1176

Commentary

Phonetica 2000;57:275–283

Received: November 2, 1999
Accepted: February 14, 2000

Imitation and the Emergence of Segments

Michael Studdert-Kennedy

Haskins Laboratories, New Haven, Conn., USA

Abstract

The paper argues that the discrete phonetic segments on which language is raised are subjective gestural structures that emerge ontogenetically (and perhaps emerged evolutionarily) from the process of imitating a quasi-continuous acoustic signal with a neuroanatomically segmented and somatotopically organized vocal machinery. Evidence cited for somatotopic organization includes the perceptual salience in the speech signal of information specifying place of articulation, as revealed both by sine wave speech and by the pattern of errors in children's early words.

'Almost every insight gained by modern linguistics, from Grimm's Law to Jakobson's distinctive features, depends crucially on the assumption that speech is a sequence of discrete entities.'

Morris Halle [1964, p. 325]

In his position paper for this volume Lindblom proposes an original and persuasive account of how 'discrete entities' (gestures, segments) may emerge from the continuous processes of speaking and listening. Supplementing two earlier papers [Lindblom, 1992, 1998], the account is unique in its attempt to solve the central issue of modern speech research without programmatic appeal either to the as yet undiscovered invariants of direct realism [Fowler, 1986] or to the specialized decoding devices concealed in a phonetic module [e.g. Liberman and Mattingly, 1985]. Since I fully approve of Lindblom's goals, agree whith his proscription of nativism, and admire the elegance of his arguments, the following comments are largely those of an *amicus curiae.*

The Intersubjectivity of Speech

The 'discrete entities' of Halle's observation are not simply a heuristic 'assumption'. Such entities are physical prerequisites of all systems that 'make infinite use of finite means' [von Humboldt, 1836/1972, p. 70]. Such systems (e.g. physics, chemistry, genetics, language) necessarily conform to the 'particulate principle of self-diversifying systems' [Abler, 1989; Studdert-Kennedy, 1998, 2000] by which discrete units from a finite set of meaningless elements (e.g. atoms, chemical bases, phonetic segments) are repeatedly sampled, permuted, and combined to yield larger units (e.g. mol-

KARGER

Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

0031–8388/00/0574–0275
$17.50/0
Accessible online at:
www.karger.com/journals/pho

Michael Studdert-Kennedy
Haskins Laboratories
270 Crown Street
New Haven, CT 06511-6695 (USA)

ecules, genes, words) higher in a hierarchy and both different and more diverse in structure and function than their constituents. Duality of patterning, the two-level hierarchy of phonology and syntax on which the unbounded semantic scope of language rests, is a special case of the particulate principle common to every physical system of unbounded diversity. Discrete units are therefore logically necessary postulates for the description of linguistic function.

What, then, is the empirical evidence for such units? Every phonetician is familiar with the fact that spectrograms do not divide the acoustic flow of speech into a sequence of discrete, invariant segments corresponding to the segments of linguistic description [Fant, 1962; Liberman et al., 1967]. Yet, perhaps because 'the sounds of the world's languages' are physical events amenable to increasingly sophisticated acoustic analysis, speech scientists have been reluctant to accept that '... there is no way to avoid the traditional assumption that the speaker-hearer's linguistic intuition is the ultimate standard that determines the accuracy of any proposed grammar, linguistic theory, or operational test...' [Chomsky, 1965, p. 21]. Many speech scientists have continued to hope that advances in speech technology or behavioral analysis may enable them to shed the introspective methods still burdening their colleagues in syntax.

There may indeed be some prospect of demonstrating the segmented structure of speech through analysis of phonetic segments into discrete gestures [Browman and Goldstein, 1992], as Lindblom [this vol.] argues (see below). But until such an analysis can be agreed upon, the only objective evidence for the discrete segments of speech are the discrete letters of the alphabet that represent a speaker-hearer's intuitions. That the fluent speech of any language can be transcribed as a string of letters and can then be recovered from that string by a reader is, in my view and despite the scholarly scepticism of some linguists [e.g. Faber, 1992], unequivocal evidence for the reality of the segment at some level of linguistic function. Spoonerisms and other segmental speech errors may lend insight into likely biophysical constraints on speaking [e.g. MacNeilage, 1985], but they add little to the alphabet by way of objective support for the segment, because [with the single exception of an electromyographic study by Mowrey and Mackay, 1990] they depend on alphabetic records of speaker-hearers' subjective judgments.

