetically reliable way has an explanation at the micro-level of biochemistry. Whether that is a valid hypothesis or not remains to be seen. Nevertheless, some such explanation is clearly needed for several reasons. First, accounting for phonemic coding as due to a 'demand' for large vocabularies is teleological and therefore unacceptable. Second, explaining it in terms of the idiosyncratic nature of our genetic endowment for language (Halle & Stevens 1979) is a theoretical possibility but premature and therefore a low priority.

5.2 The nepotism game: 'close relatives get promoted'.

Here is a simple formal exercise to illustrate the above claims. Imagine a 10-by-10 matrix with 100 cells. Our task is to choose a sequence of *n* points located in the matrix so that a 'cost criterion' is minimized. We consider two alternative definitions of 'cost': (a) For every new cell we pay 1 unit! (b) For every new coordinate specification (row or column) we pay 0.5 units!

(4) Cost minimization induces combinatorial coding!



A single item is one unit on either measure. For (a), the cost is equal to *n* units regardless of the cells selected. With (b), costs can be cut by selecting a cell in a previously activated row and/or column. As *n* (system size) increases, numerous opportunities for re-use arise. (4) shows a situation with six points sequentially chosen according to the second measure. A selected cell has been marked in black. When such a choice is made, the other cells of that row and column become available at half price (0.5 units). This is indicated by the shading. Zero cost is associated with cells at intersections of already committed rows and columns. The example in (4) costs 6 units when we pay per cell (first measure), but 2.5 units when selections are priced by coordinate specifications as in (b). Thus cost minimization forces the system to go combinatorial.

5.3 Self-segmentation and the emergence of articulatory 're-use'.

What does this exercise tell us about speech? Our answer begins by reinterpreting the matrix as a crude articulatory space. We replace rows and columns by continuous parameters, say the phase and amplitude of elementary oscillatory movement. Along a third dimension we specify the articulator performing the movement. A given point in this 3-D space represents a Gestalt motor score.

Suppose a child consistently uses forms sounding like [didi], [mɤmɤ] and [bɑbɑ]. In the articulatory space these forms are represented by three points whose coordinates specify the movement parameters: e.g., three amplitude values for the open-close movement of the jaw, two positions (front and back) for the rest/target alternation of the tongue, etc. In standard notation (but without implying any segmental organization), the jaw-tongue parameters form the following matrix:

	tongue positions	
jaw openings	_i_i	
		_ɤ_ɤ
		_ɑ_ɑ

These specifications are each linked to its own type of closure movement: d_d_, m_m_, and b_b_.

The nepotism principle (NEP) literally states that a re-combination of all these hidden "component" movements is favored by the memory constraint. If NEP were consistently and mechanically implemented, it would yield the following additional potential re-use patterns for jaw-tongue movement:

	tongue positions	
jaw openings		_ɯ_ɯ
	_ɛ_ɛ	
	_æ_æ	

Moreover, it would put a number of forms in a state of 'readiness', e.g., [dɛdɛ], [dædæ], [dɯdɯ], [dɤdɤ], [dɑdɑ], [mimi], [mɛmɛ], [mæmæ], [mɯmɯ], [mɑmɑ], [bibi], [bɛbɛ], [bæbæ], [bɯbɯ], [bɤbɤ]. Again no segmental organization is implied.

How does this re-use come about? How are the “component movements” identified? The quotation marks around “component” are important, since so far we have little reason to treat phonetic forms as anything but Gestalts.

As a first step towards an answer, we note that the vocal tract consists of several independently controllable structures. In other words, although early vocalizations do not arise from phoneme-like control signals, the system producing them is in fact anatomically ‘segmented’.

Second, we observe that, in many cases, neural representations are *somatotopically* organized (Kandel & Schwartz 1991) which means that the brain stores individual motor and sensory activities in specific locations with anatomical identity preserved (cf notion of *homunculus*). Both of these circumstances play a crucial role in the proposed self-segmentation process.

Faced with the task of producing ambient forms not yet acquired, the child must solve the problem of assembling new motor programs. NEP predicts that the speed and accuracy of imitation, spontaneous use and recall will depend significantly on whether or not the new form shares “component” movements with old forms. Assembling a new motor score is assisted by overlap with previously encoded patterns even if those patterns are part of unanalyzed wholes and have not yet been ‘defined’ as separate motor entities. For the evidence supporting this scenario see review in Lindblom (1998).

What is the mechanism underlying gestural re-use? We propose that in part the NEP bias leads the child to engage in spontaneous articulatory re-use, in part the native language favors forms that match the output of NEP. Learners can thus use NEP to find ‘hidden’ structure.

The present account resembles the role sketched previously for motor constraints in the acquisition of phonology. The argument common to both sections is that behavioral conditions make certain patterns more functional than others. Languages are molded by those functional constraints. They adapt to them, incorporating fossils of naturalness in their architecture, and by so doing they become more learnable and easier to use.

