

Chapter 1

CHARACTERISTICS OF SPEECH AS A MOTOR CONTROL SYSTEM

Vincent L. Gracco
Haskins Laboratories

The structural and functional organization of any biophysical system provides potentially important information on the underlying control structure. For speech, the anatomical and physiological structure of the vocal tract and the apparent functional nature of speech motor actions suggest a characteristic control structure in which the entire vocal tract is the smallest functional unit. Sounds are coded as different relative vocal tract configurations generated from neuromuscular specifications of characteristic articulatory actions. Sensorimotor processes are applied to the entire vocal tract to scale and sequence changes in vocal tract states. Sensorimotor mechanisms are viewed as a means to adjust speech motor output predictively in the face of continuously changing peripheral conditions. An underlying oscillatory process is hypothesized as the basis for sequential speech adjustments in which a centrally-generated rhythm is modulated according to internal (task) requirements and the constantly changing configurational state of the vocal tract.

Speaking is a complex action involving a number of levels of organization and representative processes. At a cognitive level, speaking represents the manipulation of abstract symbols through a synthesis of associative processes expressed through a sophisticated linguistic structure. At a neuromotor level, at least seven articulatory subsystems can be identified (respiratory, laryngeal, pharyngeal, lingual, velar, mandibular,

and labial) which interact to produce coordinated kinematic patterns within a complex and dynamic biomechanical environment. At an acoustic level, characteristic patterns result from complex aerodynamic manipulations of the vocal tract. The cognitive, sensorimotor, and acoustic processes of speech and their interaction are critical components in understanding this uniquely human behavior. As the interface between the nervous system and the acoustic medium for speech production/perception, speech motor processes constitute a direct link between higher level neurophysiological processes and the resulting aerodynamic/acoustic events.

In the following chapter, characteristics of the speech motor control process will be evaluated from a functional perspective emphasizing the structural and functional organization of the vocal tract and the timing characteristics associated with their continuous modulation. In contrast to perspectives which emphasize the large numbers of muscular/kinematic degrees of freedom, the current perspective is one that assumes that the overall vocal tract is the smallest unit of functional behavior. Sounds are encoded according to characteristic vocal tract shapes specified neuromuscularly and modulated through sensorimotor mechanisms to adapt to the constantly changing peripheral environment. Examination of the structural components and their interaction is consistent with this macroscopic organization as are a number of empirical observations. The functional organization is implemented by a limited number of sensorimotor control processes that scale overall vocal tract actions spatiotemporally within a frequency-modulated rhythmic organization characteristic of more automatic, innate motor behaviors.

Structural Properties

In order to describe speech from the perspective of a motor control system, a necessary step is to identify the components of the motor system to determine how their structural properties may reflect on the overall functional organization. The structures of the vocal tract include the lungs, larynx, pharynx, tongue, lips, jaw, and velum. Anatomically the vocal tract structures display unique muscular architecture, muscular connections, and muscular orientation that determine their potential contributions to the speech production process. For example, the orientation of the muscles of the pharynx, primarily the pharyngeal constrictors, is such that they generate a sphincteric action on the long axis of the vocal tract producing a change in the cross-sectional area and the tension or compliance of the pharyngeal tissues. The muscles of the velum are oriented primarily to raise and lower the soft palate separating the oral and nasal cavities. Perioral muscles are arranged such that various synergistic muscle actions

result in a number of characteristic movements such as opening and closing of the oral cavity and protruding and retracting the lips. Some of the components, such as the tongue and larynx, can be subdivided into extrinsic and intrinsic portions each of which appear to be involved in different functional actions. Intrinsic tongue muscle fibers are oriented to allow fine grooving of the longitudinal axis of the tongue and tongue tip and lateral adjustments characteristic of liquid and continuant sounds. Extrinsic tongue muscles are arranged predominantly to allow shaping of the tongue mass as well as elevation, depression, and retraction of portions of the tongue. Intrinsic laryngeal muscles are arranged to open and close the glottis reciprocally and adjust the tension of the vibrating vocal folds, whereas extrinsic laryngeal muscles are oriented to displace the entire laryngeal complex (thyroid cartilage and associated intrinsic muscles and ligaments). Generally, movements of the vocal tract can be classified into two major categories; those that produce and release constrictions (valving) and those that modulate the shape or geometry of the vocal tract. The valving and shaping actions are generally associated with the production of consonant and vowel sounds, respectively (Öhman, 1966; Perkell, 1969).