Speech segments, then, are subjective psychophysical entities, their characteristic perceptual properties (as described by distinctive feature theory, for example) analogous to loudness and pitch, or brightness and color. Yet segments are perhaps even more deeply subjective than such auditory and visual dimensions, insofar as their defining margins seem to have no reliable correlates in the acoustic signal. Acoustic discontinuities often occur, to be sure, with more or less abrupt changes in amplitude or spectrum. These discontinuities may serve as 'landmarks' around which important information concerning segments is distributed [Liu, 1996; Stevens, 1985]. But not every discontinuity marks a segment, nor is every segment marked by a discontinuity.

The nature of the difficulty is thrown into relief by attempts to transform the speech waveform into a conceptually and experimentally more tractable auditory representation. The spectrograph, although sometimes billed as an instrument for purely acoustic analysis, was, in fact, devised as a rough model of the human ear viewed as an adaptive biophysical mechanism for analyzing a waveform into its component frequencies. More sophisticated models, intended to capture nonlinearities of audition by weighting the frequency output of the analyzer according to the estimated frequency response curve of the cochlea [e.g. Evans, 1982] or the general auditory system [e.g. Schroeder et al., 1979], have clarified certain issues in speech research, and Lindblom has been a leader in their use [e.g. Liljencrants and Lindblom, 1972; Lindblom, 1986]. But whatever such models may contribute to the solution of the invariance problem, they have little to contribute to the problem of segmentation, because segments are no better marked in auditory transforms than in standard spectrograms.

In this regard, the success of Victor Zue in learning to read spectrograms is instructive [Cole et al., 1980]. As the reader will recall, Zue learned by assiduous practice for over 2,000 h to do what few, if any, others have done, namely, to transcribe the spectrograms of unknown (even nonsensical, and therefore syntactically and semantically unpredictable) utterances with remarkable accuracy. His central task was to see through the variability of differing phonetic contexts to the underlying sequence of discrete phonetic segments, *without top-down support* from syntax or meaning. That he succeeded demonstrates unequivocally that the informed eye (and, presumably, the informed ear) can indeed find reliable, even if context-conditioned, segmental markers in the fluent spectral sequence.

I use the epithet 'informed' because two pieces of information, absent from the spectrogram, were critical to Zue's success: (1) that the language was English, (2) that there were segments to be found. Even for the already segmented, but undeciphered, transcriptions of Minoan Linear B [Chadwick, 1958] and old Mayan [Coe, 1992], the identity of the language and the nature of the discrete entity encoded (word, syllable, segment) were postulates essential to their decoding. For the child learning to talk, the identity of its native language gradually unfolds in its characteristic structural regularities. But how does the child discover that fluent speech is composed of discrete, though intricately overlapping, segments?

I shall argue shortly that the child does so by discovering how it must engage its vocal machinery in order to speak, and that it speaks by repeatedly combining the discrete actions (gestures) of six functionally independent (even when mechanically interacting) articulators: lips, tongue blade (tip), tongue body, tongue root, velum, and larynx. In other words, the child discovers segments in its own body and behavior. It is the behavior implicit in an animal's morphology and environment at any given stage that drives development, just as it drives evolution [Mayr, 1982, p. 612; Studdert-Kennedy, 1991, p. 8].