6. Summary: There *is* no ’hidden’ structure!

How do children find the ‘hidden’ structure of speech? This question presupposes that ‘structure’ is something disembodied. In other words, it is seen as embedded in an incomplete, degraded, noisy and infinitely variable signal. That is the traditional, but, in our view, not necessarily the correct assumption. Instead the following approach is advocated.

Phonetic variations are far from random. They are patterned in principled ways because of perceptual distinctiveness, articulatory dynamics and VT acoustics. A cumulatively growing, exemplar-based phonetic memory should go a long way towards revealing that patterning to the child. In such a model ‘categories’ do not resemble the neat, operationally (*sic!*) defined units of classical phonemic analysis, since their correlates are likely to be strongly contextually embedded, in a sense ‘hidden’. However, over time variability would get sorted and disambiguated by context and by the cues providing semantic and situational labeling. Mapping simple, representation complex.

One source of information for perceptual labeling is articulatory. Research on non-speech offers the phonetician valuable clues as to how motor processes operate. The role of metabolic cost in solving the DOF problem is a case in point. We have made the parsimonious assumption that speech movements are organized like other movements. Therefore energetics should be relevant. From that conclusion we were led to propose a two-part hypothesis: *Easy-way-sounds-OK!* It says (1) that children initially explore their vocal resources in an energetically low-cost mode and (2) that sound patterns have adapted to reward that behavior. This is a kind 'conspiracy' that makes children stumble on motorically motivated phenomena in the ambient language. Syllabic organization is one of them. It also establishes motor links to perceptual forms (cf imitation).

A related scenario was sketched for the development of the phonemically coded lexicon. We suggested that a linguistic system with featural and phonemic recombination humors learners whose memories charge a metabolic fee for storage. If that fee increases with the number of bits (amount of information) to be stored, it follows that patterns that do not share materials (Gestalts) are costly, whereas patterns with overlap are cheaper. Somatotopic organization and VT anatomy were found to impose a segmentation of this overlap into articulator-specific parameters. This is the process that implicitly defines the '*phonetic gesture*' for the child. Metabolically controlled *re-use* is thus launched and paves the way for cognitively driven and combinatorial vocabulary growth. These considerations favor the view that phonemic coding is an adaptive emergent rather than a formal idiosyncrasy of our genetic endowment for Language.

7. Conclusion: Say no to the 'inescapable dogma'.

Emergent phonology is proposed to promote a new vision of the relationship between phonetics and phonology. By substituting it for the traditional division of labor, we would get away from what Chomsky (1964) calls the 'inescapable dogma' of 20th century linguistics: the *logical priority of linguistic form over substance.*

The distinctions between form/substance and competence/performance, having served their historical purpose, should be abandoned. There is *no split* between analog phonetics and digital phonology because, from the developmental point of view, phonology remains behavior and continues to be analog. Phonology differs qualitatively from phonetics in that it represents a new, more complex and higher level of organization of that behavior. For the child, phonology is not abstract. It represents an *emergent* patterning of phonetic substance.

References

Anderson S R (1985): *Phonology in the twentieth century*, Chicago:Chicago University Press.

Anderson F C & Pandy M G (1999): "A dynamic optimization solution for vertical jumping in three dimensions", *Computer Methods in Biomechanics and Biomedical Engineering*, 1-31.

Chomsky N (1964): "Current trends in linguistic theory" 50-118 in Fodor J A & Katz J J (eds): *The structure of language*, New York:Prentice-Hall.

Davis B L & Lindblom B (1994). "Prototype formation in speech development and phonetic variability in Baby Talk", in Lacerda F, von Hofsten C & Heiman M (in press): *Emerging Cognitive Abilities in Early Infancy*, CUP.

Estes W K (1993): "Concepts, categories, and psychological science", *Psychological Science 4*: 143-153.

Ferguson C A (1977): "Baby talk as a simplified register", 219-236 in Snow C E & Ferguson C A (eds): *Talking to children: Language input and acquisition*, Cambridge:CUP.

Fernald A (1984): "The perceptual and affective salience of mothers' speech to infants", 5-29 in Feagans L, Garvey C & Golinkoff R (eds): *The origins and growth of communication*, New Brunswick:Ablex.

Fónagy I (1983): *La Vive Voix*, Paris:Payot.

Fowler C A (1986): "An event approach to the study of speech perception from a direct-realist perspective", *Journal of Phonetics 14(1)*: 3-28.

Fowler C A (1994): "Speech perception: Direct realist theory", 4199-4203in Asher R E (ed): *Encyclopedia of Language and Linguistics*, Pergamon:New York.

Gay T, Lindblom B & Lubker J (1981): "Production of bite-block vowels: Acoustic equivalence by selective compensation", *J Acoust Soc Am 69*:802-810.

Gonzales-Lima F (1992): "Brain imaging of auditory learning functions in rats: Studies with fluorodeoxyglucose autoradiography and cytochrome oxidase histochemistry ", 39-109 in Gonzales-Lima F, Finkenstädt T & Sheich H (eds): *Advances in metabolic mapping techniques for brain imaging of behavioral and learning functions*, NATO ASI Series D:68, Dordrecht: Kluwer.