In addition to the structural arrangement of the vocal tract muscles for valving and shaping actions, mechanical properties of individual vocal tract structures provide insight into the functional organization of the speech motor control system. The dynamic nature of the tissue load against which the different vocal tract muscles contract is extremely heterogeneous. For some structures such as the lips and vocal folds, inertial considerations are minimal, while for the jaw and respiratory structures inertia is a significant consideration. The tongue and lips are soft tissue structures that undergo substantial viscoelastic deformation during speech while the jaw and perhaps the lips display a degree of anisotropic tension (Lynn & Yemm, 1971). Even seemingly homogeneous structures, such as the upper and lower lips, display different stiffness properties (Ho, Azar, Weinstein, & Bowley, 1982), possibly contributing to their differential movement patterns (Kelso, Vatikiotis-Bateson, Saltzman, & Kay, 1985; Gracco & Abbs, 1986; Gracco, 1988). Considering the structural arrangement of the vocal tract, the different muscular orientations and the vast interconnection of muscles, cartilages, and ligaments, it is clear that complex biomechanical interactions among structures are the rule. Passive or reactive changes in the vocal tract due to inherent mechanical coupling is a consequence of almost any vocal tract action, with the relative significance varying according to the specific structural components and conformational change and the speed at which adjustments occur. As a result, a single articulatory action may generate primary as well as secondary effects throughout the vocal tract. Examination of individual articulatory actions is important to determine their contribution to the

sound-producing process. However, individual articulatory actions never have isolated effects. The combination of the viscoelastic properties of the tissues, the different biomechanical properties of vocal tract structures, and the complex geometry of the vocal tract comprise a complex biomechanical environment. The kinematic and acoustic variability characteristic of speech production reflects in part the differential filtering of neural control signals by the peripheral biomechanics. Only through detailed biophysical models of the vocal tract and considerations of potential biomechanical interaction associated with various phonetic environments can the control principles of the speech motor control system be separated from structural or cognitive/linguistic influences.

Functional Organization

In order to characterize the speech motor control system accurately, and pose the motor control problem correctly, it is important to determine how the behavior is being regulated. That is, are the individual sound-influencing elements being independently controlled, or does the control structure involve larger units of behavior, and if so, what is the organizational structure? For speech, the simple observation that even an isolated vowel sound requires activity in respiratory muscles, tension and adduction of the vocal folds, adjustments in the compliance of the oropharyngeal walls, shaping of the tongue, positioning of the jaw, elevation of the velum, and some lip configuration is rather convincing evidence that speech is functionally organized at a level reflecting the overall state of the vocal tract. It is the interaction of all the neuromuscular components that provide each speech sound with its distinct character, not the action of any single component. The often-cited fact that speech production involves over 70 different muscular degrees of freedom, while perhaps anatomically factual, is a functional misrepresentation of the motor control system organization. As early as the birth cry and through the earliest stages of speech development, the infant's vocalizations involve the cooperative action of respiratory, laryngeal, and supralaryngeal muscles to produce sounds. A similar observation can be made for locomotion in that rhythmic stepping and other seemingly functional locomotion-like behaviors can be elicited well before the infant manifests upright walking (Thelen, 1985, 1986). It appears that functional characteristics of many human behaviors are present at birth or very early in the infant's development, suggesting that the "significant functional units of action" (Greene, 1972) may be innate properties of the nervous system. It is suggested that speech motor development reflects the ability to make finer and more varied adjustments of the vocal tract, not the mastering of the articulatory or muscular degrees of freedom.

