Studia Anglica Posnaniensia XXVIII, 1994

SOME ASPECTS OF THE MONITORING (REGULATION) FUNCTION IN SPEECH PRODUCTION

STANISŁAW PUPPEL

Adam Mickiewicz University, Poznań

1. Introduction

The literature on the subject of speech production agrees on the fact that successful speech production rests on the prior creation of a blueprint comprising the semantic, syntactic, and phonological characteristics of a speech utterance in the mind of a prospective speaker. However, any discussion of speech production is basically hampered by the fact that it cannot avoid being speculative. This is because of the inaccessibility of the human brain to direct experimental investigation. Yet substantial evidence that has been accumulated to date with regard to the 'speech output', expressible directly through the dynamics of motor-articulatory activity, leaves no doubt as to the temporal-linear-hierarchical organization of speech.

It is convenient to distinguish the following stages (or levels) in the production of a speech utterance (cf. Laver 1968, 1969, 1970, 1980; Puppel 1988):

- *a goal (idea) formation level (the so-called ideation function)* during which a central underlying idea is formed;
- *a program-planning level* where the processing of the idea proceeds in terms of the semantic, syntactic, and phonological rules of a given language and where individual words are retrieved from some kind of a lexical storage system;
- *a speech production mechanism level which includes a neuromotor sub-level*, where the neurolinguistic program is converted into a temporally-determined program of motor commands to the appropriate speech muscles, and an *articulatory (myodynamic) execution phase* during which the respiratory, laryngeal, and articulatory muscular structures are activated in order to carry out a well-orchestrated sequence of movements specified in the neuromotor phase.

Obviously, for speech to be *successful* (where "success" directly implies the smoothness of perception both by an interlocutor and the speaker himself and subsequent smoothness of the entire decoding operation), it has to be monitored at various stages in the speech production process. Thus, the *monitoring (feedback) function* is postulated to constitute an integral part of any model of speech production. Although one may speculate that monitoring can exist at all levels of the speech production process, e.g., during the ideation stage (a more central type of monitoring) as well as during the myodynamic phase, one must immediately admit that relatively more is known about the latter, more peripheral, type of monitoring.

Broadly speaking, the monitoring operation, whether running on the more central and cognitively more complex levels of the speech production process or exploited on the largely automatic level of specification of motor commands and subsequent movements of articulatory complexes, ensures the accuracy of the speech generation process. In this way the smooth triggering of successive speech target movements is secured. The standards of accuracy and smoothness which are required for normal fluent speech are made possible only because on the level of the speech production mechanism appropriate motor regulations are facilitated. This implies that, generally, the motor regulator must constantly scan the details of the current speech production program being processed at any time during the speech production process. However, we must also emphasize that disorders can occur involving both the program-planning stage and the speech production mechanism phase. In this case, smoothness and accuracy of speech may be affected and various manifestations of disfluency may hamper the speaker's successful production of speech.

2. Types of feedback in speech production

In order for speech to be fluent and unperturbed, the different levels of the speech production process must possess some sort of logical temporal progression, most naturally with the goal formation and program-planning stages preceding the speech production mechanism phase. Obviously, the monitoring function cannot be simply "plugged" into or after any of these levels but should be conceived of as operating at all levels simultaneously and continuously. It is equally obvious that monitoring ensures the smooth transition from one more global unit of neurolinguistic program to another, be it Boomer's (1965) 'phonemic clause', Halliday's (1963) 'tone-unit' (or 'tone group'), or Lieberman's (1967) 'breath group'. Subsequently, it is assumed that monitoring is also responsible for the operation of the sequential ordering device which, once the neurolinguistic program has been activated, controls the execution of the current program and evokes the generation of a subsequent program. Whether the program is encodable in terms of the minimal units of the size of the 'phoneme' and accompanying invariant motor commands to the articulatory musculature (cf. Harris et al. 1965; Liberman et al. 1967; Öhman 1966), or of the size of the 'syllable' (cf. Fromkin 1968, 1971; Boomer and Laver 1968; Nooteboom 1969), or simply in terms of 'targets' understood as internalized space co-ordinate systems (cf. MacNeilage 1970), or finally in terms of 'target movements' (or 'gestures', cf. Dalton and Hardcastle 1989; Mattingly 1990), can only be taken as speculative in the absence of any direct experimental findings on the nature of neural processing during the speech production process. However, despite these difficulties what is unquestionable about feedback regulations in speech production is that there exist three main sensory channels through which information is fed back to the central nervous system (CNS) in order to provide ongoing control of articulation. These include:

