

Launching Language: The Gestural Origin of Discrete Infinity

Michael Studdert-Kennedy
Haskins Laboratories
New Haven, CT 06511
U.S.A.

&

Louis Goldstein
Yale University
and
Haskins Laboratories
New Haven, CT 06511
U.S.A.

‘Human language is based on an elementary property that also seems to be biologically isolated: the property of discrete infinity.’

-Noam Chomsky (2000:3)

0.1 Introduction

‘Discrete infinity’ refers to the property by which language constructs from a few dozen discrete elements an infinite variety of expressions of thought, imagination and feeling. The property ‘seems to be biologically isolated’, because it is unique among systems of animal communication. From another point of view, however, it is not isolated at all, but rather an instance of a general principle common to all natural systems that ‘make infinite use of finite means’ (Humboldt 1836/1999: 91), including physics, chemistry, genetics and language, namely, ‘the particulate principle of self-diversifying systems’ (Abler 1989).

0.2 The Particulate Principle

According to the particulate principle, the only route to unbounded diversity of form and function is through a combinatorial hierarchy in which discrete elements, drawn from a finite set, are repeatedly permuted and combined to yield larger units higher in the hierarchy and more diverse in structure and function than their constituents. The particulate units in physical chemistry include atoms, ions, and molecules, in biological inheritance, chemical radicals, genes and proteins, in language, gestures (as will be argued below), segments, syllables, words and phrases.

A parallel between languages and genetic systems has repeatedly been remarked by physicists (e.g. Schrödinger 1944), linguists (e.g. Jakobson 1970), and biologists (e.g. Jacob 1977, Pollack 1994). Jacob, for example, wrote: ‘Linguistics has furnished genetics with an excellent model. The image which best describes heredity is that of a chemical message...written...with the combination...of just four chemical radicals. The four units...are combined and permuted infinitely, just as are the letters of the alphabet throughout the length of a text. As a phrase corresponds to a segment of text so does a gene correspond to a segment of the nucleic acid fiber’ (Jacob 1977:187). Like Jakobson (1970), Jacob emphasized that for such a system to work its basic units must themselves be devoid of meaning or function. In language, only if phonetic units have no meaning can they be commuted across contexts to form new words with new meanings.

Jacob went on to observe that the principle of combining discrete units to form successive levels of a hierarchy ‘...is not limited to language and heredity ...[but]...appears to operate in nature each time there is a question of generating a large diversity of structures using a restricted number of building blocks’ (1977:188). But Jacob did not try to explain why systems as apparently diverse as language, physics, and genetics converge on a common structural principle. That was left to Abler (1989) who first recognized correspondences among Fisher’s (1930) genetical theory of natural selection, the atomic theory of physical chemistry, and Humboldt’s (1836/1999) description of language.

Fisher (1930) reasoned that if parents' characteristics were to blend in their offspring, they would vanish in an average; variation, critical to the process of natural selection, would then decrease from one generation to the next. In fact, of course, variation is conserved, or even increased, across generations and parental characters lost in one generation may reappear unmodified in the next. From such facts Fisher (like Mendel before him) inferred that biological inheritance was necessarily effected by a particulate mechanism: unbounded biological diversity can only be maintained by permutation and combination of discrete genetic entities.

Abler (1989) saw that Fisher's logic of particulate combination applied to physics, chemistry and language no less than to genetics. Moreover, Humboldt's characterization of the language hierarchy could be extended to these other domains: all four achieve unbounded diversity by '...a synthetic [i.e. combinatorial] process...[that]...creates something...not present *per se* in any of the associated constituents' (1836/1999:67). Novel structures and functions arise at each level of a hierarchy because units do not blend and disappear, but combine as integral units to form new integral units, whose properties are not limited by, and cannot be predicted from, the properties of their constituents.

We cannot derive the fire-extinguishing properties of water from the combination of hydrogen (which burns) and oxygen (which sustains burning), nor the properties of proteins from the genes that control their formation. In language we cannot derive the meaning of a word from the phonetic elements that compose it, nor the meaning of a phrase from the lexical meanings of its words without regard to their syntax. Indeed, it is precisely because the properties of units at each new level cannot be derived from the properties of their constituents that successive levels in the language hierarchy (phonology, morphology, syntax) are independent and subject to their own characteristic rules of combination.

Thus, the particulate principle rationalizes and generalizes across diverse domains the combinatorial mechanisms and the independence of successive levels in a hierarchy that standard linguistic theory adopts as axioms of linguistic analysis. The principle is a mathematical constraint to which any system that has the property of discrete infinity necessarily conforms. That is why, despite their different modalities, signed and spoken languages arrive at analogous hierarchies of phonology (or sign formation) and syntax (Klima & Bellugi 1979). By assimilating language to other particulate domains, we do not ignore the unique properties of syntax and phonology essential to its function. We do, however, emphasize the roots of language in biophysics, and the critical importance for both lexicon and syntax of the prior evolution of phonetic capacity.

0.3 Discrete Phonetic Units as Conditions of a Lexicon and Syntax

Discrete phonetic units of some kind must have emerged relatively early in the evolution of language. For, as Bickerton remarks, '...syntax could not have come into existence until there was a sizable vocabulary whose units could be organized into complex structures' (1995:51). And a sizable vocabulary could not have come into existence until holistic vocalizations had been differentiated into categories of discrete phonetic units that could be organized into words. A critical early step into language therefore was (as it still is) the breakthrough into words, or symbolic verbal reference, by means of a particulate phonetics (Studdert-Kennedy 1998, 2000).

Less often remarked, though no less important, syntax could also not have come into existence until there was a code, a phonetic form, for short-term storage of words, independently of their meaning and syntactic function, during preparation of an utterance by a speaker and comprehension of an utterance by a listener. Independent phonetic segments, devoid of meaning, are indeed taken for granted by virtually every approach to the evolution of syntax. Berwick (1998), for example, in his account of the development of hierarchical syntactic concatenation of words by the operator ‘Merge’ in the Minimalist framework starts his derivation with a ‘bag’ of unordered words, each marked by independent phonetic, formal and semantic features. Kirby (2000), for another example, models the emergence of syntax from ‘holistic’ utterances associated with decomposable meanings. His initial utterances are semantically holistic, but consist ‘physically’ of discrete symbols, randomly concatenated into strings of ‘phonetic gestures’. Thus, a necessary condition of compositional syntax (discrete phonetic units) is included in the initial conditions: compositionality can only emerge, because ‘holistic’ utterances readily fractionate along the fault lines of their discrete components. Similarly, Wray (2000), deriving words from holistic utterances as a first step into syntax, assumes that ‘...arbitrary phonetic representation developed not in the service of words, but of complete, [semantically] holistic utterances...long before words or grammar appeared’ (293). Thus, phonetic breakpoints between portions of a semantically holistic utterance (portions that, in Kirby’s and Wray’s models, eventually become words, if they happen to correlate with presupposed breakpoints in the field of reference) are built into the utterance.