As suggested above, the characteristics of speech as a motor control system include a control structure in which the smallest functional unit is the entire vocal tract. Recent studies have demonstrated examples of large scale manipulation of vocal tract actions rather than the modulation of separate articulatory actions. As shown in Figure 1, movements of individual articulators such as the upper lip, lower lip, and jaw demonstrate timing relations such that adjustments in one structure are accompanied by adjustments in all functionally-related structures. The coordinative process reflects a constraint on articulatory actions involved in the production of a specific sound. Similar results can be observed for other more spatially remote, but functionally related articulators.

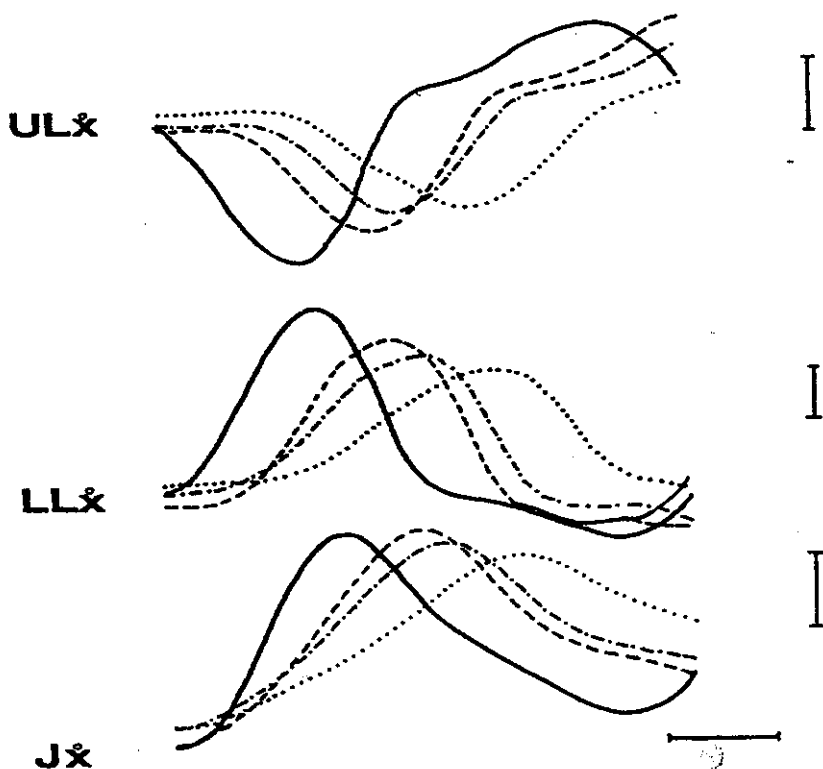


Figure 1. Upper Lip (UL), Lower Lip (LL), and Jaw (J) movement velocities associated with the first 'p' closing in 'sapapple'. Signals are aligned to the jaw opening peak velocity for the first 'a' in 'sapapple'. As the preceding vowel duration changes, the timing of the UL, LL, and J change in a consistent and unitary manner (from Gracco, 1988). Calibration bars are 50 mm/sec (vertical) and 100 ms (horizontal).

As shown in Figure 2, movements of the larynx and the lower lip demonstrate a similar timing dependency for the production of the 'f' in 'safety'. In order to generate the frication noise characteristic of the /f/, the glottal opening and labial constriction is appropriately timed. As the timing of one structure changes, the timing of the other functionally-related articulatory action also changes.

