TOWARD A NEURODEVELOPMENTAL MODEL OF PHONETICS

R. Kent
University of Wisconsin-Madison, Madison, Wisconsin

The papers presented at this Congress give strong witness to the diversification and expansion of the phonetic sciences. These are signs that the phonetic sciences are thriving and growing. But diversification and expansion also come with a certain risk to the unity of the phonetic sciences. Unity is important not only for scholarly collegiality, but also for the larger and deeper insights that a wide embrace of knowledge permits.

Where do the phonetic sciences stand as we approach the close of both a century and a millenium? My comments are necessarily selective as the field is too large to be satisfactorily summarized in a few pages. The areas of research to be addressed here are: cross-language comparisons, speech disorders, speech development, and the relation between production and perception. The objective is to show the interrelationship among these areas in a general theory.

CROSS-LANGUAGE STUDIES

One powerful leverage in phonetic research is the cross-language investigation, which has been used quite profitably to study proclivities in production and perception. One objective in these studies is to determine universal patterns or tendencies. Another is the study of two or more languages that are selected according to a specific criterion, such as presence or absence of a phonetic feature. The compilation of data from a large number of languages is perhaps the phonetic sciences' equivalent to "big science," such as the human genome study in molecular biology. In both cross-language research and genome research, one important long-range goal is a characterization of essential human traits. Large-scale cross-language studies of the kind that explore universal patterns have been a patchwork of smaller studies conducted by individual investigators or research teams working in different nations. There is nothing wrong with this enterprise, but it is likely that international coordination of some kind would facilitate the effort. There are, after all, several thousand languages to be studied. Successful models of cross-language research include the UPSID database [1] and a recent report on Long-Term-Average Speech Spectra (LTASS) for 17 languages [2]. Quantitative studies such as the latter could address several acoustic and physiologic aspects of speech to define the universal parameters of phonation and articulation. The phonetic properties of individual languages could be described within these universal boundaries.

In the aggregate, studies of speakers of different languages reveal universal patterns in the phonetic structure of speech. Some recent studies give shape to a tentative conclusion, namely that acoustic properties of speech show some evidence of universal organization, but the congruence among languages is not sufficient to suggest a universal template of acoustic-phonetic patterns. One form of acoustic data available for several languages is the formant-frequency pattern for vowels. Krull and Lindblom [3] reported that vowels labeled with the same IPA symbol are to some degree tuned to individual languages. If so, then the IPA symbols are only generally indicative of acoustic similarity among sounds from different languages. Tuning for individual languages is not necessarily evidence against the basic principles of Stevens' [4] quantal theory of speech production, which holds that nonlinearities in the articulatory-acoustic relation determine preferences in phonetic selection. But it does argue against the hardest version of such a theory in which quantal relations in themselves would determine universal selections. It appears that the formant frequencies for individual vowels in a language are adjusted to reflect the structure of the vowel system for that language. Additional evidence for a language-tuning effect was reported by Sussman et al. [5] in their derivation of locus equations for bilabial, alveolar, and velar stops in five different languages. Although the locus equations were generally similar across languages, the differences were large enough to discourage a conclusion of universal determinism.

A unifying hypothesis for further research is that different languages divide the acoustic space for vowels into similar, but not identical, phonetic regions. Languages leave a subtle imprint on the otherwise universal articulatory-acoustic relations identified in quantal theory and similar approaches [4, 6]. This idea is consistent with evidence showing the persistence of foreign accent on vowels [7] and with accounts of the influence of early language exposure on vowel perception in infants [8]. Early language exposure appears to be a powerful determinant of phonetic organization.

DISORDERS AND DEVELOPMENT

One test of a theory's power is its capacity to address problems in several interrelated domains. Phonetic theories generally have been formulated to account for the abilities of competent adult speakers. But these theories may also be called upon to account for speech errors and for the development of speech. The following remarks focus on these two areas which often have been marginalized in phonetic theory but are assuming a more central role in the contemporary phonetic sciences.

Errors in Speech Production

The errors of speech consist of two general kinds of (possibly related) errors: (1) the lapses made by normal talkers or listeners, and (2) the clinical errors made by individuals with disorders of speech or hearing. Phonetics has been largely concerned with normal error-free performance, but the investigation of errors has been a productive pursuit. Error-free adult speech is difficult to analyze into its formative units partly because of its relatively seamless nature. But the errors of speech help to make the seams transparent. These "accidents of nature" can be exploited to gain a glimpse into the systems that underlie speech production and perception.