Where, then, do these phonetic breakpoints come from? What is the physical basis for phonetic segments? The standard units, consonants and vowels, will not do, because they and their descriptive features are purely linguistic and therefore precisely what an evolutionary account must explain. What we require is a prelinguistic unit of motor action that takes on linguistic form and function as it is put to communicative use.

0.4 The Nature of Phonetic Units

Modern humans, speaking English at a comfortable rate, produce 120-180 words/minute, or 10-15 phonetic segments/second. On the face of it, such a rate makes prohibitive demands on both listening and speaking. For the speaker the rate would seem to exceed by far that of any other motor activity (with the possible exception of skilled musical performance); for the listener, 10-15 units/second are close to the rate at which acoustic pulses merge into a low-pitched buzz. The puzzle begins to resolve itself, however, once we recognize that the perceptually salient units of speech are not phoneme-sized phonetic segments, but syllables, and that syllables, carriers of speech melody and rhythm, are intricate patterns of simultaneous or overlapping gestures.

‘Such rates can be achieved only if separate parts of the articulatory machinery—muscles of the lips, tongue, velum, etc.—can be separately controlled, and if...a change of state for any one of these articulatory entities, taken together with the current state of others, is a change to another phoneme...it is this kind of parallel processing that makes it possible to get high-speed performance with low-speed machinery’ (Lieberman, Cooper, Shankweiler & Studdert-Kennedy 1967:446). Similarly, sign languages and finger spelling depend on rapid sequences of movement distributed across arm, wrist, and fingers.

Evidently, particulate language, whatever its modality, requires an integral neuro-anatomical system of discrete, independently movable parts that can be coordinated to effect rapid sequences of motor action, and it is the actions of these individual parts, the gestures, that constitute the basic units of language. We do not, of course, have to suppose that the particulate structure of either the vocal apparatus or the hand evolved specifically for language. The task, rather, is to understand how referential communication exploited, refined and, as it were, emancipated into language a particulate machinery that had already evolved in primates for other purposes.

Our initial hypothesis, following MacNeilage (1998), is that speech, *as a motor function*, draws on phylogenetically ancient mammalian oral capacities for sucking, licking, swallowing and chewing. Sucking (the defining mammalian behavior), licking, and tongue action in preparing food for swallowing, presumably initiated neuroanatomical differentiation of the mammalian tongue; the evolution of speech carried the process further, differentiating tongue tip, tongue body and tongue root into independent organs of phonetic action. Cyclical lowering and raising of the jaw for mastication laid the neural ground for early homologs of the repetitive lip and tongue smacks, calls and cries that we observe in modern monkeys and apes. From these, by hypothesis, arose the hominid protosyllable, precursor of the modern child's unit of babble, and of the modern adult's unit of rhythm and melody. *Such a protosyllable can be viewed as a gesture, that is, as constriction and release of one of the vocal organs, set in the context of an overall vocal tract posture and combined with phonatory action.*

Thus, early hominids chanced on the particulate principle by adopting for communicative use an apparatus already divided neuroanatomically into discrete components. Constrictions of the distinct parts could be associated with distinct meanings. At first, utterances would necessarily have been holistic, employing (perhaps repeated) constrictions of a single vocal organ, superimposed on a characteristic posture of the vocal tract as a whole. Actions of distinct organs could not yet be permuted and combined to form different utterances. Pressure for differentiation would have grown as demands for lexical increase were placed on an apparatus equipped for a limited number of discrete speech actions. The initial impetus for 'reuse' of articulators (and so for the emergence of combinatorial mechanisms) would then have come from the simple fact that articulators were few in number (cf. Lindblom 1998,2000).

Once gestures of distinct organs (and their neurophysiological support) had evolved as discrete, combinable units, expansion of the phonological system could have occurred by sociocultural processes in at least two ways without any further genetic change. First, expansion of the set of discrete gestures could have resulted from differentiating the actions of a given organ into distinct types through a process of mutual attunement, or accommodation, between individual speakers/hearers. Second, increasingly complex gestural structures could have emerged, somewhat as we now observe the process in the developing child.

In what follows, we begin by justifying the gesture as the basic unit of phonological structure, and then go on to show how the basic set of gestures might have been expanded and elaborated in specific languages by sustained speaker-hearer interactions.

0.5 Articulatory Phonology

Within the approach of articulatory phonology (e.g. Browman & Goldstein, 1992; 1995a), the most primitive (atomic) units of phonological structure are hypothesized to be units of articulatory action, or *gestures*. Since units of articulatory action can be defined and in principle observed even in the absence of a mature phonological system, this approach allows the description of phonology and pre-phonological vocal behavior with the same set of primitives. This stands in sharp contrast to most theories of phonological structure, in which the units (features) are defined relationally, in *opposition* to one another (e.g. Jakobson, Fant, & Halle 1951/1963), and therefore cannot exist, as such, in the absence of a developed system. From the perspective of articulatory phonology, we can observe the development of a phonological system in the child (and by inference, the broad course of its evolution in the species) as atomic action units come to be combined systematically into the ions and molecules (roughly, segments and syllables) of mature phonological form.

The principal theoretical motivation behind articulatory phonology is to reconcile the apparently incompatible phonological and physical descriptions of speech. The phonological description views speech as composed of a small number of discrete (particulate), permutable, context-independent units, while the physical description finds continuous, context-dependent variation in a large number of articulatory, acoustic, auditory, or neural variables. The discovery of this incompatibility (through the development of techniques for physical analysis of speech in the mid twentieth century) led theoreticians to remove phonological units from the domain of direct physical observation altogether by hypothesizing that they correspond to mental (cognitive) events but *not* physical ones. This view is well captured by Hockett's (1958) familiar Easter egg analogy. Phonological units, in the mind of the message-sender, are represented by a row of brightly colored, but unboiled, Easter eggs. The act of talking involves smashing the eggs through a wringer. The message-receiver then must act like a forensic expert, reconstructing the original egg sequence from the bits of shell and knowledge of possible sequences.

