EFFECTS OF EFFERENT AND AFFERENT INTERFERENCE ON SPEECH PRODUCTION: IMPLICATIONS FOR A GENERATIVE THEORY OF SPEECH MOTOR CONTROL

Thomas Gay and Michael Turvey, Haskins Laboratories, New Haven, Connecticut, U.S.A.

One might claim that speech production proceeds in open-loop fashion: for a given speech sound a motor program prescribes a standard set of instructions to the musculature. Against this claim, however, is the fact that the backdrop of articulatory states into which the standard instructions would be inserted is itself not standardized. The initial conditions (or contexts) for the articulatory gestures yielding a given speech sound vary considerably (cf. MacNeilage, 1970). This is a most notable feature of speakers: within reasonable limits they are capable of producing the necessary configurations of articulatory maneuvers for the sounds of speech even though the departure points for those configurations are ever-varying. Moreover, it appears that the configuration of gestures underlying a desired speech sound can be generated with virtually no experimentation and without the benefit of auditory monitoring. Lindblom and his co-workers (Lindblom et al., 1978) have shown that speakers fitted with bite blocks can produce isolated vowels within the range of variability for normal vowel production, and that satisfactory formant matching occurs on the first pitch pulse of the first attempt.

Clearly, the adaptive, generative nature of articulation is not captured by the notion of open-loop control. Consequently, speech investigators have turned to the claim that the control of speech is closed-loop. In closed-loop explanations, a sensory referent is proposed that relates either to the environmental goal of the articulatory gestures, such as a spatial target or an acoustic pattern, or to the movement-producing commands. (The interpretation of sensory referent as a spatial target is currently the more popular interpretation.) The comparison of sensory feedback with the sensory referent yields an error signal that provides the basis for adjusting the lower level motor mechanism(s) responsible for controlling the referent. Over successive comparisons an increasingly closer match between the feedback sensory signal and the desired sensory referent is achieved.

While a closed-loop mechanism can, in principle, adjust motor instructions to variable initial conditions to attain the referent, it is not immediately obvious how a feedback mechanism that _gradually_ approaches a desired result could underly the _immediate_ adjustment to context evidenced in everyday speaking and underlined by the phenomenon reported by Lindblom and his colleagues. What is needed is a mechanism that: (1) can produce the appropriate articulatory gestures in the face of variable and often novel initial conditions, and (2) can do so without trial and error.

On first thought, these two criteria are met by model-referenced control. Here, the closed-loop mechanism tied to the peripheral speech apparatus is modeled centrally so that motor commands and their sensory consequences can be simulated for the current conditions of the peripheral speech apparatus. The simulated motor commands that result in a match between the simulated sensory feedback and the sensory referent are then realized as actual motor commands. In principle, the predictive simulation of model-referenced control could underly the immediate readjustment phenomenon (Lindblom et al., 1978). There is, however, a potentially serious drawback to any closed-loop explanation: While an error signal can index how near the collective action of a number of muscles is to the desired consequence, it does not prescribe in any straightforward way how the individual muscles are to be adjusted to give a closer approximation to the referent (Fowler and Turvey, in press).

There is another mechanism, very different from closed-loop control, that meets the two criteria noted above. The rationalization and evidence for this mechanism - referred to as a coordinative structure - has been presented elsewhere in some detail (Fowler, 1977; Fowler, Rubin, Remez and Turvey, in press; Turvey, Shaw and Mace, in press). A rough sketch must suffice for current purposes.

Consider a set of several (relatively) independent muscles. As an aggregate, the muscles would exhibit a large number of degrees of freedom and would rely on a source external to themselves for their control. The number of degrees of freedom can effectively be reduced by functionally linking the muscles so that they mutually determine one another's states in a systematic fashion. But such linkage control would, in large part, be internal to the set of muscles. Such functional linkages, that render an aggregate of relatively independent muscles into a single autonomous unit, may be conceived of as equations-of-constraint written, as it were, on the ascending and descending neural pathways.

To identify some important features of this latter system, let us compare it with closed-loop control in relation to the problem of uttering a vowel under conditions of efferent and afferent interference. In the closed-loop perspective, to produce a given vowel is to specify a particular spatial target as referent. In the coordinative structure perspective just outlined, to produce vowels is to organize the articulators into a single, autonomous system according to a particular equation (or set of equations) of constraint; and, to produce a given vowel is (perhaps) to parameterize that system in a particular way (cf. Fowler, 1977).