We should not, however, identify a gesture too closely with an individual articulator, because at each instant of speech every articulator is engaged in action (or in inaction) complementary to the current primary articulator, and is thereby contributing to the overall vocal tract configuration [Mattingly, 1991]. The rise and fall of each configuration then specifies the domain of a given gesture [Carré and Chennoukh, 1995; Fowler and Smith, 1986]. Because it is the vocal tract configuration that structures the acoustic signal, gestures implicit in a configuration are accessible, in principle, to any listener, even one who cannot talk, such as a cerebral palsied child. Provided afferent and central cognitive processes are unimpaired and damage is confined to efferent processes, a child can evidently apprehend a gesture's contrastive phonological function [e.g. Cossu et al., 1987]. Thus, even a child who cannot talk may discover segments implicit in its own body and behavior, and is thereby launched into language.[1]

[1] Kluender et al. [1987] conditioned Japanese quail to discriminate /dv/ syllables from /bv, gv/ syllables, spoken by an adult male, and to generalize their responses across novel vowel contexts. The quail were not tested, however, for generalization across syllable position from CV to VC. We have no reason therefore to suppose that they discriminated anything more than one class of holistic syllables differing in onset from two others. Whatever bearing the study may have on the issue of invariance, it has little bearing on the emergence of segments.

Imitation and Segments

Vocal imitation, unique among primates to humans [Hauser, 1996, pp. 650–651], is the mechanism by which a child builds its lexicon. *Imitating a word* entails analysis of the sound pattern into its underlying articulatory components and reassembly of the components in their correct temporal sequence, *a process that introduces segmentation by transducing continuously variable sound into a pattern of discrete gestures.* Lindblom [this vol.], correctly in my view, grounds the segmental mechanism in the neuroanatomical segmentation of the vocal machinery and the likely somatotopic organization of its neural control. But he has little to say about how the perceptuomotor link between sound and gesture is established. How does sound get into the muscles? There have been two main classes of answer: (1) by conditioning, (2) by a specialized (not necessarily innate) mechanism. Let us briefly consider examples of each of these.

Exemplar Theory

In his exemplar model of speech perception, Johnson [1997, p. 147] defines an exemplar as '... an association between a set of auditory properties and a set of category labels... [T]he set of *category labels* includes any classification that may be important to the perceiver, and which was *available at the time the exemplar was stored* – for example, the linguistic value of the exemplar, the gender [sic] of the speaker, the name of the speaker, and so on' [my italics]. Presumably, the 'linguistic value' of an exemplar includes its component phonetic categories. The question then is whether the auditory properties of a spoken exemplar can be labeled phonetically without specifying its articulation.

I believe the answer to this question is 'no'. The name of a speech sound is not an arbitrary label like a letter of the alphabet or a word. We can certainly change the written label for /b/ to /d/, or the name ('category label') for a canine creature from *dog* to *cat*. But we cannot change the name of the sound we transcribe as /b/ to the name of the sound we transcribe as /d/, because the name of a speech sound is the spoken sound itself. If a listener is to store an exemplar speech sound as a phonetic segment, it would seem that the 'category label' specifying the 'linguistic value' associated with the auditory properties of the sound must include its articulatory value. In this way a speech sound exemplar becomes a segmentally structured phonetic event, simultaneously both auditory and articulatory.

Perhaps Johnson [1997, p. 153] himself believes this, since he goes on to explain how '... an exemplar model can, in principle, also be used to give an account of the production-perception link'. He proposes a conditioning, or associationist, account of the link as '... based on one's own speech' [p. 154]. Apparently, among the category labels the infant associates with the auditory properties of its own speech are the gestures it makes in pronouncing the sounds. (Here, it would seem, is the origin of the subjectivity of speech in exemplar theory.) The infant's own speech sounds are then 'ego exemplars', and '... the gestural knowledge derived or generated while listening to others is based on ego exemplars' [p. 154.].

If this is so, the child learning to talk derives articulatory instructions (gestures) for pronouncing new words by recognizing an auditory match between its own sound exemplars and the sounds that compose the adult word it intends to imitate. This may seem plausible, but there are several difficulties. First, given the lack of spectral overlap between the speech of an adult and the speech of, say, a 1-year-old child, how can the child match its own sounds with an adult's without some process of the normalization, or demodulation [Traunmüller, 1994], that exemplar theory rejects? Second, if gestural 'category labels' only become associated with the auditory properties of a speech sound exemplar when the child hears its own speech, how do new sounds enter the child's spoken repertoire? Only by random search, it would seem. Indeed, Lindblom [this vol.] writes of children who '... stumble on motorically motivated phenomena in the ambient language'. Random search within a motorically constrained articulatory space seems plausible enough for the hominid evolutionary path into speech. For the modern child, however, we have evidence that the search is not random, but rather is guided by an early developing somatotopically organized mode of perception. Before I consider this, let me turn briefly to an alternative approach to the perceptuomotor link.