While contemporary views of the relation between the phonological and physical descriptions of speech are less colorful than Hockett's, they are similar theoretically. Phonological units are seen as exclusively mental (e.g. Pierrehumbert 1990), and implementation rules (e.g. Keating 1990) are posited to translate these discrete mental representations to their physical interpretations. Since there are no generally accepted constraints to which implementation rules must adhere (Clements 1992), the relation between the phonological and physical structures of speech can even be construed as arbitrary.

Apart from its inelegance, this bipartite view of phonology and speech has substantive problems. First, as Fowler (1980) has argued, it would be extremely odd for a communication system to evolve in which the essential properties of the system's units are destroyed in the act of conveying them. Second, as Browman and Goldstein (1995a) point out, reciprocal constraints prevail between the structure of phonological systems and the physical properties of phonological units, exactly as would be expected, if the phonological and physical descriptions of speech were, in fact, low and high dimensional descriptions of a single self-organized complex system (e.g. Kugler & Turvey 1987). For example, languages that contrast voicing in labial and coronal stops often lack a voiced dorsal stop. This tendency can be attributed to the relative difficulty of sustaining phonation under the aerodynamic conditions imposed by the different closures (Ohala 1983). More generally, suspension of contrast (neutralization) can often be understood with reference to physical

constraints. At the same time, physical properties associated with some phonological unit may vary cross-linguistically as a function of the nature and number of contrasting units in the language's inventory. For example, the amount of context-dependence exhibited by a particular vowel may vary as a function of the number of other vowels in a language's system (Manuel 1990). Thus, certain constraints on the physical properties of speech stem from the structure of phonological inventories, and certain constraints on phonological inventories stem from physical properties of speech.

0.5.1 The gestural hypothesis: units of information as units of action

Articulatory phonology resolves the apparent incompatibility between the phonological and physical (phonetic) descriptions of speech by hypothesizing that speakers deploy discrete, context-independent units of action, permuting and combining them to form the contrasting words of the language. Thus, articulatory phonology equates the most basic units of phonological information with units of vocal tract action. In this view, the apparent continuity of speech stems from the fact that measurements of speech record the *results* of the activities that go on in speakers' vocal tracts when they talk, but not the activities themselves.

As an example of an action unit, consider the word 'bad.' If we examine what takes place in the vocal tracts of speakers of English as they produce this word, we expect to see changes in the moment-to-moment positions of all the vocal tract articulators, and differences in that pattern of changes across speakers and across contexts in which the word is spoken. However, we also expect something to be the same across speakers and contexts. At the beginning of the word, the two lips come together to form a seal: the upper lip is displaced downward and the lower lip is displaced upward, due both to their own muscular activations, and to the raising of the mandible. This pattern of activities constitutes a distinct action unit, a lip closure *gesture*. By hypothesis, this same action unit is deployed by speakers when they produce words like 'mat,' 'pint,' and 'lob,' in coordination with other gestural actions.

A lip closure gesture is not identical to a traditional phonological segment or feature. It is clearly 'smaller' than a segment, as several different segments involve lip closure gestures (/p/, /b/, /m/). It is also 'larger' than a feature, as it would be specified as [labial], [-continuant], [+consonantal]. However, it is a viable choice as the basic permutable element (atom) of phonological structure, because larger structures (syllables, words) can then be viewed as molecules composed of (coordinated) gestural atoms, and segments can be viewed as coherent substructures of gesture (corresponding to ions in the chemical analogy) that recur in many distinct molecules. Not only is such a view of phonological structure possible, but Browman and Goldstein (e.g. 1992, 1995b) have argued that certain kinds of phonological alternations (casual speech, allophony due to syllable position) are better described with gestural structures than with traditional feature-based representations. Gafos (2002) has shown the necessity of positing gestural structures in accounting for some cases of more abstract allomorphy.

0.6 Bases for discrete action units: Organs of the vocal tract

The nature of the action units posited by articulatory phonology and the basis for their participation can provide some insight into how such a system could have evolved. The human vocal tract can be viewed as composed of a small number of relatively independent constricting devices: the lips, tongue tip/blade, tongue body, tongue root, larynx, and velum. While these devices, or *organs* as we shall refer to them, share some common articulatory components (for example, jaw movement contributes to lip, tongue tip, and tongue body constrictions), each of these organs can constrict without causing a concurrent constriction of one of the other organs. Action units are hypothesized within articulatory phonology to be constriction gestures performed by one of these organs.

Gestures of different organs ‘count’ as different and provide the basis for phonological contrast. Thus, the word ‘bad’ begins with a gesture of the lips organ. It contrasts with the word ‘dad’ that begins with a constriction of the tongue tip organ instead of the lips organ, and also with the word ‘pad’ that adds a gesture of the larynx organ to that of the lips. Organs are then functional synergies or coordinative structures composed of several articulators. A gesture of the lips organ involves a functional synergy comprising the upper lip, lower lip, and jaw articulators.

Of course, not all contrasting speech gestures differ in the organ employed. For example, the words ‘tick,’ ‘thick,’ and ‘sick’ all begin with gestures of the tongue tip organ, and most vowels are composed of tongue body gestures (with or without a lips gesture). Gestures of a given organ may also differ in the metric properties of the constrictions that they create: for example, in the degree of constriction (‘tick’ vs. ‘sick’), or in the location within the vocal tract of the constriction (‘sick’ vs. ‘thick’). However, evidence from the nature of phonological systems and from phonological development in children suggests that *the most basic phonological contrasts are those that involve different organs*.

0.6.1 Discrete organs in phonologies.

The phonological system of almost every language includes a contrast between Labial (lips), Coronal (tongue tip) and Dorsal (tongue body) stops, and a contrast of Nasality (velum) in stops is almost as common. These are all between-organ contrasts. The importance of organs can also be seen in the phonological alternations that languages exhibit. Within the theory of Feature Geometry, features are arranged in a hierarchy that captures their behavior in phonological processes, such as spreading (assimilation) and delinking (neutralization of contrast). A higher feature node dominates its dependents in the sense that when a given feature node is involved in an operation (such as spreading), all the dependent nodes necessarily participate in the operation as well. The converse is not true. Feature nodes that refer to the organ employed stand near the top of this feature hierarchy. For example, in McCarthy’s (1988) geometry, Nasal, Laryngeal, and Place (in turn branching into Labial, Coronal, Dorsal, and Pharyngeal) are branches at the top of the feature tree.

