

## SOME DYNAMIC ASPECTS OF SPEECH PRODUCTION

STANISŁAW PUPPEL

*Adam Mickiewicz University, Poznań*

### 1. Introduction

The title of the paper suggests a "dynamic" approach to the phenomenon of speech production. One may ask whether it is adequate? And if it is adequate, what could possibly be meant by the notion "dynamic" with regard to speech production? A standard definition of the notion "dynamics" that can be found in a handy desk dictionary (e.g. *Webster's New Collegiate Dictionary*, 1980 : 352) is formulated as follows: "a branch of mechanics that deals with forces and their relations primarily to the motion but sometimes also to the equilibrium of bodies". Obviously, forces do not apply in a vacuum. Rather, we can speak of forces which act in mechanical structures. The human body may also be envisaged in that way and may thus be regarded as a complex biomechanical system which is moved by biomechanical forces, especially muscular forces. In this respect, the human body is subject to the study of a branch of physics called "biomechanics".

The production of speech is but one small — though incredibly intricate — part of human biomechanical behaviour. It requires the skillful coordination of a serially ordered sequence of continuously changing muscular movements in the respiratory, laryngeal and articulatory domains of the human "speech apparatus". Thus, speech production can be viewed as movements of bodies (or various coordinated muscular structures) through space. If these movements are to be carried out as planned by the speaker, safeguarding against any possible and undesirable errors is secured by a heavy reliance on constant feedback from the speaker's senses of sight, hearing and balance. These senses are *exteroceptive* in nature and they generally inform the speaker of the relation of the body to the world outside it. Quite obviously, speech production is unique in this respect because the speaker is not so much informed about the position of his body in the outside world as he is informed about the behaviour of the muscular structures in the "inner space" of the vocal tract which is placed entirely inside the speaker's body. Thus, the speaker has to rely on *somesthetic*

feedback which, while involving such aspects as touch, limb position and speed of muscular displacement, informs him about the interior of the body.

Closing our introductory remarks we must not forget that for most people, speech production has an intended character, i.e. in normal humans it is consciously subjected to the uniquely human capacity for employing complex symbolic representations in an expression of meaning. This clearly cultural aspect provides a "teleological" framework for the study of speech production. However, in this paper we wish to discuss the more physical aspects of speech production (e.g. the continuous nature of speech gestures) with regard to the fundamental notion of "target sound" and its role in the dynamics of speech production.

## 2. "Aims" of the air stream mechanisms

Speech as a voluntary muscular and articulatory activity employs movements of the vocal tract in order to control air flow and change the shape of the vocal tract simultaneously so that the resultant output is a continuous sequence (or stream) of sounds. It follows from the above that the intention to speak implies making sounds. Furthermore, it should be stressed that the positions and movements of articulators should not be taken as ends in themselves but the means of making sounds. Sound production is the result of coordinated activities of three major air stream mechanisms whose "local" functions are as follows:

- (a) the *respiratory mechanism* provides compressed air as a source of power for the production of speech;
- (b) the *laryngeal mechanism* (also referred to as the phonatory mechanism) functions as a converter of energy of compressed air into kinetic energy of acoustic vibrations; and
- (c) the *supralaryngeal mechanism* (also called the supraglottal mechanism) finally modifies air flow and hence shapes the sounds which come from the speaker's mouth.

Together, the three mechanisms contribute synergistically to the production of target sounds, most notably "articulatory targets".

## 3. Articulatory targets

The targets a speaker seeks to achieve are speech sounds. These are assumed to have their mental (ideal) "iconic" representations and are thus known by the speaker as aggregates of spatial, phonational, aerodynamic, and acoustic targets which the speaker aims to produce. This he does by activating the air stream mechanisms through patterns of muscle contractions which result in a series of fixed (or static) positions and transitions between them. What we are

proposing at this point is that no articulatory target can be regarded solely as a *static target*, i.e. a fixed position of muscular structures in three-dimensional space. Rather, we should define articulatory targets as both a fixed position and an accompanying *dynamic target* understood as movement(s) through three-dimensional space. In fact, the articulation of any sound, whether in isolation or in context, is characterized by executions of movements of muscular structures with a varying degree of swiftness. That is why no target can be described by making exclusive recourse to its static portion but must also include information on the (preceding and following) trajectory through the inner space of the vocal tract. This way of defining the speech sound is particularly useful in describing a continuum of successive positions between which there occur transitions comprising information on both the preceding and following sound segments (Fig. 1).



Fig. 1

(adapted from Daniloff et al., 1980)

where:

A, B, C, and D represent target sounds, and  $T_{AB}$ ,  $T_{BC}$ , and  $T_{CD}$  represent transitions between sounds.