Sequencing errors have been studied for insights into the "slippage" points of speech production. Analyses of these errors help in understanding how segments, syllables and other units are organized, if units such as these are useful at all. But there is a worrisome possibility that the corpora of speech errors collected over the years contain a major flaw that could lead to the collapse of the entire literature. The flaw was identified by Mowrey and MacKay [9], who identified three potentially serious limitations in the traditional perceptual classification of speech errors. First, listener normalization, an integral aspect of the perceptual process, may override the detection of subtle errors. Second, errors at a fine phonetic level often cannot be detected reliably. Third, transcription techniques are not available to code highly or subtly anomalous sounds. The remedy is to use a suitable technology and a suitable phonetic representation that can register fine variations in speech motor behavior.

Similar objections have been raised against some traditional perceptual analyses of neurogenic communication disorders such as verbal apraxia, dysarthria, and aphasia. Verbal apraxia is an object lesson in the misuse of perceptual techniques of speech analysis. The early influential descriptions of this disorder emphasized the predominance of phonemic substitution errors. These errors were interpreted to mean that the disorder was one of phonology or of phonemic selection. But subsequent studies, carried out with more refined acoustic, physiologic, and perceptual methods, have shown that the earlier analyses missed important errors at the phonetic or motoric levels [10,11,12].

The closer examination of errors in normal speech and neurogenic speech disorders points to a conceptualization in terms of gestures. These component movements are vulnerable to errors in timing and coordination. Some of these errors may give rise to apparent phonemic errors such as substitution or addition. But more subtle problems affecting individual movements and their combination may be the best focus for understanding the full range of errors in speech. Gestures offer an economical description of the multi-articulate nature of speech and can account for a variety of errors. The next phase in the study of neurogenic speech disorders may reflect a shift from reliance on global phonetic descriptors (such as broad transcription) to movement-based accounts (such as derivations of gestural scores).

Development of Speech

"There is overwhelming evidence that the emergence of coordinated movements is intimately tied both to the growth of the musculoskeletal system and to the development of the brain" [13, p. 966].

The foregoing quotation summarizes a developmental perspective on skilled movement. Speech is one form of skilled movement. As such, a central task in the understanding of speech production is to show how the coordinated movements of speech relate to the growth of the physiological system of speech and to the development of the neural systems that regulate speech movements. Seen in this way, the understanding of speech development is key to speech production and perception as a human faculty. Knowing how speech develops may be the fundamental discovery that unifies the various facets of phonetic study. The child who is developing speech is faced with two problems: (a) learning to perceive the phonetic code of the languages, and (b) learning to produce the motor patterns of the language in accord with the perceptual code.

Action theory has been highly influential in recent formulations of motor control in both developing and mature organisms. Action theory applied to speech has in the main addressed the adult speaker [14]. Action theory has found widespread application to virtually every kind of motion performed by muscular systems, including locomotion, reaching, and prehension. Action theory has emphasized particularly (a) the use of coordinative structures to solve the degrees of freedom problem common to multi-articulate systems, and (b) the biomechanical task specificity of motor responses. Action theory succeeds quite well in these two domains but it has had much less to say about the neural systems that regulate behavior.

A neurodevelopmental alternative to action theory is Edelman's theory of neuronal group selection [15,16,17]. Briefly, this theory states that repertoires of interconnected neuronal groups are established developmentally. Synaptic strengths are modified as the result of experiences including learning. Initially, large repertoires of variant neural circuits are established by selectional mechanisms in the developing embyro. These are called primary repertoires. Continuing selectional events serve to enhance neuronal responses that have adaptive value for the organism. These are called secondary repertoires. An important concept in Edelman's theory is "re-entrant signalling," an exchange of signals along parallel and reciprocal connections among neuronal groups. The theory of neuronal group selection has been used to account for perception, motor responses, language, and consciousness. This theory places an emphasis on neural processes that is either lacking or vaguely described in accounts of action theory. The selectionist account offers a promising framework for a neural theory of phonetics, some aspects of which will be addressed in the balance of this paper.

PRODUCTION AND PERCEPTION

One of the most recalcitrant problems in the phonetic sciences is the unification of production and perception. But there is little agreement on how unity can be acheived. One indication of the sluggish progress is the nearly separate development of theories of speech perception and speech production. There are some notable attempts to bring perception and production under the same theoretical umbrella, especially the motor theory of speech perception [18,19]. Another attempt at unification is the theory of event perception applied to speech [20], which found theoretical resonance with action theory [14].