0.6.2 Discrete organs in children’s systematic errors

Evidence for the somatotopic decomposition of the vocal tract into distinct organs comes from systematic correspondences between children’s and adult’s utterances. Corpora

of such correspondence ‘errors’ have little to offer on this topic, however, if they simply classify and tabulate errors according to standard phonetic features. Olmsted (1971), for example, collected perhaps the largest such corpus: over 3,000 errors from half-hour interviews with 100 English-speaking children, aged from 1;3 to 4;6. He classified consonant errors by place, friction, voice and nasality, and found place errors to be far the most common. If we rescore the data (Olmsted 1971: 71, Table 13), however, classifying responses by whether the correct organ (lips, tongue tip, tongue body, velum) was employed or not, friction errors are now most frequent, with organ and voice errors roughly equal. Yet we still cannot tell whether any particular error resulted from paradigmatic substitution of a wrong gesture or from syntagmatic shift of the correct gesture to a wrong point in the utterance (Studdert-Kennedy & Goodell 1995, Studdert-Kennedy 2002). For an adequate test of the somatotopic hypothesis we need a different type of corpus. First, we need data from children during that narrow developmental window in which they are attempting their first words, before they have gained enough control over the amplitude and timing of their gestures to make random adult-like speech errors. Second, we need full transcriptions of a child’s actual utterances and their presumed targets, so that we can determine the nature of the errors.

Data of this kind have come from several sources, although they have not been analyzed in terms of gestures (e.g. Menn 1983, Vihman & Velleman 1989). Ferguson & Farwell (1975), for example, showed that standard phonetic transcriptions of a child’s earliest words (first 50) are quite variable phonetically: the initial consonant in a given word (or set of words) produced by the child is transcribed as different phonetic units (‘segments’) on different occasions. Ferguson and Farwell (1975) argue that the variability is too extreme for the child to have a coherent phonemic system, with allophones neatly grouped into phonemes, and that the basic unit of the child’s production must therefore be the word, not the segment. If we assume further, and not incompatibly, that the child is not producing segments at all, but rather (gross) constriction gestures of the vocal organs that vary in their exact degree and location, it turns out that children are remarkably consistent in the organs they move for a given word (Studdert-Kennedy 2002). Thus, children appear to be acquiring a relation between actions of distinct organs and lexical units very early in the process of developing language. This relation is apparently not mediated by a phonological unit specified for additional features. At this stage, it is simply organ identity itself that infants employ to differentiate lexical items.

0.6.3 Discrete organs in vocal and facial imitation.

The preceding discussion takes for granted a capacity for vocal imitation, or vocal learning, that is unique among primates to humans (Hauser 1996). The role of imitation in language acquisition is often discounted, because children developing syntax do very much more than repeat what they have heard. Whatever more they do depends, however, on their first building a repertoire of words by copying their companions. For the modern child (as perhaps for the early hominid) vocal imitation is the enabling step into a particulate phonetics without which a particulate (or compositional) syntax could never develop (Studdert-Kennedy 2002). To imitate a spoken utterance, imitators must first find in the acoustic signal information specifying which organs moved, as well as where, when and how they moved with respect to each other, and must then engage their own articulatory

systems in a corresponding pattern of action. Such a skill requires, as suggested above, that the vocal tract be represented somatotopically in the brain (Lindblom 1998,2000). How might such a representation have evolved?

Here the facts of facial imitation, also unique among primates to humans (Hauser 1996), lend insight. Particulation of the mouth and face into distinct organs can be seen in the facial mimicry of which very young infants, even neonates (1 hour old), have been shown to be capable (Meltzoff and Moore, 1997). In these experiments, when the experimenter performs an action with some facial organ (e.g. lips, tongue, eyes), the infant produces some movement of the *same organ*, even though the movement may not be the same as that of the adult (at least not on the first try). For example, if the experimenter protrudes the tongue from the mouth sideways, the infant will also protrude the tongue, but perhaps straight out. Thus, for neonates, the first step into facial mimicry is to match organs: infants evidently individuate their own oro-facial organs and identify them with another's. Organ identification is a fundamental part of Meltzoff and Moore's (1997) model of early facial mimicry for which the infant has only optic information about the model and only kinesthetic information about his/her own face.

Organ identification, as we have just seen, is also the first step in a child's early attempts to imitate words. The close link between facial expression and vocal tract configuration was indeed remarked by Darwin (1872/1998:96), and is evidenced in modern humans by audiovisual interactions in the perception of speech and by their capacity for lip- or speech- reading. Neural support for facial and vocal imitation may come from 'mirror neurons' of the type discovered for manual action in the macaque monkey by Rizzolatti, Fadiga, Gallese and Fogassi (1996). Such neurons discharge not only when an animal engages in an action, but also when it observes another animal engage in the same action.

We do not have space to develop the argument here (see Studdert-Kennedy 2002), but we are encouraged by the salience of facial expression in human communication to hypothesize that: (i) facial imitation evolved in the mimetic culture of early hominids (Donald, 1991) by duplicating a mirror system already established for manual action in lower primates, and (ii) vocal imitation evolved by coopting and extending the facial mirror system with its characteristic somatotopic organization. Evidence consistent with activation of a mirror neuron system in the perception of speech has indeed already come from transcranial magnetic stimulation studies (Fadiga, Craighero, Buccino and Rizzolatti 2002), and we believe that our hypotheses concerning the mechanisms of facial and vocal imitation should be amenable to further experimental tests by brain imaging and other techniques.

In any event, it is evidently the capacity for vocal imitation that supports mutual attunement of individuals' speech. And it is by attunement, we hypothesize, that the universal set of organ-based gestures is enlarged and elaborated in specific languages.

0.7 Expansion of Phonological Systems by Attunement among Individuals: Particulating Gestures with a Common Organ

Within articulatory phonology, gestures are modeled as task-dynamical systems that regulate the formation of constrictions by one of the vocal tract organs (e.g. Saltzman and Munhall, 1989). The dynamical specification for each gesture includes a spatial goal, which is the equilibrium position for the (point attractor) dynamics. For example, in tongue tip gestures, a goal is specified for two vocal tract variables: tongue tip constriction degree

(distance of the tongue tip from the palate), and tongue tip constriction location (location of the constriction along the palate/upper tooth surface). So the oral constriction gestures at the beginning of words like ‘tick, thick, sick’ differ in degree and location, which, of course, define continuous metric spaces. How do these spaces come to be partitioned into discrete regions (or points) that can serve as contrasting gestures?