(a) a *tactile* feedback channel which accumulates and transmits information from mechanoreceptors (placed in the superficial layers of oral mucosa) concerning the precise localization of contact between articulatory muscular structures, the strength and pressure of contact, direction of movement as well as the timing of these contacts;

(b) a *proprioceptive* feedback channel which conveys information mainly from muscle spindles, Golgi tendon organs and joint receptors (situated in the speech musculature) concerning the length of muscle fibres, degree of stretch of the fibres, velocity of stretch, direction of movement of muscle, rate, direction and extent of joint movement as well as rate of change of muscle length;

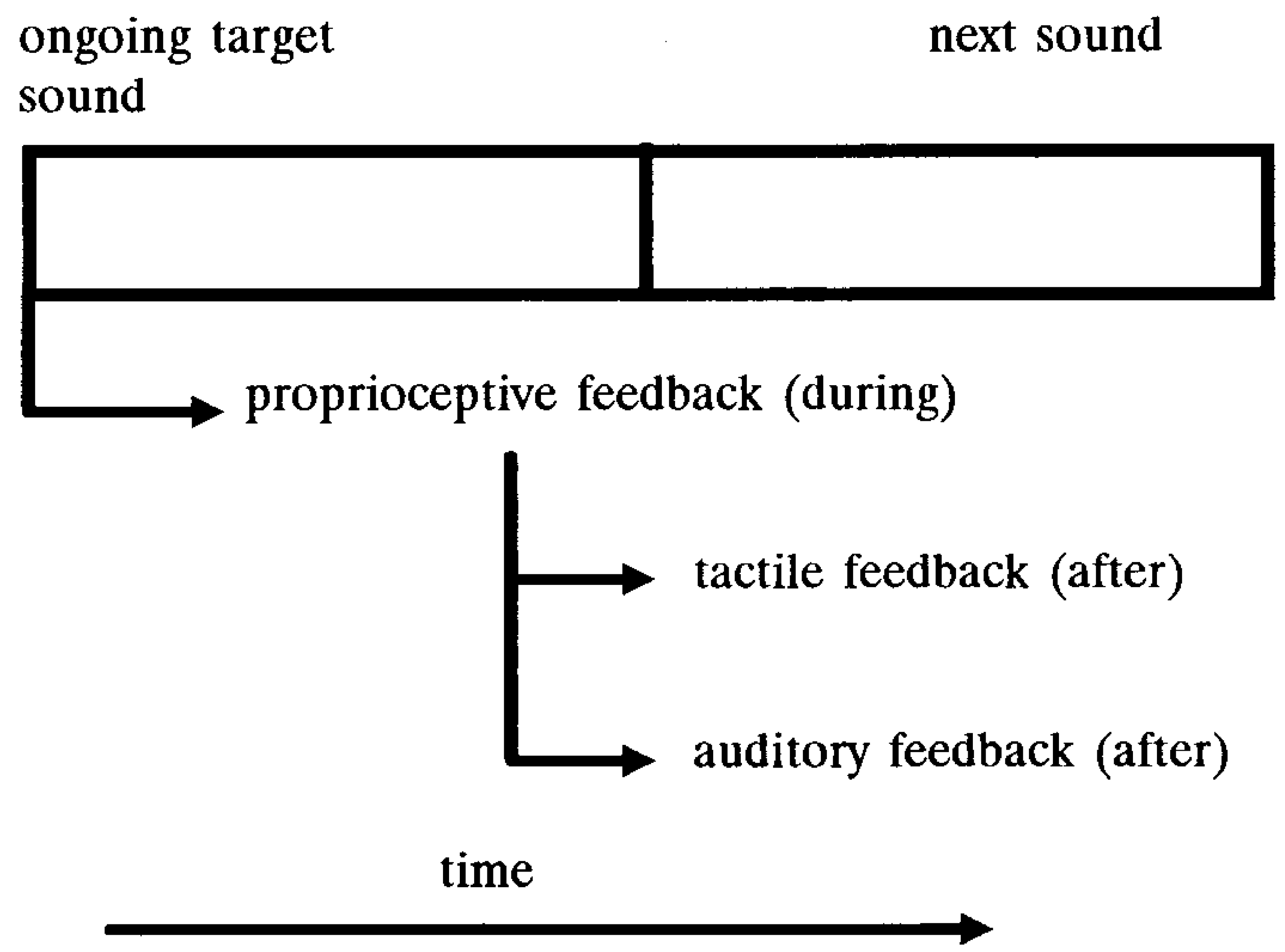
(c) an *auditory* feedback channel which transmits information from hair-cells in the Organ of Corti (situated in the inner ear ,i.e., cochlea) concerning the acoustic properties of the generated speech sounds, such as intensity, frequency, periodicity of vibrations, duration, and direction of sounds.