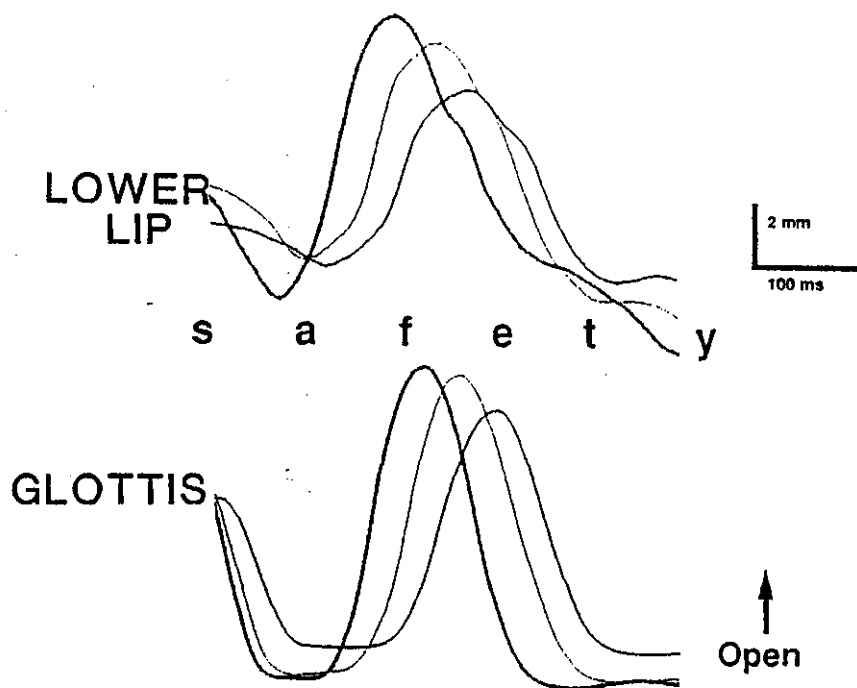


Figure 2. Lower lip closing and glottal opening for three repetitions of the word 'safety' produced in a carrier sentence "Its a _____ again"; signals were aligned to the peak glottal opening for the 'ts' in 'Its'. As the lower lip closing movement for 'f' varies, the timing of the glottal opening (devoicing) also varies (from Gracco & Löfqvist, 1989). Similar to Figure 1, the timing of the oral and laryngeal actions appears to be adjusted as a unit.

Similarly, for movements associated with resonance-producing vowel events, timing constraints can be observed between laryngeal voicing and jaw opening associated with tongue positioning for a vowel (Figure 3). Here, the laryngeal action associated with phonation and the change in jaw positioning to assist the tongue in vowel production demonstrate similar coordinative interdependency.

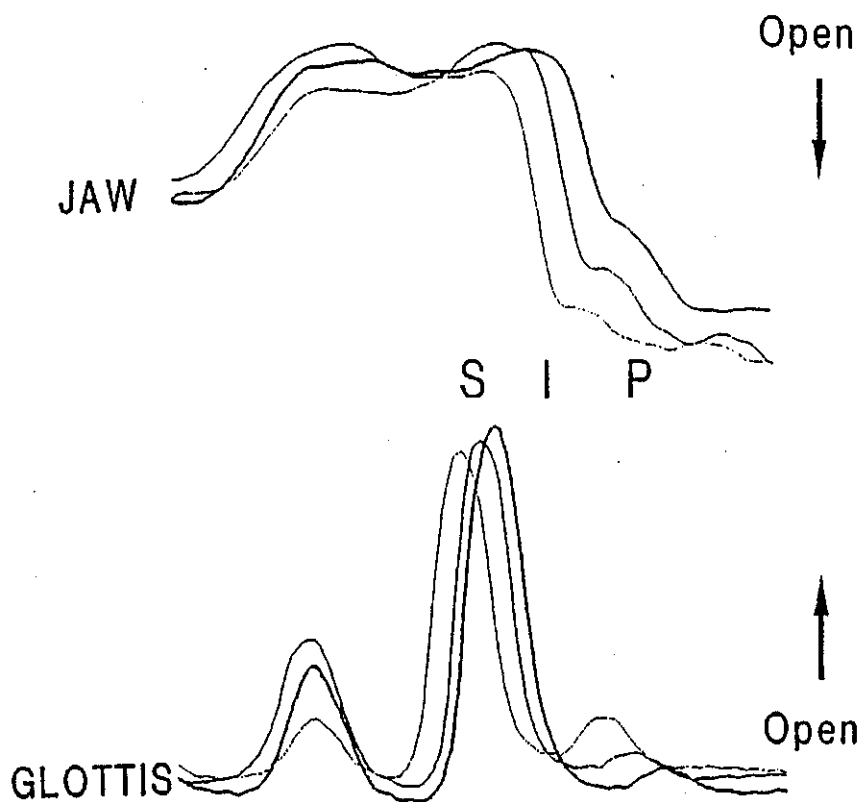


Figure 3. Timing relations between the glottal closing and the jaw opening associated with the vowel in 'sip'. As the glottal opening/closing associated with the 's' and subsequent vowel varies, the jaw opening (noted by the downward movement) also varies proportionally (from Gracco & Löfqvist, 1989). The signals are aligned as in the previous figure.