Ultimately, the integration of production and perception should be evident in the same neural mechanisms that explain phonetic development. One possibility for sensory-motor interaction is the proximity of motor and sensory neurons in the CNS representation of various parts of the body. Huang et al. [21] found that the auditory area of Crus II of the cerebellar hemisphere in both rat and cat is surrounded by orofacial somatosensory receptive fields. Furthermore, the cerebellar granule cells in the posterior vermis and the hemispheres exhibited phasic responses to auditory stimuli. This could mean that the cerebellum is involved in event timing, a possibility supported by clinical studies showing that a primary feature of the speech of persons with cerebellar disease is disordered timing. The speech disorder, known as ataxic dysarthria, is commonly described as having altered patterns of syllable timing or stress [22,23]. Moreover, an hypothesis of cerebellar regulation of event timing has been advanced by Keele and Ivry [24] who proposed that a major cerebellar function is to provide temporal computations that underlie a variety of perceptual and motor tasks. One role of the cerebellum in speech may be as a neural timekeeper for the registration of sensory and motor information.

Some cortical neurons also respond to more than one type of stimulation. Bruce, Desimone, and Gross [25] reported that the majority of neurons in the superior temporal sulcus of macaques were sensitive to more than one modality. Only 41% were exclusively visual. The remainder responded to visual-auditory, visual-somesthetic, or visual-auditory-somesthetic stimulation. Neurons that respond to more than one modality may be the means for a convergence of bisensory or multisensory peripheral stimulation. Separate sensory channels can dissociate stimuli by modality, but multisensory convergence in the CNS allows integration across modalities [26]. This convergence is highly relevant to speech, which has a plurimodal sensory foundation of audition, taction, kinesthesia, and baroreception [27].

Sensory-motor integration also could

be based on neural representations formed in a complex neural circuitry. Song birds are a good animal model for speech for several reasons, especially in that many avian species learn their songs through exposure and practice. Song learning apparently involves a comparison of the young bird's calls with those of the adult conspecific birds. Williams [28] described the neural circuitry underlying birdsong development as having multiple loops, multiple modalities, multiple representations of song, and multiple neural mechanisms. She identified two primary branches of the control system. The first branch descended from the control center to the motoneurons for respiration and vocalization. The nuclei of this branch continue to grow during song development and seem to be indispensable for adult song. The nuclei of this branch have long and variable latencies. The second branch forms a recursive loop between two nuclei of the descending branch. It contains one nucleus that reaches its maximum size in early song learning and then declines. This branch has short latencies with little variability. Auditory activity can be recorded in both branches, including the motoneurones controlling the vocal organ. Williams proposed that the short-latency recurrent branch allows for comparisons of different song representations. If these results can be extended to speech development, the implications include the following.

(1) Different parts of the neural circuitry reach their maximum sizes at different ages. This maturational variation could be related to the concept of a "sensitive period" for speech development [29,30]. Potential for language learning may be related to the maturational gradients of the complex neural circuitry.

(2) The neural circuitry that supports speech sound learning may contain auditory, motor, tactile, and kinesthetic representations. The existence of these different representations may explain the robustness of speech production in the face of various attempts to reduce or modify sensory feedback [27,31].

(3) Recurrent branches provide a means for the comparison of these different representations of speech, although the use of these comparative mechanisms varies with stage of development and task demands. Sensory disruption is more damaging to the speech of children than to the speech of adults [27], which may reflect the adults' capability to switch easily and effectively between various representations. Speech development may involve the acquisition of different neural representations and the facility to select among them as task conditions change.

This perspective is consistent with Edelman's [15,16,17] theory of neuronal group selection, and particularly with his idea of re-entrant signaling. Speech may be represented as a number of neuronal "maps" that combine different kinds of sensory and motor information. In its global nature, speech is defined by the totality of these maps and their interactions. More narrowly, speech can be defined by interactions among selected maps. Therefore, speech is auditory-visual (as the McGurk effect [32] demonstrates), and tactuo-motor (as in the haptic communication system employed by users of Tadoma, who can understand speech from tactile cues gathered from a hand placed over the talker's face and neck [33]). This idea also accords with models of auditory processing that emphasize the temporal properties of the global neural response [34]. A primary advantage of re-entrant signaling is that it provides a temporal coherence of related sensory and motor information. This coherence is a fundamental neural correlate of phonetic events. It also offers a useful interpretation of speech rhythm as the basic temporal plan for the coordination of sensory and motor information in speech production and perception.

POINTS OF CONVERGENCE

Principles that apply to different domains in the phonetic sciences have good potential for the delineation of universal tendencies. One important convergence is that the sounds prominent in babbling also are prominent in the world's languages [30,35]. Furthermore, these sounds also tend to be preserved in neurogenic speech disorders [36,37]. These universal patterns can be explained by a combination of factors including: the developmental anatomy of the vocal tract, nonlinearities in the articulatory-acoustic relation, evolutionary influences expressed as primary movement repertoires, and perceptual contrast.