The above sensory channels are 'equals and partners' in securing detailed sensory information and enabling ongoing control of the articulatory process. They are also jointly involved in providing a continuous 'report' (or 'plot') on the synchronization of articulatory movements to be compared with the idealized target gestures that a speaker retrieves from his/her long-term memory in the program-planning phase of the speech production process. However, it must be emphasized that these three types of sensory feedback are activated at different stages in the completion of an articulatory target gesture. The differences may be manifested in the following ways:

- tactile monitoring is primarily involved in those types of articulations which are based on crucial contact between particular speech organs, e.g., in a class of sounds referred to as 'lingual', that is, where a meticulous contact between the tongue and the palate is required;
- auditory monitoring is primarily involved in vowel articulations, that is, where there occurs very little contact between the articulatory musculature. It is also significant in providing feedback on the realization of the timing of vocalic gestures as well as suprasegmental features, such as pitch and intensity;

- proprioceptive monitoring is significant for all articulatory activities where precise positioning of the articulatory structures is required. This may be true with regard to such articulations as 'trills' or 'grooved' fricatives [s] and [s] as well as with regard to height/backness distinctions in the production of vowels.

At this point it must also be emphasized that the three types of feedback differ as to the swiftness of their monitoring operation. For example, the proprioceptive monitor system operates via nerve fibres which are characterized by a larger diameter, thus securing an extremely fast conduction velocity of neural impulses to the CNS. The tactile system involves neural fibres which are much smaller in diameter and thus the velocity of conduction of neural impulses is relatively slow. Finally, the auditory monitoring system, relying on the transmission of sound through the air and the bones of the skull, is the slowest of the three. Because of the marked differences in the speed of operations, one may speculate that they are effective at different moments of the production of target gestures. Namely, the proprioceptive system exerts control over the production process *during* the execution of a given sound, that is, during the movement itself, while the tactile and auditory systems come into action *after* contact has been made for that sound. One may illustrate the different stages in the operation of the three types of sensory feedback by means of the following diagram:



3. The mechanism of regulation in speech production

It is highly probable that the monitoring function is not utilized merely to provide closed-loop control of the speech production process in a chain of associations, whereby each stage of the planning level awaits the error signal of the previous stage to equal zero. Rather, it has been postulated (cf. Fairbanks 1954; Dalton and