Some preliminary evidence further suggests that certain physiological changes associated with the production of emphatic stress result in an increase in the actions of all portions of the vocal tract rather than being focused on one specific articulator (Fowler, Gracco, & Vatikotis-Bateson, 1989). In the presence of a potentially disruptive mechanical disturbance applied to one of the contributing articulators there is a tendency for the timing of all articulators to readjust (Gracco & Abbs, 1988). The timing of individual articulators is apparently not adjusted singularly but reflects a system level organization (see Löfqvist & Yoshioka, 1981, 1984; Tuller,

Kelso, & Harris, 1982, for other examples). It is not clear how general these observations are with regard to all speech sounds in all possible contexts. For example, the lip/jaw and laryngeal/supralaryngeal coordination observed in Figures 1 and 2 is modified when the sound is at the beginning of a word, apparently reflecting a change in the functional requirements of the task. The importance of these kinds of observations is not the specific observable pattern but the presence of characteristic patterns that are used for time-dependent articulatory adjustments.

Speech motor patterns reflect characteristic ways of manipulating the vocal tract, in the presence of a constant pressure source, to generate recognizable and language-specific acoustic signals (Ohala, 1983). The process through which such functional cooperation occurs has been described for many motor tasks in various contexts, with the assumption that the control actions involve the assembly of functional units of the system organized into a larger systems known as synergies or coordinative structures (Bernstein, 1967; Gelfand, Gurfinkel, Tsetlin, & Shik, 1971; Fowler, 1977; Fowler, Rubin, Remez, & Turvey, 1980; Turvey, 1977; Kugler, Kelso, & Turvey, 1980, 1982; Saltzman, 1979, 1986). In keeping with the interactive structural configuration outlined previously and the apparent functional nature of the task itself, a modification of this perspective is offered. Speaking appears to involve coordinative structures (or synonymously motor programs; see Abbs, Gracco, & Cole, 1984; Gracco, 1987) available for all characteristic vocal tract actions associated with the sound inventory of the language. It is not the case, however, that a coordinative structure or a motor program is a process but a set of sensorimotor specifications identifying the relative contribution of the vocal tract structures to the overall vocal tract configuration (see Abbs et al., 1984; Gracco, 1987). As such, coordinative structures may be more rigidly-specified than previously thought and the distinction between a flexible coordinative structure and a hard-wired motor program algorithm may be more rhetorical than real (cf. Kelso, 1986, for discussion of differences). In this regard, two observations are of note. When the contribution of jaw movement is eliminated, by placing a block between the teeth, jaw-closing muscle actions are still present (Folkins & Zimmermann, 1981). Further, in response to jaw perturbation, both functionally-specific responses and non-functional responses are observed, such as upper lip muscle increases when the subjects are not producing sounds requiring upper lip movement (Kelso, Tuller, Vatikiotis-Bateson, & Fowler, 1984; Shaiman, 1989). Together, these observations reflect on specific aspects of the speech motor control process and suggest that speech production may rely to some degree on fixed neuromuscular specifications. The presence of jaw muscle actions when jaw movement is eliminated is consistent with the previous suggestion that speech motor control is a

holistic process involving the entire vocal tract. The presence of upper lip muscle increases (albeit small) when the sound being produced does not involve the upper lip, reflects on the underlying control process. The interaction of the phasic stimulus (from the perturbation) with activated motoneurons will produce the functionally-specific compensatory response. If the motoneurons are inactive, or slightly active, the phasic stimuli would result in small increases in muscle activation levels without any significant movement changes. This is a much simpler control scheme in that certain interactions and functionally-specific responses are a consequence of the activation of specific muscles and the actual synaptic interactions in the neural system controlling various vocal tract structures (Gracco, 1987). The advantage of this perspective is that certain properties of speech production result from the physiological organization and focus the functional organization of the speech motor control system on the neural coding of speech sounds and the characteristic sensorimotor processes that modulate and sequence vocal tract configurations.