As we have indicated above, each target sound can be defined as an aggregate of ideal spatial (both static and dynamic), phonational, aerodynamic, and acoustic parameters. Consider the pair of target sounds [s] and [z]. In order to produce them, the speaker must "know" that the air should be directed egressively, that the vocal folds should be abducted for the voiced sound and adducted for the voiceless one, that the front part of the tongue should traverse the space of the oral cavity and position itself against the alveolar ridge (together with a slight fronting movement of the mandible) in order to produce friction noise, and that the friction noises obtained appropriately characterize the acoustic patterns of the target sounds in question. The speaker must also know how to coordinate the intricate muscular actions in a synergistic manner. Finally, each target sound may be described in terms of its aerodynamic parameters, e.g. in terms of pulmonic and intra-oral pressure. In the particular case of [s] and [z], pressure depends on the generated resistance which, in turn, depends on the length of the constriction as well as on the cross-section area (cf. Scully, 1971). It is assumed here that a knowledge of these four parameters decisively contributes to the speaker's mental representation of any target sound as a dynamic entity in the sense that these parameters are its intrinsic components.

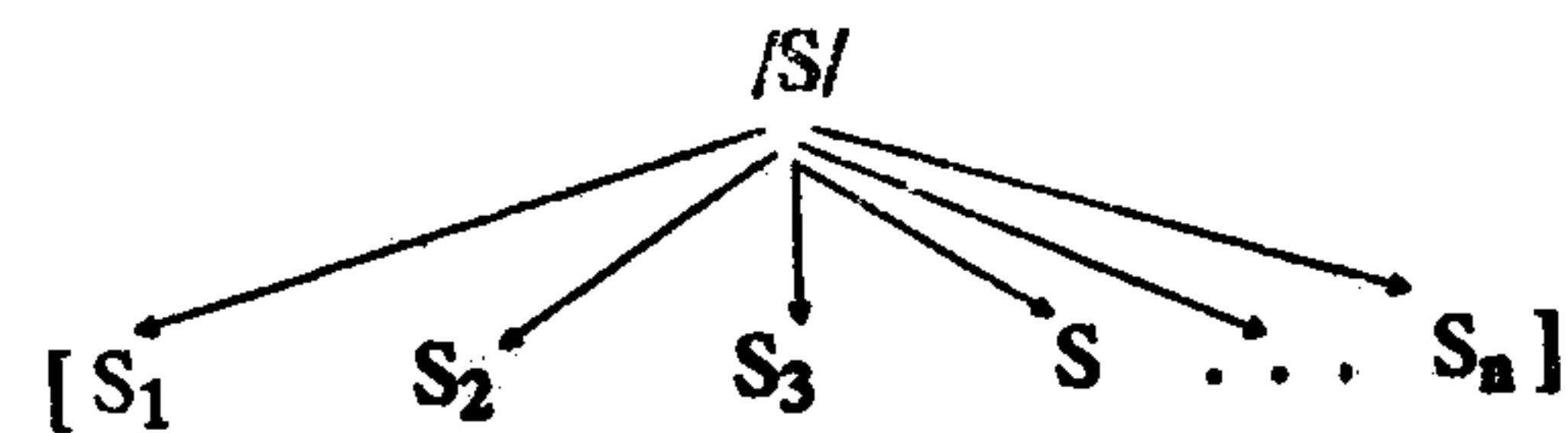
#### 4. Precision of muscular movements in articulation

Let us now consider the reasons why humans are so immensely skillful at producing a stream of sounds. The first reason appears to be connected with the fact that every speaker employs a complex feedback system (which is partly exteroceptive and mostly somesthetic) which informs him about the intensity and precision of movements of muscular structures in action. For example, speakers are informed about the length, tension, and velocity of stretch of various muscle fibres which participate in the muscular movements within the inner space of the vocal tract (cf. Smith, 1971; also Matthews, 1977). Furthermore, these movements are monitored by listening to the quality of the sounds produced. As a result of this tight cooperation between the two monitor systems, speakers can both *feel* and *hear* their intended articulatory movements. Obviously, any damage to this complex surveillance system always results in a varying degree of severe impairments in speech production (cf. Lenneberg, 1967; Haggard et al., 1971; Leonard, 1973).

Another reason which appears to bear on the precision of sound production concerns the size of the space in which these movements occur. The space of the vocal tract is indeed small in size and therefore it only allows for a limited number of possible muscular movements within its confines. The fact of certain muscular discreteness is well-documented in the various phoneme inventories, in particular in the traditional manner of articulation and place constraints on these inventories (see, for example, the UCLA Phonological Segment Inventory Database, 1981). It must also be kept in mind that every normal and healthy speaker performs the act of producing a given muscular movement pattern for an exceedingly large number of times. This last factor contributes decisively to the fact that on the muscular level speech is a well-practiced (overlearned) and highly automatic (i.e. stereotyped) behaviour.