One possible answer is that partitioning emerges when multiple vocal tracts *attune* themselves to one another by mutual vocal mimicry. While we can describe speech actions as events within an individual (I-language), the actions must also be shared by the members of a speech community (E-language) (Chomsky 1986). Attunement can be observed in phonological development as early as six months of age when infants begin to modify their speech actions (babbling) so as to approximate quantitative properties of the speech to which they are exposed. Of course, particulation of a continuum by attunement to the ambient language depends on the continuum being already partitioned in that language. From an evolutionary perspective we must ask where this partitioning comes from in the first place.

0.7.1 Consonant gestures

We can test the possible evolutionary role of attunement in partitioning a continuum by means of computational simulations in which none of the participants (or ‘agents’) begins with the continuum broken into categories. Browman and Goldstein (2000, forthcoming) report the results of such simulations in which all agents begin by producing the points along some continuum at random, with uniform density distribution along the continuum. The agents attune by comparing their gestures to the gestures they perceive their partner(s) producing, and if gestures match within some criterion, agents increase the likelihood of re-producing the matching continuum points. The outcome of repeated cycles of interchange depends on the nature of the articulatory-acoustic mapping. If the articulatory continuum is homogeneous, but the acoustic ‘continuum’ includes regions of spectral discontinuity, non-random partitioning can emerge. For example, suppose that the acoustic consequences of a tongue tip constriction degree continuum have the step-like, quantal character described by Stevens (1989): ‘stable’ regions of silence, turbulent noise and low amplitude vocalic phonation, in which there is little acoustic change as constriction degree is varied, bordered by “unstable” regions of rapid acoustic change. When agents recover their partner’s constriction degree from the acoustics of such a system in the presence of noise, their behavior converges on three ‘categories’ corresponding to stops, fricatives, and glides. That is, the initial random distribution of constriction degree across the continuum changes to a tri-modal distribution, with peaks in the stable regions. Only these stable regions afford attunement. Thus, the discrete categories are density distribution maxima that emerge when randomly acting vocal tracts attune to one another, under constraints of the articulatory-acoustic mapping. They are attractors of the dynamical system of attunement.

On the other hand, if both articulatory and acoustic continua are homogeneous, the outcome of attunement is that agents simply converge on a narrow range of values along the articulatory continuum. This might be the case for within-organ contrasts such as constriction location among tongue tip consonant gestures (e.g. dental, alveolar, retroflex), for example. Organ differentiation and articulatory-acoustic mappings will lead a language

to evolve a stop gesture produced with the tip of the tongue. However, as speakers attune, they converge on a range of locations for tongue tip stops that is narrower than the range of anatomically possible tongue tip stops. Such a narrowed range could then become a point of reference for other gestures that might emerge along the continuum, and so be subject to competition with these new gestures for motor and perceptual efficiency. Notice that, on this account, gestures are, in no sense, targets aimed at by the emerging system. They arise from random search, driven by pressures on speakers to enlarge their shared lexicons and to ‘sound like one another’.

0.7.2 Vowel gestures

Vowels are not entirely straightforward from the perspective of constricting organs. On the one hand, all vowels involve gestures of the tongue body, with differences between vowels being analyzable as differences of constriction location (palatal, velar, uvular, pharyngeal) or degree. On the other hand, the vowel /u/ involves a lip gesture in addition to a tongue body gesture, and it would be possible to analyze the vowel /a/ as involving a tongue root gesture in addition to a tongue body gesture. Thus, the most basic vowels found in almost all human languages (/i, a, u/) can be viewed as gestures that employ distinct organs. With these vowels as reference points other vowels could emerge by attunement, as described above.

De Boer (2001) has proposed a related process of vowel evolution. He presents simulations of how vowel systems like those of human languages (in the number of vowels and their distribution) can self-organize through the interaction of agents attempting to imitate one another within the constraints of a modern vocal tract. None of the agents possesses a vowel system at the outset, and vowels are added to growing systems both at random and in response to imitative failures. Strikingly, pressure to add vowels to a system does not come from the need for linguistic contrast (which has no role in these simulations), but from the need to optimize imitative success. However, the form of the simulation is such that it presupposes the concept of discrete units, so that the issue of where agents get the idea that vowels are discrete points within their articulatory capabilities is never addressed.

Another approach to the emergence of vowels, viewed as tongue body gestures that differ parametrically, comes from the theoretical work of Carré and Mrayati (1995). These authors show that the vocal tract as an acoustical transmission system can be divided into distinctive regions that have the property of maximal acoustic efficiency: for constrictions within these regions, small changes in vocal tract area result in large perturbations of formant frequencies away from the values for a neutral tube. They explain the evolution of vowel systems on this basis: vowel gestures are characterized by maximally efficient (acoustically) constrictions and thus fall within the distinctive regions. Carré and Mrayati (1995) do not themselves distinguish within- from between-organ differentiation of gestures. For them, the basis for articulation is purely acoustic. However, while the relevant attunement simulations have not been done, it is not hard to imagine that differential acoustical efficiency would cause the attunement process to partition tongue body constrictions into the locations (although not the degrees) that are observed in vowel systems.

0.8 Expansion of Phonological Systems by Development of Complex Gestural Structures

Evidence for another possible way of expanding the phonology of a language comes from the order of acquisition of English consonants. As we have seen, gestures in articulatory phonology are atoms that combine with other gestures to form molecules corresponding to larger phonological units such as segments, syllables, feet, or words. Combining gestures involves establishing a pattern of (temporal) coordination among them. Words such as ‘cab’ and ‘back’ are composed of the same gestures coordinated in different ways. Segment-sized phonological units (vowels or consonants) can be viewed as ions, combinations of atoms that recur in many different molecules. Crucially, segmental ions differ in their internal complexity. Some consonant segments are composed of single gestures, while others are complex, requiring the coordination of multiple gestures. This difference in complexity has implications for language change (cf. Lindblom and Maddieson, 1988), but also, as argued below, for the course of phonological development, and so, by hypothesis, for the evolution of phonological structures (cf. Kohler 1998).