Hardcastle 1989) that feedback and speech production systems can 'scan ahead' and jointly predict the positioning and coordination of the articulatory structures required for the production of subsequent target gestures. In order for the action of 'extrapolating' into the future components of an utterance to be as smooth as possible, time is essential. Therefore, one can imagine that feedback information, both during and after the speech event, has to travel along afferent fibres back to the CNS with great speed the degree of which differs, however, for the different types of sensory feedback (e.g., proprioceptive feedback was measured at 12-15 msec., cf. Rushworth 1966) with the ultimate goal of ensuring an appropriate correction procedure when an error has been detected.

We can hypothesize that the correction procedure consists of the following processes:

- *inputting* motor commands to the speech musculature;
- *monitoring* the on-going speech events, i.e., collecting information from sensory receptors situated in the oral mucosa, the speech muscles and in the cochlea, and sending it via afferent fibres back to the CNS;
- *comparing* the overall sensory information on the actual speech gestures with the idealized target schemas;
- *modifying* motor commands on the neuromotor level, with the latter functioning as a motor regulator.

Of special significance in facilitating the smooth production of a sequence of articulatory targets is the process of comparison which, if disrupted, may be effectively prevented from guaranteeing sensorimotor coordination. The process of comparison is hypothesized to comprise the intended (idealized) target schema and the actual speech gesture. In this way learning and performing a motor skill may be regarded as essentially the discovery and preservation of optimal self-organization (cf. Kerr, 1982). Obviously, in a normal speaker, fluency of motor performance is safeguarded due to an uninterrupted flow of information to the comparator. On the other hand, a number of sensori-motor speech pathologies which impair normal articulatory coordination may be generated by a disruption between the comparator and incoming sensory information.

4. Some pathologies in speech production

It is quite obvious that successful speech production requires the structural integrity of the speech production mechanism as well as a smooth recoding (transmission) of the program-planning operations into the neuro-motor-articulatory act. Moreover, it follows from what has been said above that normal speech production is also dependent upon the continuous presence of the monitoring function which relays information via afferent nervous pathways to the CNS. Obviously, any damage to the central nervous system, peripheral nervous system, or both, may be manifested in different kinds of language-speech disorders. They come under the common name of 'neurogenic communicative disorders' (cf. Dworkin and Hartman

1988), of which 'apraxia of speech' and 'dysarthria' are particularly noteworthy. More precisely, the impairments in the monitoring function may be caused by lesions in the sensory and motor areas of the dominant (left in most cases) cerebral hemisphere. Below we briefly consider these disorders.

a. Apraxia of speech
Apraxic speech disturbances are regarded as resulting not so much from weakness and slowness of articulatory musculature as from faulty operations at a higher programming/execution level of the speech production process, notably the neuromotor level. Apraxia of speech is predominantly characterized by global lack of coordination between the lingual, mandibular, labial, palatal, and laryngeal programs so that an apraxic speaker demonstrates unpredictable and inconsistent articulatory errors. Most typically they include: substitutions, additions, omissions of sounds, repetitions, and prolongations of articulatory gestures. The sites of lesions resulting in apraxic speech have been identified as Broca's convolution or dominant supplementary motor cortex.

b. Dysarthria
Dysarthria is the name assigned to a group of speech disorders caused mainly on the myodynamic level of the speech production process. They are characterized by a lack of fine muscular coordination in the execution of articulatory gestures due to paralysis, paresis, weakness and slowness of muscular activities. Typical and demonstrable effects include imprecise and slow-laboured articulatory movements due to damage to some cranial nerves, breathy phonation, hypernasality, harsh vocal quality, decreased loudness, prolonged syllables, and excessive pausing as well as generally imprecise vowel and consonant articulations.