#### 5. Variability in speech production

What we can observe in speech is the occurrence of alternate renderings of the underlying segment. Why do these allophonic variations occur? In what follows we shall be concerned with possible sources of allophonic variation. Basically, the situation phoneticians encounter is as follows:



where:

/S/ is an ideal (abstract) segment, and  
[S<sub>1</sub> ... S<sub>n</sub>] are target allophones

A standard account of the situation illustrated in Fig. 2 involves looking at the string of target allophones as a class of equivalent sounds whose differences are not distinctive. Furthermore, as a group they are representative of the underlying ideal segment. The problem, however, should not be foreclosed at that, for in the speech of any speaker one can observe two kinds of allophones. The first of these, so-called "extrinsic" allophones, are those that clearly belong to the sound pattern of a given language. For instance, the velar/alveolar alternation in English: the velar variant being produced in word-final position, as in *bill*, and the alveolar variant being produced in word-initial position, as in *line*. However, both can be easily interchanged by the native speaker of English which follows from the fact that the speaker has a fully conscious control over their rendition. As members of the sound pattern of the language, they can be used to mark word (and syllable) position, juncture boundaries, etc. Thus, we may say that the source of extrinsic variability is linguistic in nature.

The second type of allophones, so-called "intrinsic" allophones, do not appear to be a part of the sound pattern of the language. Rather, they reflect the limitations and preferred modes of operation of the *speech production mechanism* (SPM) which can be defined as the peripheral neuro-motor-articulatory system responsible for the actual production of sounds and their coarticulated complexes (cf. Puppel, 1988). The limitations of the SPM reflect the limitations of the vocal tract as a biomechanical system and necessarily follow from the fact that the articulating bodies can only move so far and so fast. For example, movements of the body of the tongue are limited in the inner space of the vocal tract, and so is the speed of its movements. Likewise, the vocal folds cannot go into action unless there exists a pressure differential above and below them. A fine embodiment of limitation of the SPM is the so-called "carry-over" coarticulation. On the other hand, speech production falls within the preferred modes of operation of the SPM which, in turn, is a part of the preferred and most natural (i.e. economical) modes of behaviour in all bio-systems. This last point is built on the assumption that all living systems must be subject to the same principles (e.g. that the swiftness and precision of an action is mass dependent, or that a system behaves in such a way as to have as low an energy cost as possible, etc.). In speech production a good example of intrinsic allophonic variation representing a preferred mode of operation of the SPM as a biomechanical system is the production of a velar stop adjacent to a rounded vowel, as in *cool*. In this word, the velar stop consonant is articulated as rounded, which indicates that a neuro-motor-articulatory adjustment of rounding occurs prior to the round vocalic segment. This phenomenon is known under the name of "anticipatory" coarticulation. It follows from the above example that *context* is perhaps the most important source of intrinsic variation. The other important source is the mechano-inertia of the vocal tract structures which have differing degrees of inertia, elasticity, and resistance.

## 6. Concluding remarks

In this paper I have suggested a dynamic treatment of speech production which also prompts a "systems" approach. It is believed that the production of speech is a complex interactive process involving not only the phonological sub-system of language but also symbolic, motor, and feedback functions. They all seem to be at work in the performance of speech. The aspect essential to the operation of the production of serially ordered speech sounds is the existence of the phenomenon of speech targets. They are — performance-wise — the final authority against which all other aspects of speech production should be judged. And although they can be modified in continuous speech due to a number of contextual (coarticulatory and mechano-inertial) limitations, this superbly flexible behaviour is fundamentally directed by a handful of ideal target sounds as primitive goals. In this respect, the existing phonetic variability does not appear to be totally random but is rigidly regulated by both linguistic (phonological) and non-linguistic (SPM constraints) factors.

## REFERENCES

- Daniloff, R., Schuckers, G. and L. Feth. 1980. *The physiology of speech and hearing*. Englewood Cliffs, N. J. Prentice-Hall, Inc.
- Dickerson, W. B. 1977. "Language variation in applied linguistics". *ITL, A Review of Applied Linguistics*. 35. 43—66.
- Haggard, M. P., Corrigall, J. M. and A. G. Legg. "Perceptual factors in articulatory defects". *Folia Phoniatica*. 23. 33—40.
- Lenneberg, E. H. 1967. *Biological foundations of language*. New York: John Wiley and Sons, Inc.
- Leonard, L. B. 1973. "The nature of deviant articulation". *Journal of Speech and Hearing Disorders*. 38. 156—161.
- Matthews, P. C. B. 1977. "Muscle efferents and kinesthesia". *British Medical Bulletin*. 33. 137—142.
- Puppel, S. 1988. *Aspects of the psychomechanics of speech production*. Poznań: Wydawnictwo Naukowe UAM.
- Scully, C. 1971. "A comparison of /s/ and /z/ for an English speaker". *Language and Speech*. 14. 187—200.
- Smith, T. S. 1971. *A phonetic study of the function of the extrinsic tongue muscles*. *UCLA Working Papers in Phonetics*. 18. Los Angeles: University of California.
- UPSID: *UCLA Phonological Segment Inventory Database*. 1981, *UCLA Working Papers in Phonetics*. 53. Los Angeles: University of California.
- Webster's New Collegiate Dictionary*. 1980. Springfield, Mass.: G. & C. Merriam