While the order of acquisition of individual consonant segments in English-learning children varies widely across individuals, Dinnsen (1992) finds that all children’s consonant inventories (in a sample of 40), at any point in time, fall into one of the five classes of segments shown here in (1a-e). These classes exhibit increasing levels of complexity: Dinnsen defines the classes as elaborations of a feature “tree,” with additional features being recruited at each level. Here we will see that the succession of levels follows from the inter-gestural coordination required to produce the segment types of each level.

- (1) a stops (voiceless unaspirated), nasals, glides and [h]
- b all segments in (a), plus voicing contrast in stops
- c all segments in (b), plus fricatives and affricates
- d all segments in (c), plus one liquid ([l] or [r])
- e any English segment

The consonants in level a are just what we would expect if children were producing poorly controlled, uncoordinated gestures of vocal tract organs. Note that (i) the almost interchangeable appearance of stops and glides of the same organ in the phone classes of children’s first 50 words (Ferguson and Farwell 1975) suggests that degree of constriction is not being systematically controlled; (ii) the fact that stops are voiceless unaspirated indicates that oral and laryngeal gestures are not being coordinated—voiceless unaspirated is the expected phonatory consequence of a stop closure in the absence of any consonant-specific laryngeal adjustment; (iii) a gesture of the larynx (glottal abduction) in the context of regular phonation will be perceived as [h]; (iv) a nasal stop will be perceived if a child *happens* to lower the velum and effect an oral constriction at roughly the same time: a wide range of overlaps can be perceived as nasal. This interpretation is consistent with the appearance of oral and nasal stops of the same organ within phone classes of Ferguson and Farwell (1975).

The appearance of aspirated stops at level b heralds the ability to coordinate gestures in a stable fashion: production of aspirated stops requires coordination of an oral constriction gesture with glottal abduction. Interestingly, level b tends to occur exactly when the first CVC words (as opposed to CV) are found in children’s productions (Stoel-

Gammon, 1985) Note that a CV word can be produced by a single organ forming a constriction and release without any precise coordination of consonant and vowel gestures. By contrast, a CVC word requires precise intergestural coordination—either consonant gestures to vowel, or consonant gestures to each other. Otherwise, the temporal structure of a CVC syllable will not be perceived.

The voiceless fricatives and affricates of level c could not be expected any earlier than the aspirated stops of level b, as they require laryngeal-oral coordination. But why do fricatives come later? Note first that the most common early fricative in children's inventories is [s] for which the positioning of the lower teeth (by action of the mandible) must be coordinated with the tongue tip constriction against the alveolar ridge, a problem in coordination that does not arise in the production of stops or glides. A second possible reason for the delay is that fricatives and affricates require more precise control of constriction degree than do stops or glides, in order to form a constriction just narrow enough to produce turbulence.

At level d, the child begins to produce a liquid ([r] or [l], but not both). English liquids involve coordination of multiple oral constriction gestures which is presumably more difficult than coordination of oral and laryngeal gestures seen at level b. Evidence for liquids as a coordination problem comes from children's common "errors" of omission. For example, [w] for [r] follows from omission of the tongue tip gesture, [j] for [l] from omission of the tongue body retraction gesture. On this account, other kinds of segments (non-liquids) that require coordination of multiple oral constriction gestures should also be acquired relatively late, as is evidently the case for click consonants (Herbert, 1990). Conversely, liquid consonants that do *not* require coordination of multiple oral constrictions, could be acquired relatively early, as seems to be the case for the trilled [r] of K'iché (Ingram 1992).

At level e, a second liquid is acquired. If coordinated gestures of mechanically linked organs is difficult (level d), then imposing more than one pattern of coordination on those organs should be even more difficult.

In summary, empirical regularities in the order of acquisition of English consonant segments support the hypothesis that gestures are the basic units of phonological structure and that phonological development is a cumulative, hierarchical process by which gestures are combined into larger and increasingly complex structures.

0.9 Summary and Conclusions

According to the particulate principle of self-diversifying systems, a necessary condition for the evolution of discrete infinity, a property on which every language, spoken or signed, is based, was evolution of a finite set of discrete phonetic units. Since production is logically prior to perception, a necessary condition for the evolution of these units was an integral neuroanatomical system of potentially independent movable parts (such as the hand, the face, or the vocal apparatus) that could be differentiated to effect rapid, coordinated strings of discrete motor actions, or gestures.

Under pressure for mutually intelligible imitative exchange, the vocal apparatus differentiated evolutionarily into six neuromotorically independent structures (lips, tongue tip, tongue body, tongue root, velum, and larynx) capable of effecting discrete changes in vocal tract configuration. By hypothesis, these organs provide the universal biologically

determined base for discrete units of phonetic action in every language. Different languages then expand and elaborate the basic gestural phonology by sociocultural processes of attunement among speakers through mutual vocal mimicry, resulting in differentiation of gestures produced by a given organ and coordination of gestures into larger complex structures. The imitative processes of learning to speak and of mutual attunement among speakers of a language may have evolved by means of specialized, somatotopically organized systems of mirror neurons of the type observed for manual action in the macaque monkey.

Key Further Readings

In the preceding paper, we took for granted that early evolutionary steps into language entailed differentiation of the primate vocal apparatus and its neural support, under pressure for lexical growth, but we did not consider proximate pressures (e.g. toward ease of articulation and perceptual distinctiveness) that must have shaped the morphology, modes of action and acoustic outputs of the apparatus.

For an assessment of the role of an enlarged pharynx, relative to that of other primates, in increasing phonetic scope, see Fitch (2000). For an account of the role of acoustic contrast in differentiating the vocal tract, see Carré and Mrayati (1995). For discussion of language-independent neuromuscular and biomechanical constraints on the evolution of a basic, universal set of consonants and vowels, see Kohler (1998).

A crucial step in speech evolution, according to MacNeilage, was adoption of the open-close alternation of the primate mandible (previously used for chewing and for communicative lipsmacks) as the 'frame' for syllables. Subsequent expansion entailed differentiation of the syllable into its segmental 'content', consonants and vowels. For the 'frame/content' theory of speech production, see MacNeilage (1998).

