MODERN METHODS OF INVESTIGATION IN SPEECH PRODUCTION

Osamu Fujimura, Bell Laboratories, Murray Hill, New Jersey 07974
U.S.A.

Introduction

The natural process of speech production may be discussed on several levels beginning with the cortical level and ending with the acoustic signals. The higher the level, the less applicable direct physical measurements are. Recent efforts by psychologists are focused on temporal aspects of motor control, in an attempt to infer basic mechanisms of cortical programming and its execution. Techniques such as adaptation and reaction time measurements are now being used for direct observation of speech production processes (e.g. Sternberg et al. 1978), and it is hoped that such techniques in combination with powerful physical measurements of speech articulation processes will trigger a new development in this area of research.

Several interesting proposals have been made about the basic principle of articulatory dynamics trying to relate abstract and discrete phonological codes to the temporal structures of continuous speech phenomena (see Kent and Minifie 1977 for a review). The notion of coarticulation (Öhman 1967) still requires a general definition in relation to the basic process of concatenating well-defined phonetic units (Fujimura and Lovins 1978). Information on actual movements of the principal organs is badly needed for such a study. Relatively large amounts of data obtained from the same subject are necessary to cope with an inherent variability of speech production phenomena.

In what follows, we shall try to review recent work on physiological or physical (but not acoustic) observations. Due to the severe space limitation, reference can be made only to a small subset of the representative examples.

Physiological Studies - Muscle Controls

The general question here is which muscle plays the principal role of implementing motor commands for a given phonetic gesture, viz. an elementary articulatory event. Electromyographic studies with use of hooked-wire electrodes, for example by Hirose and Gay (1972), have revealed that the glottal abduction reflecting the devoicing gesture is related to activity of the posterior cricoarytenoid muscles, whereas glottal adduction is achieved by several different muscles in varied ways depending on linguistic (and paralinguistic) functions.

Hirano (1977) recently studied the anatomy and physiology of the vocal cords using various advanced techniques such as electron-microscopy, histochemistry, electromyography, electric nerve stimulation, high speed motion picture, and mechanical measurements, applied to both human and animal larynxes. He arrived at a cover-body approximation of the vocal cords, which reminds us of the earlier account by Svend Smith. Baer has provided a detailed study of excised canine larynxes, and Titze and his coworkers are contributing a new computerized model of the vocal cord vibration process.

Lingual muscles are difficult to study, but the rather limited information obtained by EMG measurement is indispensable for inferring muscular functions relative to specific phonetic gestures. Of particular importance is the use of computational models simulating the tongue deformation as the result of muscular contraction patterns. A three-dimensional static model using the finite-element method has been initiated by Kiritani and substantially extended by Kakita. The role of orosensory patterns in defining targets of articulatory gestures has been discussed by Stevens and Perkell (1977) in relation to the quantal nature of speech. Controlled interference, by such techniques as anesthesia and bite block, has been used to study the effects of feedback on articulatory gestures. The complexity of speech physiology and the highly experienced human strategies in speech behavior tend to make an interpretation of the results of such experiments rather difficult, but some interesting findings are available (Lindblom et al. 1978). A servo-mechanistic technology for controlling mechanical load for dynamically specified load impedance can be used for a control-theoretical analysis of natural articulatory systems. According to Abbs and coworkers, the frequency response of feedback loop systems for articulators seems to allow actively controlled movements of articulatory organs via brainstem feedback, but there are occasions in speech articulations where so-called ballistic-type inertia-controlled movements of articulators are observed (Fujimura 1961). On the other hand, a dynamic palatographic study suggested feedback-controlled tongue tip movements for apical stop gestures (Fujimura et al. 1973b).

<u>Physical States of Organs</u>

There are several stages of information mapping between physiologic motor control, the resultant muscular contraction patterns and the sound output signals. An efficient computational procedure for studying the relation between vocal tract area functions and formant patterns has been proposed by Mathews and coworkers (Atal et al. 1978). There is considerable interaction between the source and the vocal tract, and this situation can be computer-simulated by a composite vocal-cord vocal-tract system (Flanagan et al. 1975).

The physiologic control of the larynx is parametric in the sense that usually gross average states of the larynx rather than details of vibratory changes of the vocal cord shapes within each voice fundamental period are adjusted. The fiberscopic technique developed by Sawashima and Hirose (1968) or its stereoscopic version is appropriate for studying such parametric states of the larynx. Much knowledge has been gained by the use of the fiberscope. In particular, the state of the glottal aperture during the oral closure for stop consonants with various types of laryngeal control is now relatively well known for languages such as Korean, French, Hindi, Tibetan, as well as English, Swedish, and Japanese. Electric measurement of the glottal state is also useful for phonetic studies.