Suppose that a speaker impeded by a bite block is requested to utter a given vowel. The model-referenced version of closed-loop control assumes that the condition of the speech apparatus is sensed and motor commands together with sensory feedback are simulated to determine what needs to be done given these conditions. The coordinative structure perspective simply notes that if some parts of the system are 'frozen' the other parts will, by virtue of the equation(s) of constraint, automatically assume values tailored to that of the frozen part and appropriate for producing the vowel.

Suppose now a speaker is interfered with not by a bite block but by anesthetization of parts of the speech apparatus and, as before, is requested to utter a given vowel. In this situation model-referenced control must suffer to the extent that sensory information about initial conditions is not available. In short, anesthetization should impair vowel production considerably more than a bite block restriction. From the coordinative structure perspective, however, anesthetization and bite block should be equivalent in that neither one alone should seriously perturb vowel production. For some members of a coordinative structure to be 'uninformed' about the states of other members is not important; as long as all members of the structure can vary, equilibration according to the equation(s) of constraint will occur and vowel production will be successful. However, we suspect that if some members cannot vary (due to a bite block) and their values are not communicated within the system (due to anesthetization), then fulfilling the equation(s) of constraint will not be possible and successful vowel production would be seriously hindered. The experiment that follows is a preliminary appraisal of these notions. In it, both efferent and afferent variables were either interfered with directly or controlled indirectly during the production of several isolated vowels. Both acoustic and electromyographic measures were used to determine how speech performance is affected when the linkage among these variables is both partially and completely disrupted.

Method

Subjects were two adult male native speakers of American English, one phonetically trained (WE) and the other phonetically naive (SJ). The speech material consisted of the isolated vowels, /i,a,u/. Four separate articulatory variables were controlled directly and one was controlled indirectly. A bite block and an artificial palate were used to produce direct efferent interference, and anesthesia of the temporomandibular joint (TMJ) and oral mucosa were used to produce direct afferent interference. Two different bite blocks were used, one 23 mm long and the other 3 mm long. The longer bite block was used to fix jaw position for the close vowels /i/ and /u/, and the shorter bite block was used for the open vowel /a/. An acrylic artificial palate was constructed from upper mouth casts of both subjects. This prosthesis was approximately 10 mm thick at the midline, 3 mm thick along its edge, and 5 cc in volume. It extended from the posterior surface of the central incisors to approximately 8 mm anterior to the soft palate. Jaw position afference from the mechanoreceptors of the temporomandibular joint was eliminated by 2 ml of xylocaine injected directly into the joint capsules, bilaterally. Oral mucosa sensation was eliminated by spraying the entire oral cavity with a benzocaine solution.

The recording procedures were as follows: First, the subjects produced three triads of each vowel spontaneously, with the bite block, the artificial palate, and the bite block and artificial palate, in that order. Anesthesia was then applied in two steps. For one subject (WE), the joint was anesthetized first, while for the other subject the topical anesthesia was applied first. In each case, the entire vowel sequence was repeated after each anesthetization. The experiments were run with the subjects seated in front of a microphone. For one subject (WE), electromyographic recordings from the genioglossus (tongue) and orbicularis oris (lip) muscles were obtained using conventional hooked wire techniques. All data were recorded on magnetic tape for later analysis.

Results

For both subjects, the effects of the experimental conditions were variable and evident only for /i/ and /u/; the formant frequencies of /a/ were virtually unaffected by either mechanical interference or anesthesia. Apparently, only pharyngeal cavity variables are relevant to /a/. For both /i/ and /u/, articulatory performance was unaffected by anesthesia alone, the artificial palate under all conditions of anesthesia, and the bite block under normal conditions and under incomplete anesthesia. Performance was affected, however, and dramatically so, when the bite block was introduced either alone or in combination with the artificial palate under complete anesthesia. These effects were substantial not only perceptually, but at the acoustic and muscle activity level as well. For example, first and second formant frequencies of the first spontaneous /i/ produced by subject WE were 275 and 2275 Hz. These values were approached for all experimental combinations except the TMJ + topical anesthesia + bite block and TMJ + topical anesthesia + bite block + artificial palate conditions where first and second formant frequencies shifted to 375 and 2050 Hz and 425 and 1600 Hz, respectively. Formant shifts were also evident for subject SJ, although to a slightly lesser degree. The EMG data dramatically illustrate these effects. Figure 1 shows the genioglossus EMG for the spontaneous bite block and TMJ + topical anesthesia + bite block conditions for the vowel /i/. The top trace shows the genioglossus muscle activity for /i/ produced spontaneously. The middle trace shows the corresponding EMG for the simple bite block condition. The increase in activity here is expected because the tongue has farther to move from a fixed-open position toward its target. Note also that the increase in activity is present at onset, before any online feedback mechanism would have time to generate an adjusted movement. The lower record corresponds to the TMJ + topical anesthesia + bite block condition. It shows virtually no activity. Absence of muscle activity was the rule for all tokens within this condition as well as for the TMJ + topical anesthesia + bite block + artificial palate condition. Apparently, fixation of both the efferent and afferent variables resulted in an inability to produce any coordinated movement; hence, a neutral tongue position and a tendency toward schwa.