Another property of speech that is evident from several vantage points is what Fujimura [38] referred to as the "inherently multidimensional principle of speech coordination" (p. 218). This principle is a kernel concept in theories of speech production [14] but is also recognized in various nonlinear phonologies [38] and articulatory phonology [39]. This property of the motor speech system appears to offer valuable insights into speech development [40] and speech disorders [41], and may help to resolve some the problems that have been described with corpora of segment errors in normal and disordered speech [9,12]. Multidimensionality can be expressed as the regulation of component gestures. The concept of gesture is becoming increasingly central in discussions of speech motor control, phonology, speech development in children, and speech disorders. Smith [31] noted, "The convergence of theories onto the notion of a basic articulatory gesture suggests that it is an idea with intrinsic merit" (p. 261). But she added: "Despite the general appeal of the idea, a disproportionate amount of theory development has occurred in relation to the data available that clearly support the notion" (p. 261). This lack of proportion between theory and data has been a handicap, but the new technologies and the increasing availability of databases [38] may help to strike the balance in the next millenium.

ACKNOWLEDGMENTS

This research was supported by NIH research grants DC00319 and DC0082 from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health.

REFERENCES

[1] Maddieson, I. (1984), Patterns of sound. Cambridge, England: Cambridge University Press.

[2] Byrne, D., Dillon, H., Tran, K., Arlinger, S., et al. (1994). An international comparison of long-term average speech spectra. Journal of the Acoustical Society of America, vol. 96, pp. 2108-2120.

[3] Krull, D., & Lindblom, B. (1992). Comparing vowel formant data cross-linguistically. PERILUS (Phonetic Experimental Research, Institute of Linguistics, University of Stockholm), No. XV, pp. 7-15.

[4] Stevens, K. (1989). On the quantal nature of speech. Journal of Phonetics, vol. 17, pp. 3-45.

[5] Sussman, H. M., Hoemeke, K. A., & Ahmed, F. S. (1993). A cross-linguistic investigation of locus equations as a phonetic descriptor for place of articulation. Journal of the Acoustical Society of America, vol. 94, pp. 1256-1268.

[6] Badin, P., Perrier, P., Boe, L.-J., & Abry, C. (1990). Vocalic nomograms: Acoustic and articulatory considerations

upon formant convergences. Journal of the Acoustical Society of America, vol. 87, pp. 1290-1300.

[7] Munro, M.J., Flege, J.E., & MacKay, I.A.R. (in press). The effects of second language learning on the production of English vowels. Applied Psycholinguistics.

[8] Kuhl, P.K., Williams, K.A., Lacerda, F., Stevens, K.N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. Science, vol. 255, pp. 606-608.

[9] Mowrey, R.A., & MacKay, I.R.A. (1990). Phonological primitives: Electomyographic speech error evidence. Journal of the Acoustical Society of America, vol. 88, pp. 1299-1312.

[10] Itoh, M. Sasanuma, S., Tatsumi, I.F., Murakimi, S., Fukusako, Y., & Suzuki, T. (1982). Voice onset time characteristics in apraxia of speech. Brain and Language, vol. 17, 193-210.

[11] Ziegler, W. (1987). Phonetic realization of phonological contrasts in aphasic patients. In. J.H. Ryalls (Ed.), Phonetic approaches to speech production in aphasia and related disorders (pp. 163-179). Boston: College-Hill.

[12] McNeil, M.R., & Kent, R.D. (1990). Motoric characteristics of aphasia and apraxic speech. In G.R. Hammond (Ed.), Advances in psychology: Cerebral control of speech and limb movements (pp. 349-387). Amsterdam: North Holland.

[13] Sporns, O., & Edelman, G.M. (1993). Solving Bernstein's problem: A proposal for the development of coordinated movement by selection. Child Development, vol. 64, pp. 960-981.

[14] Kelso, J.A.S., Saltzman, E.L., & Tuller, B. (1986). The dynamical persective on speech production: Theory and data. Journal of Phonetics, vol. 14, pp. 29-59.

[15] Edelman, G. (1993). Neural Darwinism: Selection and Reentrant signaling in higher brain function. Neuron, vol. 10, pp. 115-125.

[16] Montague, P.R., Gally, J.A., & Edelman, G.R. (1991). Spatial signaling in the development and function of neural connections. Cerebral Cortex, vol. 1, pp. 199-220.

[17] Friston, K.J., Tononi, G., Reeke, G.N., Jr., Sporns, O., & Edelman, G.M. (1994). Value-dependent selection in the brain: Simulation in a synthetic neural model. Neuroscience, vol. 59, 229-243.