Facial Imitation

How infants imitate facial gestures that they cannot see themselves perform is a question about which we have learned a great deal over the past two decades, largely through the sustained research program of Meltzoff and Moore [for a fairly recent review, see Meltzoff and Moore, 1997]. I have no space to consider this work in detail, but several aspects of their findings and current model are of interest for an understanding of vocal imitation. In particular, Meltzoff and Moore propose a supramodal representation system mediating between perception and action, and an intermodal mechanism (as opposed to a conditioned association) for generating an imitative response. Among the theoretical concepts they invoke are: (1) organ identification, a mechanism for identifying the body part(s) to be moved; (2) body babbling, a process analogous to vocal babbling, that maps muscle movements onto 'organ-relation end states', analogous to vocal tract configurations; (3) a cross-modal metric of equivalence based on 'organ relations [that] render commensurate the seen but unfelt act of the adult and the felt but unseen facial act of the infant'.

How much of the system is innate, and how much develops epigenetically in response to the social environment, is as much a question for facial imitation as for speech. But at the heart of the facial system is the infant's (and the adult's) capacity to recognize correspondences in organ relations between self and conspecific other. This would seem necessarily to depend on some demodulating mechanism that renders self and conspecific other structurally and functionally isomorphic. Among the many possibilities the facial work raises for speech, then, is that the cross-modal (auditory-to-articulatory) metric of equivalence for speech may be mediated by vocal tract configuration rather than by normalization of the auditory signal. Perhaps this could be accomplished by auditory-to-body-part neural links analogous to the visual 'mirror neurons' discovered by Rizzolatti et al. [1996] at the University of Parma.

Mirror Neurons

Recently, Rizzolatti et al. [1996] have reported what they call 'mirror neurons' in macaque cortex: neurons that fire not only when a monkey grasps or manipulates food, but also when it sees a human experimenter do the same. Firing is specific to the act of grasping, and does not occur when the monkey sees an experimenter pick up food with a tool. These perceptuomotor neurons lie in an area of macaque cortex arguably homologous with Broca's area. Data from transcranial magnetic stimulation and positron emission tomography studies demonstrate a mirror system for manual grasping also in humans. Rizzolatti and Arbib [1998, p. 190], reviewing this evidence, postulate 'a fundamental mechanism for action recognition' in both monkey and human.

Obviously, mirror neurons must be part of a complex network engaged both in acting and in monitoring the acts of others – not only of conspecifics, it would seem, but of other animals, such as humans, with similar gross morphology. But these neurons offer no solution to the problem of 'how light gets into the muscles', because they give no hint of how the perceptuomotor link is made. Nor, I should emphasize, is there yet evidence for such neurons in the speech system. Nonetheless, work on mirror neurons takes another experimental step toward understanding the neural basis for the social empathy characteristic of primate species [cf. Brothers et al., 1990]. And these neurons are of particular interest for vocal imitation because they seem to be organized not only by function – grasping, manipulating, eating, and so on –, but also somatotopically. With this in mind, let us turn to children's early words.

Early Words

Elsewhere I have sketched a development sequence for the origin of segments [Studdert-Kennedy, 1987, 1991; Studdert-Kennedy and Goodell, 1995] that draws its exemplar-type model from the account proposed by Lindblom et al. [1984]. On this account, the initial unit of linguistic action is the holistic word [Ferguson and Farwell, 1975]. The word is said to be holistic, even though it is spoken as a sequence of discrete gestures, because gestures are not yet represented as independent phonetic elements that can be marshaled for use in an unbounded set of other contexts. As an automatic consequence of sorting and stacking phonetically similar words, independent gestures eventually emerge, and recurrent patterns of co-occurring gestures are then gradually integrated into segments. Lindblom [e.g. 1992, 1998] has characterized such an emergent process far more elegantly and concisely than I can, both in this volume with his NEP model and in other papers.