5. Summary

It can be seen from the foregoing discussion that the production of fluent and normal speech can only be properly comprehended if one views speech production as a complex multi-stage encoding operation accompanied by constant surveillance on the part of the monitoring function. Traditional phonetic approaches to the problem of the generation of speech render a fundamentally unsatisfactory view of running speech as they fail to consider the significance of the feedback operation. The role of this operation, differentially affected under various pathological conditions, is of vital importance to the coordinative style of articulatory processes in normal interactions by way of the speech code.

REFERENCES

Andrew, B.L., ed. 1966. *Control and innervation of skeletal muscle*. Edinburgh: Livingstone.

Boomer, D.S. 1965. "Hesitation and grammatical encoding". *Language and Speech* 8. 148-58.

Boomer, D.S. and Laver, J.D.M. 1968. "Slips of the tongue". *British Journal of Disorders of Communication* 3. 2-12.

Dalton, P. and Hardcastle, W.J. 1989². *Disorders of fluency*. London: Cole and Whurr Limited.

Dworkin, J.P. and Hartman, D.E. 1988. *Cases in neurogenic communicative disorders*. Boston: Little, Brown and Company.

Fairbanks, G. 1954. "A theory of the speech mechanism as a servosystem". *Journal of Speech and Hearing Disorders* 19. 133-9.

Fromkin, V.A. 1968. "Speculations on performance models". *Journal of Linguistics* 4. 47-68.

Fromkin, V.A. 1971. "The non-anomalous nature of anomalous utterances". *Language* 47. 27-52.

Fromkin, V.A., ed. 1980. *Errors in linguistic performance. Slips of the tongue, ear, pen, and hand*. New York: Academic Press.

Halliday, M.A.K. 1963. "The tones of English". *Archivum Linguisticum* 15. 1-18.

Harris, K.S., Lysaught, G.F. and Schvey, M.M. 1965. "Some aspects of the production of oral and nasal labial stops". *Language and Speech* 8. 135-47.

Kerr, R. 1982. *Psychomotor learning*. Philadelphia: Saunders College Publishing.

Laver, J. 1968. "Phonetics and the brain". *Work in progress*. (Department of Phonetics and Linguistics, University of Edinburgh) 2. 63-75.

Laver, J. 1969. "The detection and correction of slips of the tongue". *Work in progress*. (Department of Phonetics and Linguistics, University of Edinburgh) 3. 1-12.

Laver, J. 1970. "The production of speech". In Lyons, J., ed. 1970. 53-75.

Laver, J. 1980. "Monitoring systems in the neurolinguistic control of speech production". In Fromkin, V.A., ed. 1980. 287-305.

Liberman, A.M. 1967. "Some observations on a model for speech perception". In Wathern-Dunn, W., ed. 1967. 68-87.

Lieberman, P. 1967. *Intonation, perception, and language*. Cambridge, Mass.: MIT Press.

Lyons, J., ed. 1970. *New horizons in linguistics*. London: Penguin.

MacNeilage, P.F. 1970. "Motor control of serial ordering of speech". *Psychological Review* 77. 182-96.

Mattingly, I.G. 1990. "The global character of phonetic gestures". *Journal of Phonetics* 18. 445-52.

Nooteboom, S.G. 1969. "The tongue slips into patterns". In Sciarone, A.G., ed. 1969. 114-32.

Öhman, S.E.G. 1966. "Coarticulation in VCV utterances: spectrographic measurements". *Journal of the Acoustical Society of America* 39. 151-68.

Puppel, S. 1988. *Aspects of the psychomechanics of speech production*. Poznań: Wydawnictwo Naukowe UAM.

Rushworth, G. 1966. "Some functional properties of deep facial affarents". In Andrew, B.L., ed. 1966. 125-33.

Sciarone, A.G., ed. 1969. *Nomen, Leyden studies in linguistics and phonetics*. The Hague: Mouton.

Wathern-Dunn, W., ed. 1967. *Models for the perception of speech and visual form*. Cambridge, Mass.: MIT Press.