[18] Liberman, A.M., Cooper, F.S., Shankweiler, D., & Studdert-Kennedy, M. (1967). Perception of the speech code. Psychological Review, vol. 74, pp. 4311-4361.

[19] Liberman, A.M., & Mattingly, I.G. (1985). The motor theory of speech perception revised. Cognition, vol. 21, pp. 1-36.

[20] Fowler, C.A. (1986). An event approach to the study of speech perception from a direct-realist perspective. Journal of Phonetics, vol. 14, pp. 3-28.

[21] Huang, C., Hsiao, C.F., Yang, B., & Mu, H. (1991). Auditory receptive area in the cerebellar hemisphere is surrounded by somatosensory areas. Brain Research, vol. 541, pp. 251-256.

[22] Darley, F.L., Aronson, A.E., & Brown, J.R. (1969). Cluster of deviant speech dimensions in the dysarthrias. Journal of Speech and Hearing Research, vol. 12, pp. 462-496.

[23] Kent, R.D., Netsell, R. and Abbs, J. (1979). Acoustic characteristics of dysarthria associated with cerebellar disease. Journal of Speech and Hearing Research, vol. 22, pp. 627-648.

[24] Keele, S.W., and Ivry, R. (1990). Does the cerebellum provide a common computation for diverse tasks? In A. Diamond (Ed.), The development and neural bases of higher cognitive functions. Annals of the New York Academy of Sciences, vol. 608, pp. 179-211.

[25] Bruce, C., Desimone, R., & Gross, C.G. (1981). Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. Journal of Neurophysiology, vol. 46, pp. 369-384.

[26] Stein, B.R., & Meredith, M.A. (1990). Multisensory integration: Neural and behavioral solutions for dealing with stimuli from different sensory modalities. In A. Diamond (Ed.), The development and neural bases of higher cognitive functions, Annals of the New York Academy of Sciences, vol. 608, pp. 51-70.

[27] Kent, R.D., Martin, R.E., & Sufit, R.L. (1990). Oral sensation: A review and clinical prospective. In H. Winitz (Ed.), Human communication and its disorders: A review, Vol. 3 (pp. 135-192). Norwood, NJ: Ablex.

[28] Williams. H. (1989). Multiple representations and auditory-motor interactions in the avian song system. In M. Davis, B.L. Jacobs & R.I. Schoenfeld (Eds.), Modulation of defined vertebrate neural circuits. Annals of the New York Academy of Sciences, vol. 563, pp. 148-164.

[29] Lenneberg, E. (1967). Biological foundations of language. New York: Wiley.

[30] Locke, J. (1993). The child's path to spoken language. Cambridge, MA: Harvard University Press.

[31] Smith, A. (1992). The control of orofacial movements in speech. Critical Reviews in Oral Biology and Medicine, 3, 233-267.

[32] McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. Nature, 264, 746-749.

[33] Reed, C.M., Durlach, N.I., Braida, L.D., & Schultz, M.C. (1989). Analytic study of the Tadoma method: Effects of hand position on segmental speech perception. Journal of Speech and Hearing Research, vol. 32, pp. 921-929.

[34] Ghitza, O. (1988). Temporal non-place information in the auditory-nerve firing patterns as a front-end for speech recognition in a noisy environment. Journal of Phonetics, vol. 16, pp. 109-124.

[35] MacNeilage, P.F. (1994). Prolegomena to a theory of the sound pattern of the first spoken language. Phonetica, vol. 51, pp. 184-194.

[36] Klich, R.J., Ireland, J.V., & Weidner, W.E. (1979). Articulatory and phonological aspects of consonant substitutions in apraxia of speech. Cortex, vol. 15, pp. 451-470.

[37] Marquardt, T.P., Reinhart, J.B., & Peterson, H.A. (1979). Markedness analysis of phonemic substitution errors in apraxia of speech. Journal of Communication Disorders, vol. 12, pp. 481-494.

[38] Fujimura, O. (1990). Methods and goals of speech production research. Language and Speech, vol. 33, pp. 195-258.

[39] Browman, C.P., & Goldstein, L. (1992). Articulatory phonology: An overview. Phonetica, vol. 49, pp. 155-180.

[40] Thelen, E. (1991). Motor aspects of emergent speech: A dynamic approach. In N.A. Krasnegor et al. (Eds.), Biological and behavioral determinants of language development (pp. 339-362). Hillsdale, NJ: Erlbaum.

[41] Weismer, G., Tjaden, K., & Kent, R.D. (in press). Can articulatory behavior in motor speech disorders be accounted for by theories of normal speech production? Journal of Phonetics.