Halle M & Stevens K N (1979): "Some reflections on the theoretical bases of phonetics", 335-353 in Lindblom B & Öhman S (eds): *Frontiers of speech communication research*, London:Academic Press.

Hoyt D F & Taylor C R (1981): "Gait and the energetics of locomotion in horses", *Nature 292*: 239.

Johnson K & Mullenix J (1997): "Complex representations used in speech processing: Overview of the book", 1-8 in Johnson K & Mullenix J (eds): *Talker variability in speech processing*, Academic Press.

Kandel E, Schwartz J & Jessel T (1991): *Principles of neural science* (3rd edition), New York:Elsevier.

Kluender K R, Diehl R & Killeen P (1987): "Japanese quail can learn phonetic categories", *Science 237*: 1195-1197.

Kuhl P K, Andruski J E, Chistovich I A, Chistovich L A, Koshevnikova E V, Ryskina V L, Stolyarova E I, Sundberg U & Lacerda F (1997): "Cross-language analysis of phonetic units in language addressed to infants", *Science 277*: 684-686.

Ladefoged P (1990): "Some reflections on the IPA", *J Phonetics 18:*335-346.

Liberman A & Mattingly I (1985): "The motor theory of speech perception revised," *Cognition 21:* 1-36.

Lindblom B (1983): "Economy of speech gestures", in P F MacNeilage (ed): *Speech Production*, 217-246, Springer Verlag:New York.

Lindblom B (1996): "Role of articulation in speech perception: Clues from production", *J Acoust Soc Am 99(3)*:1683-1692.

Lindblom B (1998): "Systemic constraints and adaptive change in the formation of sound structure", 242-264 in Hurford J, Studdert-Kennedy M & R Knight C (eds): *Approaches to the evolution of language,* Cambridge:CUP.

Lindblom B & Sundberg J (1971): "Acoustical consequences of lip, tongue, jaw and larynx movement", *J Acoust Soc Am 50*:1166-1179.

Lindblom B, Lubker J & Gay T (1979): "Formant frequencies of some fixed-mandible vowels and a model of speech programming by predictive simulation", *J Phonetics 7:* 147-162.

Lindblom B & Lubker J (1985): "The speech homunculus and a problem of phonetic linguistics", 169-192 in Fromkin V A (ed): *Phonetic Linguistics*, Academic Press: London.

Lindblom B & Maddieson I (1988): "Phonetic universals in consonant systems", 62-78 in Hyman L M & Li C N (eds): *Language, Speech and Mind*, London & NewYork:Routledge.

Lindblom B, Brownlee S, Davis B & Moon S-J (1992): "Speech transforms", *Speech Communication 11*: 357-368.

Lindblom B, Davis J, Brownlee S, Moon S-J & Simpson Z (1999): "Energetics in phonetics and phonology", to appear in Fujimura O et al (eds): *Linguistics and Phonetics*, Ohio State University.

MacNeilage P F & Davis B L (1993): "Motor explanations of babbling and early speech patterns", 341-352 in Boysson-Bardies B, de Schonen S, Jusczyk P, MacNeilage P F & Morton J (eds): *Changes in speech and face processing in the first year of life*, Kluwer:Dordrecht.

Maeda S (1991): "On articulatory and acoustic variabilities", *J of Phonetics 19*:321-331.

McArdle W D, Katch F I & Katch V L (1996): *Exercise physiology*, 4th ed, Baltimore:Williams&Wilkins.

Miller G A (1977): *Spontaneous apprentices,* Seabury Press: New York.

Ohala J J (1980): "Chairman's introduction to the Symposium on Phonetic Universals in Phonological Systems and their Explanation", 184-185 in

Proceedings of the Ninth International Congress of Phonetic Sciences 1979, Institute of Phonetics, University of Copenhagen.

Ohala J J (1990): “The phonetics and phonology of aspects of assimilation”, 258-275 in Kingston J & Beckman M (eds): *Papers in laboratory phonology: Vol 1. Between grammar and the physics of speech,* Cambridge:CUP.

Pandy M G (1999): “Simulating human gait: role of energetics”, to be presented at *Speech communication and language development*, Stockholm University, June 17-19.

Studdert-Kennedy M (1987): “The phoneme as a perceptuo-motor structure”, In Allport A, MacKay D G, Prinz W & Scheerer E (eds): *Language perception and production*. London: Academic Press, 67-84.

Studdert-Kennedy M (in press): “Evolutionary implications of the particulate principle: Imitation and the dissociation of phonetic form from semantic function”, In Knight C, Studdert-Kennedy M & Hurford J R (eds): *The emergence of language: Social function and the origins of linguistic form*, Cambridge:CUP.

Sundberg U (1998): *Mother tongue – Phonetic aspects of infant-directed speech*, Ph D dissertation, Stockholm University.

Acknowledgements

This research is supported by grant number BCS-9901021 from the National Science Foundation.