Lindblom, also concerned with derivation of phonetic segments and gestures from non-linguistic precursors, sees phonological evolution as a tug-of-war between speaker and listener, fought through words and their phonetic components, as they compete for survival within a language. Lindblom conceptualizes the strengths of phonetic segments or gestures as ratios of articulatory cost to perceptual benefit, minimized by competition and summed across the phonological system to a least squares criterion. He quantifies articulatory cost, in principle, in terms of the metabolic costs of learning, remembering and executing a phonetic unit, in such a way that *reuse* of a unit reduces its cost. Thus, discrete gestures and combinatorial mechanisms emerge automatically, not only because gestures are executed by a small number of discrete organs, as argued above, but also because gestural cost-benefit ratios are reduced by repeated use of a small number of units. For full development of these ideas, see Lindblom (1992, 1998, 2000).

Finally, for selected readings on articulatory phonology, see the following: basic motivation (Browman and Goldstein, 1986), application to sound change (Browman and Goldstein, 1991; Gick, 1999), fluent speech (Browman and Goldstein, 1990; Zsiga, 1995), syllable structure and prosody (Browman and Goldstein, 2000; Byrd and Saltzman, 1998; Gick, in press).

Acknowledgements. Preparation of this paper was supported in part by Haskins Laboratories and by grants NICHD HD-01994, NIDCD DC-00403 to Haskins Laboratories.

References

- Abler, W. (1989), 'On the particulate principle of self-diversifying systems'. *Journal of Social and Biological Structures*, 12: 1-13.
- Berwick, R. (1998), 'Language evolution and the Minimalist Program: the origins of syntax', in J.R. Hurford, et al., 320-340.
- Bickerton, D. (1995), *Language and Human Behavior*. (Seattle, WA: University of Washington Press).
- Browman, C.P. and Goldstein, L. (1986), 'Towards an articulatory phonology'. *Phonology Yearbook*, 3: 219-252.
- Browman, C.P. and Goldstein, L. (1990), 'Tiers in articulatory phonology, with some implications for casual speech'. In J. Kingston & M.E. Beckman (Eds.), *Papers in laboratory phonology I: Between the grammar and physics of speech*. (Cambridge: Cambridge University Press), 341-376.
- Browman, C.P. and Goldstein, L. (1991), 'Gestural structures: Distinctiveness, phonological processes, and historical change'. In I.G. Mattingly & M. Studdert-Kennedy (eds.), *Modularity and the motor theory of speech perception*. (Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers), 313-338.
- Browman, C.P. and Goldstein, L. (1992), 'Articulatory phonology: An overview'. *Phonetica*, 49: 155-180.

- Browman, C. P., and Goldstein, L. (1995a), 'Dynamics and articulatory phonology', in T. van Gelder & R. F. Port (Eds.), *Mind as Motion* (Cambridge, MA: MIT Press) 175-193.
- Browman, C. P. and Goldstein, L. (1995b), 'Gestural syllable position effects in American English', in F. Bell-Berti & L. J. Raphael (Eds.), *Producing speech: Contemporary issues. For Katherine Safford Harris*. (Woodbury, NY: AIP Press), 19-34.
- Browman, C.P. and Goldstein, L. (2000), 'Competing constraints on intergestural coordination and self-organization of phonological structures'. *Les Cahiers de l'ICP, Bulletin de la Communication Parlée*, 5: 25-34.
- Byrd, D. (1996), 'Influences on articulatory timing in consonant sequences'. *Journal of Phonetics*, 24: 209-244.
- Byrd, D. and Saltzman, E. (1998), 'Intragestural dynamics of multiple phrasal boundaries'. *Journal of Phonetics*, 26: 173-199.
- Carré, R. and Mrayati, M. (1995), 'Vowel transitions, vowel systems, and the Distinctive Region Model', in C. Sorin, J. Mariani, H. Méloni and J. Schoentgen (eds.) *Levels in Speech Communication: Relations and Interactions* (Amsterdam: Elsevier) 73-89.
- Chomsky, N. (1986), *Knowledge of Language* (New York: Praeger).
- Chomsky, N. (2000), *New Horizons in the Study of Language and Mind*. (Cambridge: Cambridge University Press).
- Clements, G. N. (1992), 'Phonological primes: features or gestures?' *Phonetica*, 49: 181-193.
- Darwin, C. (1872/1998), *The Expression of the Emotions in Man and Animals* (3rd edition). (Oxford: Oxford University Press).
- de Boer, B. (2001), *The Origins of Vowel Systems*. (Oxford: Oxford University Press).
- Dinnsen, D. (1992), 'Variation in developing and fully developed phonetic inventories', in C.A. Ferguson, et al., 423-435.
- Donald, M. (1991), *Origins of the Modern Mind*. (Cambridge, MA: Harvard University).
- Fadiga, L., Craighero, L., Buccino, G. and Rizzolatti, G. (2002), 'Speech listening specifically modulates the excitability of tongue muscles: a TMS study' *European Journal of Neuroscience*.

- Ferguson, C.A. and Farwell, C.B. (1975), 'Words and sounds in early language acquisition'. *Language*, 51: 419-439.
- Ferguson, C.A., Menn, L. and Stoel-Gammon, C., (1992) *Phonological Development: Models, Research, Implications*. (Timonium, MD: York Press).
- Fisher, R. A. (1930), *The Genetical Theory of Natural Selection*. (Oxford: Clarendon).
- Fitch, W.T. (2000), 'The phonetic potential of nonhuman vocal tracts: comparative cineradiographic observations of vocalizing animals'. *Phonetica*, 57: 205-218.
- Fowler, C. A. (1980), Coarticulation and theories of extrinsic timing. *Journal of Phonetics*, 8: 113-133.
- Gafos, A. (2002), A grammar of gestural coordination. *Natural Language and Linguistic Theory*.
- Gick, B. (1999), A gesture-based account of intrusive consonants in English. *Phonology* 16: 29-54.
- Gick, B. (in press), 'Articulatory correlates of ambisyllabicity in English glides and liquids'. in J. Local, R. Ogden and P. Temple (eds), *Papers in laboratory phonology VI: constraints on phonetic interpretation*. (Cambridge University Press, Cambridge).
- Hauser, M.D. (1996), *The Evolution of Communication*. (Cambridge, MA: MIT Press).
- Herbert, R. K. (1990), 'The relative markedness of click sounds: Evidence from language change, acquisition, and avoidance'. *Anthropological Linguistics*, 32: 120-138.
- Hockett, C.F. (1958), *A Course in Modern Linguistics*. New York: MacMillan.
- Humboldt, W. von. (1836/1999), *On Language*. Translated by Peter Heath and edited by Michael Losonsky. (Cambridge: Cambridge University Press).
- Hurford, J.R., Studdert-Kennedy, M. and Knight, C. (eds.) (1998), *Approaches to the Evolution of Language*. (Cambridge: Cambridge University Press).
- Ingram, D. (1992), 'Early phonological acquisition: A cross-linguistic perspective'. in C.A. Ferguson, et al., 423-435.
- Jacob, F. (1977), 'The linguistic model in biology', in D. Armstrong and C.H. van Schoonefeld (eds.), *Roman Jakobson: Echoes of his Scholarship*, (Lisse: de Ridder), 185-192.