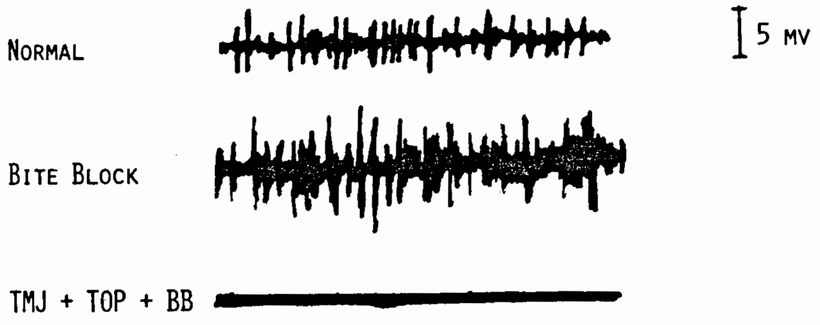

Figure 1

EMG activity for three experimental conditions.

This motor disorganization, however, was relatively short-lived. In each case learning took place and the normal vowel targets were reached after several trials. Table 1 shows the formant frequency values for /i/ produced under the most extreme experimental condition for each of nine token repetitions, for subject WE. Again, measurements were made at the time of the first glottal pulse. For even this extreme condition, complete acoustic compensation was attained by the sixth trial where vowel targets approached those of the spontaneously produced vowel.

Table 1

First and second formant frequencies for the vowel /i/ produced by subject WE under the most extreme experimental conditions (TMJ + topical anesthesia + bite block + artificial palate). Normal vowel target values are: F1 = 275 Hz, F2 = 2275 Hz.

	TRIALS								
	1	2	3	4	5	6	7	8	9
F1	425	500	475	325	325	275	300	300	275
F2	1600	1700	1900	2050	2150	2175	2225	2225	2250

Conclusions

The main finding of this experiment was that interference with either an efferent or afferent variable alone did not affect the production of isolated vowels; however, simultaneous interference with both efferent and afferent variables seriously altered vowel production. It is our view that these findings demonstrate both the necessity of a generative approach to speech production modeling and the utility of a coordinative structure mechanism for the control of speech movements. First, the experimental conditions produced novel physical and sensory situations that were met with immediate and successful articulatory responses. An open-loop model based on stored experiences cannot explain the success of these responses. Second, from the coordinative structure perspective, the finding that afferent interference does not affect vowel production unless an efferent variable is frozen is consistent with the view that these muscles are functionally linked across efference and afference in such a way that control can be taken over by either system when the other is fixed.

References

Fowler, C.A. (1977): "Timing control in speech production", Indiana University Linguistics Club, Bloomington, Indiana.

Fowler, C.A. and M.T. Turvey (1978): "Skill acquisition: An event approach with special reference to searching for the optimum of a function of several variables", to appear in Information Processing in Motor Control and Learning, G. Stelmach (ed.), New York: Academic Press (in press).

Fowler, C.A., P. Rubin, R.E. Remez and M.T. Turvey (1978): "Implications for speech production of a general theory of action", Language Production, B. Butterworth (ed.), New York: Academic Press.

Lindblom, B., J. Lubker and T. Gay (1978): "Formant frequencies of some fixed-mandible vowels and a model of speech motor programming by predictive simulation", JPh (in press).

MacNeilage, P.F. (1970): "Motor control of serial ordering of speech", Psychological Review 77, 182-196.

Turvey, M.T., R. Shaw and W. Mace (1978): "Issues in the theory of action: Degrees of freedom, coordinative structures and coalitions", in Attention and Performance VII, ed. by J. Requin, Hillsdale, N.J.: Erlbaum.