Evidence for the gesture as an independent unit of function in young children is hard to come by, partly because children tend to avoid words that they cannot pronounce [Vihman, 1991; Vihman and DePaolis, 2000] and so are surprisingly accurate in their pronunciation, partly because the period during which children make nonrandom and interpretable speech errors tends be a narrow window around the end of the 1st year when they are attempting their first words. Nonetheless, systematic analysis of such errors strongly supports the gesture as the child's initial intrasyllabic unit of phonetic action [Studdert-Kennedy and Goodell, 1995].

A remarkable fact about early consonantal errors is that they tend to be errors of gestural timing or amplitude rather than of place of articulation. This is remarkable because, according to standard speech lore, place of articulation is significantly more susceptible to degradation by noise and filtering than are manner and voicing [Miller and Nicely, 1955]. Yet, if we look at the word-initial phone classes (the set of interchangeable segments with which a child attempts a given word) for the children of Ferguson and Farwell [1975], labials tend to be exchanged with labials, alveolars with alveolars, velars with velars. Errors of place seldom occur. Table 1 makes the point with data I have tabulated from 4 children in the Stanford Child Phonology Project [Vihman, 1996, Appendix C]: 80% of single-feature errors are on voicing or manner, 20% on place of articulation.

Table 1. Initial consonants in early words of 4 English-learning children in three half-hour sessions during transition fom babbling to word use (13–16 months approximately): tabulation of data from Stanford Child Phonology Project in Appendix C of Vihman [1996]

Target	Number correct	Number of single feature errors			Number attempted
		place	voicing	other	
Bilabial [p, b, m, w]	88	3	46	3	140
Alveolar [t, d, n, r, l, s, z]	46	14	25	1	86
Velar [k, g]	27	4	9	0	40
Total	161	21	80	4	266
% of errors		20	76	4	

Taking a hint from fact and theory in facial imitation and from the macaque's mirror neurons we may speculate that the acoustic speech signal, like the optic face, specifies the 'organs' (articulators) that are to be activated more clearly than the amplitude and relative phasing of their activation. Such a hypothesis is consistent with the surprising intelligibility of both fricative speech [Kuhn, 1975] and sine wave speech [Remez et al., 1994]. The sound source for fricative speech is uniform frication exciting the oral front cavity resonance; in sine wave speech all the acoustic elements characteristic of vocal sound production are replaced by a set of time-varying sinusoids that track the changing resonances, and so the changing configurations, of the vocal tract. Perhaps it is because these bizarre forms of speech preserve information about which articulators were engaged (that is, about place of articulation) that they are so readily intelligible.

Conclusion

Exemplar-based learning models are attractive because, as Lindblom [this vol.] remarks, they undertake to solve the problems of speech invariance and segmentation by statistical accumulation rather than by ad hoc hypothetical decoding mechanisms. In this respect exemplar models formalize a style of approach that Lindblom has been following for many years. Yet there is little reason to suppose that segments automatically emerge from the statistical stacking of auditory exemplars, none of which is segmented at the time of storage. What seems to be missing, then, from Lindblom's emergent phonology is an explicit mechanism by which auditory patterns make contact with the neuroanatomically segmented vocal machinery that produces them. Perhaps studies of social interaction in other primates, such as macaques, and of imitation in other modalities, such as facial imitation, will provide the key.

Acknowledgments

My thanks to René Carré, Randy Diehl, John Kingston, and Robert Remez for instructive comments and discussion. Preparation of the paper was supported in part by Haskins Laboratories.

References

Abler, W.L.: On the particulate principle of self-diversifying systems. J. soc. Biol. Struct. *12:* 1–13 (1989).

Brothers, L.; Ring, B.; Kling, A.: Response of neurons in the macaque amygdala to complex social stimuli. Behav. Brain Res. *41:* 199–213 (1990).