- Jakobson, R. (1970), 'Linguistics', in *Main Trends of Research in the Social and Human Sciences*, Vol 1. (Paris/The Hague: UNESCO/Mouton), 437-440.
- Jakobson, R., Fant, C.G.M. and Halle, M. (1951/1963), *Preliminaries to Speech Analysis*. (Cambridge, MA: MIT Press).
- Keating, P. A. (1990), "The window model of coarticulation: Articulatory evidence", in J. Kingston and M. E. Beckman (eds.), *Papers in laboratory phonology I: Between the grammar and physics of speech*. (Cambridge: Cambridge University Press), 451-470.
- Kirby, S. (2000), 'Syntax without natural selection: How compositionality emerges from vocabulary in a population of learners', in C. Knight, et al., 303-323.
- Klima, E. & Bellugi, U. (1979), *Signs of Language*. (Cambridge: Harvard University Press).
- Knight, C., Studdert-Kennedy, M., and Hurford, J.R. (eds.) (2000), *The Evolutionary Emergence of Language*. (Cambridge: Cambridge University Press).
- Kohler, K. (1998), 'The development of sound systems in human language', in J.R. Hurford, et al., 265-278.
- Kugler, P. N., and Turvey, M. T. (1987), *Information, natural law, and the self-assembly of rhythmic movement*. (Hillsdale, NJ: Lawrence Erlbaum Associates).
- Liberman, A.M., Cooper, F.S., Shankweiler, D.P. and M. Studdert-Kennedy. (1967), 'Perception of the speech code'. *Psychological Review*, 74: 431-461.
- Lindblom, B. (1992), 'Phonological units as adaptive emergents of lexical development'. In C.A. Ferguson, et al., 131-163.
- Lindblom, B. (1998), 'Systemic constraints and adaptive change in the formation of sound structure'. In Hurford, et al. 242-264.
- Lindblom, B. (2000), 'Developmental origins of adult phonology: the interplay between phonetic emergents and the evolutionary adaptations of sound patterns'. *Phonetica*, 57: 297-314.
- Lindblom, B. and Maddieson, I. (1988), Phonetic universals in consonant systems, in L.M. Hyman & C.N. Li (eds.) *Language, Speech and Mind*. (London and New York: Routledge), 62-78.
- MacNeilage, P.F. (1998), 'The Frame/Content theory of evolution of speech production', *Behavioral and Brain Sciences*, 21: 499-511.

- Manuel, S. Y. (1990), 'The role of contrast in limiting vowel-to-vowel coarticulation in different languages', *Journal of the Acoustical Society of America*, 88: 1286-1298.
- McCarthy, J. J. (1988), 'Feature geometry and dependency: A review', *Phonetica*, 45: 84-108.
- Meltzoff, M. and Moore, K. (1997), 'Explaining facial imitation: A theoretical model'. *Early Development and Parenting*, 6: 179-192.
- Menn, L. (1983), 'Development of articulatory, phonetic and phonological capabilities'. in B. Butterworth, (ed.). *Language Production, II*. (London: Academic Press), 3-30.
- Ohala, J. (1983), 'The Origin of Sound Patterns in Vocal Tract Constraints'. in P. MacNeilage (ed.), *The Production of Speech*. (New York: Springer-Verlag), 189-216
- Olmsted, D.L. (1971), *Out of the Mouths of Babes: Earliest Stages in Language Learning*. (The Hague: Mouton).
- Pierrehumbert, J. B. (1990), 'Phonological and phonetic representation'. *Journal of Phonetics*, 18: 375-394.
- Pollack, R. (1994), *Signs of Life*. Boston: Houghton-Mifflin.
- Rizzolatti, G., Fadiga, L., Gallese, V. and Fogassi, L. (1996), 'Premotor cortex and the recognition of motor actions.' *Cognitive Brain Research*, 3, 131-141.
- Saltzman, E. L., and Munhall, K. G. (1989), A dynamical approach to gestural patterning in speech production. *Ecological Psychology*, 1: 333-382.
- Schrödinger, E. (1944), *What is Life?* (Cambridge: Cambridge University Press).
- Stevens, K. N. (1989), 'On the quantal nature of speech'. *Journal of Phonetics*, 17: 3-45.
- Stoel-Gammon, C. (1985, 'Phonetic inventories, 15-24 months: a longitudinal study. *Journal of Speech and Hearing Research*, 18, 505-512.
- Studdert-Kennedy, M. (1998), 'The particulate origins of language generativity: from syllable to gesture', in J.R. Hurford, et al., 202-221.
- Studdert-Kennedy, M. (2000), 'Evolutionary implications of the particulate principle: Imitation and the dissociation of phonetic form from semantic function', in C. Knight, et al., 161-176.
- Studdert-Kennedy, M. (2002), 'Mirror neurons, vocal imitation, and the evolution of particulate speech', in M. Stamenov and V. Gallese (eds.). *Mirror Neurons and the Evolution of the Brain and Language*. (Amsterdam: John Benjamins), 207-227.

- Studdert-Kennedy, M. & Goodell, E. W. (1995), 'Gestures, features and segments in early child speech'. in B. de Gelder, and J. Morais, (eds.). *Speech and Reading*. (Hove: Erlbaum (UK), Taylor & Francis) 65-88.
- Vihman, M. M., & Velleman, S. (1989), 'Phonological reorganization: A case study'. *Language and Speech*, 32: 149-170.
- Wray, A. (2000), 'Holistic utterances in protolanguage: the link from primates to humans', in C. Knight, et al., 285-302.
- Zsiga, E.C. (1995), 'An acoustic and electropalatographic study of lexical and postlexical palatalization in American English'. In B. Connell and A. Arvaniti (eds.), *Phonology and phonetic evidence; Papers in laboratory phonology IV*, (Cambridge: Cambridge University Press), 282-302.