There have been several methods proposed and tested in the past decade for observing tongue movements: dynamic palatography, its extension to palato-lingual distance measurements, magnetic as well as ultrasonic measurements. The most direct and informative method for observation of tongue movement is the use of x-rays for lateral views of the tongue. There were two factors that made radiographic measurements impractical for obtaining a large quantity of speech data: radiological disturbance and the time-consuming frame-by-frame analysis. A new computer-controlled x-ray microbeam system was devised to overcome these technical difficulties (Fujimura et al. 1973a). A full-scale system is now in operation at the University of Tokyo (Kiritani et al. 1975), and is producing useful data about movements of metal pellets placed on selected points of the articulators. Computer programs have been designed and implemented at Bell Laboratories in order to give the experimenter an efficient tool for interactive data analysis. An automatic algorithm has been devised which, according to specified phonetic symbols, identifies the time domains where relevant articulatory activities (and sound characteristics) are found. This system is useful both for assisting the experimenter in retrieving relevant parts of data, and for testing hypotheses about inherent characteristics of individual phonetic events.

An independent measurement of area functions by acoustic input impedance measurement has been proposed (Sondhi and Gopinath 1972).

Statistical Processing of Production Data

Through purely statistical processes, elementary component gestures of the tongue have been derived (Ladefoged 1977; Kiritani 1977). This inductive method gives us phenomenologically derived "phonetic coordinates" for describing articulatory characteristics of classes of phonetic units, classes defined by the particular choice of the speech material used for this data processing. How the results relate to our linguistic experience is an interesting question. Shirai and Honda (1977), along with other groups, assumed a simple dynamic model of the articulator movements to determine the parameters that characterize the natural system. These methods are useful particularly when we need preliminary guidelines for designing components of a larger-scaled deductive experimentation -- synthesis by rule with a comprehensive scope of simulation of human speech production.

References

Atal, B. S., J. J. Chang, M. V. Mathews, and J. W. Tukey (1978): "Inversion of Articulatory-to-Acoustic Transformation in the Vocal Tract by a Computer-Sorting Technique", JASA 63, 1535-1555.

Flanagan, J.L., K. Ishizaka, and K.L. Shipley (1975): "Synthesis of Speech from a Dynamic Model of the Vocal Cords and Vocal Tract", Bell Syst Tech J 54(3), 485-506.

Fujimura, O. (1961): "Bilabial Stop and Nasal Consonants: A Motion Picture Study and its Acoustical Implications", JSHR 4, 233-247.

Fujimura, O., S. Kiritani, and H. Ishida (1973a): "Computer-Controlled Radiography for Observation of Movements of Articulatory and Other Human Organs", Comput Biol Med 3, 371-384.

Fujimura, O., I. F. Tatsumi, and R. Kagaya (1973b): "Computational Processing of Palatographic Patterns", JPh 1(1), 47-54.

Fujimura, O. and J. B. Lovins (1978): "Syllables as Concatenative Phonetic Units", in Syllables and Segments, A. Bell and J. B. Hooper (eds.), North-Holland Pub. Co.

Hirano, M. (1977): "Structure and Vibratory Behavior of the Vocal Folds", in Dynamic Aspects of Speech Production, M. Sawashima and F. S. Cooper (eds.), 13-30, U. Tokyo Press.

Hirose, H. and T. Gay (1972): "The Activity of the Intrinsic Laryngeal Muscles in Voicing Control -- an Electromyographic Study", Phonetica 25, 140-164.

Kent, R. D. and D. Minifie (1977): "Coarticulation in Recent Speech Production Models", JPh 5, 115-133.

Kiritani, S., K. Itoh, and O. Fujimura (1975): "Tongue-Pellet Tracking by a Computer-Controlled X-ray Microbeam System", JASA 57(6,2), 1516-1520.

Kiritani, S. (1977): "Articulatory Studies by the X-ray Microbeam System", in Dynamic Aspects of Speech Production, M. Sawashima and F. S. Cooper (eds.), 171-190, U. Tokyo Press.

Ladefoged, P. N. (1977): "The Description of Tongue Shapes", in Dynamic Aspects of Speech Production, M. Sawashima and F. S. Cooper (eds.), 209-222, U. Tokyo Press.

Lindblom, B., R. McAllister, and J. Lubker (1978): "Compensatory Articulation and the Modeling of Normal Speech Production Behavior", in Articulatory Modeling and Phonetics, R. Carré, R. Descout, and M. Wajskop (eds.), 148-161, G.A.L.F. Groupe de la Communication Parlée.

Öhman, S. E. G. (1967): "Numerical Model of Coarticulation", JASA 41(2), 310-320.

Sawashima, M. and H. Hirose (1968): "New Laryngoscopic Technique by Use of Fiber Optics", JASA 43(1), 168-169.

Shirai, K. and M. Honda (1977): "Estimation of Articulatory Motion", in Dynamic Aspects of Speech Production, M. Sawashima and F. S. Cooper (eds.), 279-302, U. Tokyo Press.

Sondhi, M. M. and B. Gopinath (1972): "Determination of Vocal-Tract Shape from Impulse Response at the Lips", JASA 49, 1867-1873.

Sternberg, S., S. Monsell, R. L. Knoll, and C. E. Wright (1978): "The Latency and Duration of Rapid Movement Sequences: Comparisons of Speech and Typewriting", in Information Processing in Motor Control and Learning, G. E. Stelmach (ed.), 117-152, Academic Press.

Stevens, K. N. and J. S. Perkell (1977): "Speech Physiology and Phonetic Features", in Dynamic Aspects of Speech Production, M. Sawashima and F. S. Cooper (eds.), 323-341, U. Tokyo Press.