Browman, C.P.; Goldstein, L.: Articulatory phonology: an overview. Phonetica *49:* 155–180 (1992).

Carré, R.; Chennoukh, S.: Vowel-consonant-vowel modeling by superposition of consonant closure on vowel-to-vowel gestures. J. Phonet. *23:* 231–241 (1995).

Chadwick, J.: The decipherment of Linear B (Cambridge University Press, Cambridge 1958).

Chomsky, N.: Aspects of the theory of syntax (MIT Press, Cambridge 1965).

Coe, M.: Breaking the Maya code (Thames and Hudson, New York 1992).

Cole, R.A.; Rudnicky, A.I.; Zue, V.M.; Reddy, D.R.: Speech as patterns on paper; in Cole, Perception and production of fluent speech, pp. 3–50 (Erlbaum, Hillsdale 1980).

Cossu, G.; Urbinati, M.L.; Marshall, J.C.: Reading without speech and writing without arm movements. 5th Eur. Workshop on Cognitive Neuropsychol., Bressanone 1987.

Evans, E.F.: Representation of complex sounds at cochlear nerve and cochlear nucleus levels; in Carlson, Granström, The representation of speech in the peripheral auditory system, pp. 27–42 (Elsevier, Amsterdam 1982).

Faber, A.: Phonemic segmentation as epiphenomenon: evidence from the history of alphabetic writing; in Downing, Lima, Noonan, The linguistics of literacy, pp. 111–134 (J. Benjamins, Philadelphia 1992).

Fant, C.G.M.: Descriptive analysis of the acoustic aspects of speech. Logos *5:* 3–17 (1962).

Ferguson, C.A.; Farwell, C.B.: Words and Sounds in early language acquisition. Language *15:* 419–439 (1975).

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

Fowler, C.A.; Smith, M.R.: Speech perception as 'vector analysis': an approach to the problems of invariance and segmentation; in Perkell, Klatt, Invariance and variability in speech processes, pp. 123–139 (Erlbaum, Hillsdale 1986).

Halle, M.: On the bases of phonology; In Fodor, Katz, The structure of language, pp. 324–333 (Prentice-Hall, Englewood Cliffs 1964).

Hauser, M.D.: The evolution of communication (MIT Press, Cambridge 1996).

Humboldt, W. von: Linguistic variability and intellectual development; translated by G.C. Buck and F.A. Raven (University of Pennsylvania Press, Philadelphia 1836/1972).

Johnson, K.: Speech perception without speaker normalization: an exemplar model; in Johnson, Mullenix, Talker variability in speech processing, pp. 145–165 (Academic Press, New York 1997).

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

Kuhn, G.M.: On the front cavity resonance and its possible role in speech perception. J. acoust. Soc. Am. *58:* 428–433 (1975).

Liberman, A.M.; Cooper, F.S.; Shankweiler, D.P.; Studdert-Kennedy, M.: Perception of the speech code. Psychol. Rev. *74:* 431–461 (1967).

Liberman, A.M.; Mattingly, I.G.: The motor theory of speech perception revised. Cognition *21:* 1–36 (1985).

Liljencrants, J.; Lindblom, B.: Numerical simulation of vowel quality systems: the role of perceptual contrast. Language *48:* 829–862 (1972).

Lindblom, B.: Phonetic universals in vowel systems; in Ohala, Jaeger, Experimental phonology, pp. 13–44 (Academic Press, New York 1986).

Lindblom, B.: Phonological units as adaptive emergents of lexical development; in Ferguson, Menn, Stoel-Gammon, Phonological development: models, research, implications, pp. 131–163 (York Press, Timonium 1992).

Lindblom, B.: Systemic constraints and adaptive change in the formation of sound structure; in Hurford, Studdert-Kennedy, Knight, Approaches to the evolution of language, pp. 242–264 (Cambridge University Press, Cambridge 1998).

Lindblom, B.: MacNeilage, P.I.; Studdert-Kennedy, M.: Self-organizing processes and the explanation of phonological universals; in Butterworth, Comrie, Dahl, Explanations for language universals, pp. 181–203 (Mouton, New York 1984).

Liu, S.A.: Landmark detection for distinctive feature-based recognition. J. acoust. Soc. Am. *100:* 3417–3430 (1996).

MacNeilage, P.F.: Serial-ordering errors in speech and typing; in Fromkin, Phonetic linguistics, pp. 193–201 (Academic Press, New York 1985).

Mattingly, I.G.: The global character of phonetic gestures. J. Phonet. *18:* 445–452 (1991).

Mayr, E.: The growth of biological thought (Belknap Press of Harward University, Cambridge 1982).

Meltzoff, A.N.; Moore, M.K.: Explaining facial imitation: a theoretical model. Early Dev. Parenting *6:* 179–192 (1997).

Miller, G.A.; Nicely, P.E.: An analysis of perceptual confusions among some English consonants. J. acoust. Soc. Am. *27:* 338–352 (1955).

Mowrey, R.A.; Mackay, I.R.A.: Phonological primitives: electromyographic speech error evidence. J. acoust. Soc. Am. *88:* 1299–1312 (1990).

Remez, R.E.; Rubin, P.E.; Berns, S.M.; Pardo, J.S.; Lang, J.R.: On the perceptual organization of speech. Psychol. Rev. *101:* 129–156 (1994).
Rizzolatti, G.; Arbib, M.A.: Language within our grasp. Trends Neurosci. *21:* 188–194 (1998).
Rizzolatti, G.; Fadiga, L.; Gallese, V.; Fogassi, L.: Premotor cortex and the recognition of motor actions. Cognitive Brain Res. *3:* 131–141 (1996).
Schroeder, M.R.; Atal, B.S.; Hall, J.L.: Objective measures of certain speech signal degradations based on masking properties of human auditory perception; in Lindblom, Öhman, Frontiers of speech communication research, pp. 217–229 (Academic Press, London 1979).
Stevens, K.N.: Evidence for the role of acoustic boundaries in the perception of speech sounds; in Fromkin, Phonetic linguistics, pp. 243–255 (Academic Press, New York 1985).
Studdert-Kennedy, M.: The phoneme as a perceptuomotor structure; in Allport, MacKay, Prinz, Scheerer, Language perception and production, pp. 67–84 (Academic Press, London 1987).
Studdert-Kennedy, M.: Language development from an evolutionary perspective; in Krasnegor, Rumbaugh, Schiefelbusch, Studdert-Kennedy, Biological and behavioral determinants of language development, pp. 5–28 (Erlbaum, Hillsdale 1991).
Studdert-Kennedy, M.: The particulate origins of language generativity: from syllable to gesture; in Hurford, Studdert-Kennedy, Knight, Approaches to the evolution of language, pp. 202–221 (Cambridge University Press, Cambridge 1998).
Studdert-Kennedy, M.: Evolutionary implications of the particulate principle: imitation and the dissociation of phonetic form from semantic function; in Knight, Studdert-Kennedy, Hurford, The emergence of language: social function and the origins of linguistic form, pp. 161–176 (Cambridge University Press, Cambridge 2000).
Studdert-Kennedy, M.; Goodell, E.W.: Gestures, features and segments in early child speech; in deGelder, Morais, Speech and reading, pp. 65–88 (Erlbaum, Taylor & Francis, Hove 1995).
Traunmüller, H.J.: Conventional, biological and environmental factors in speech communication: a modulation theory. Phonetica *51:* 170–183 (1994).
Vihman, M.M.: Ontogeny of phonetic gestures; in Mattingly, Studdert-Kennedy, Modularity and the motor theory of speech perception, pp. 9–84 (Erlbaum, Hillsdale 1991).
Vihman, M.M.: Phonological development (Blackwell, Oxford 1996).
Vihman, M.M.; DePaolis, R.A.: The role of mimesis in infant language development: evidence for phylogeny?; in Knight, Studdert-Kennedy, Hurford, The emergence of language: social function and the origins of linguistic form (Cambridge University Press, Cambridge 2000).