Neural Coding of Speech Motor Actions

The coding of speech is viewed as the process by which overall vocal tract states are "represented and transformed by the nervous system" (see Perkel & Bullock, 1968). This coding is similar to what has previously been identified as the selection of muscular components associated with a specific motor act (cf. Evarts, Bizzi, Burke, DeLong, & Thach, 1972). In the following, the selection of characteristic vocal tract states will be evaluated with respect to two components of the hypothetical specification process, although the actual neural coding is viewed as a single process and is only presented separately for the purpose of clarity. As stated previously, the actions of the vocal tract are designed to either valve the air stream for different consonant sounds or to shape the geometry of the vocal tract for different vowel and vowel-like sounds. Considering the place of articulation for vowels and consonants naturally results in categorical distinctions which are apparent acoustically and aerodynamically (Stevens, 1972). However, rather than dichotomizing these apparently discrepant processes, it is suggested that valving and shaping can be conceptualized as a single physiological process. That is, speech sounds are coded according to overall vocal tract states which include primary articulatory synergies. When the appropriate muscles are activated, the resulting force vectors create characteristic actions resulting in vocal tract states which act to valve the pressure or change the geometry without creating turbulence producing constrictions. It is the orientation of the activated muscle fibers, the activation of synergistic and antagonistic muscles, and the fixed boundaries of the vocal tract (the immobile maxilla) that result in the

achievement of characteristic shapes or constriction locations; certain muscular synergies can only result in certain vocal tract configurations. For example, selection of certain upper and lower lip muscles (orbicularis oris inferior and superior, depressor anguli oris, mentalis, depressor labii inferior) will always result in the approximation of the upper and lower lips for 'p', 'b', or 'm'. The magnitude or timing of the individual muscle actions may vary, but bilabial closure will always involve the activation of upper and lower lip muscles; otherwise bilabial closure could not be attained. Similarly, changing the focus of neural activation to regions representing lower lip muscles (orbicularis oris inferior and mentalis, with primary focus in mentalis) results in movements consistent with labiodental constriction for 'f' and 'v' achieved against the immobile maxillary incisors (Folkins, 1976). Different relative contributions of extrinsic and intrinsic tongue muscles result in various shapes and movements on the tongue tip, blade, and body, resulting in characteristic constrictions or shapes as a consequence. Constriction location and constriction degree are useful categories to describe different speech sounds because they specify what is distinctive to each phonetic segment. Control over the vocal tract configuration through the development of finer control over the neuromuscular organization provides a more reasonable description of the speech acquisition process because the entire vocal tract is manipulated, not just the distinctive attributes for each sound. The neuromotor differences in consonant and vowel sounds appear to be reflected in other characteristics of the control process.

One such characteristic involves the compliant states of the vocal tract consistent with the level of tension in the tissue walls. The importance of tissue compliance can be inferred from a number of observations. A major physical difference between voiced and voiceless consonants is in the level of air pressure associated with their production. Voiceless sounds are generally produced with higher vocal tract pressures than their voiced counterparts. The pressure difference, which has significant aerodynamic and acoustic consequences, results from changes in the tension in the pharyngeal and oral cavities as well as from pressure from the lungs (Müller & Brown, 1980). For example, subjects engaged in producing speech while simultaneously engaged in a valsalva maneuver (forceful closing of the glottis thereby eliminating the lung contribution) were able to maintain voiced/voiceless intraoral pressure differences apparently resulting from changes in the overall compliance of the vocal tract walls (Brown & McGlone, 1979). Together with experimental evidence that kinematic and electromyographic characteristics of lip and jaw movements are insufficient to differentiate voiced and voiceless sounds (Lubker & Parris, 1970; Harris, Lysaught, & Schvey, 1965; Fromkin, 1966), it appears that a major factor in generating voicing and voicelessness is the

specification of overall vocal tract compliance. Two possible compliant states of the vocal tract are sufficient to categorize most speech sounds; low compliance associated with voiceless consonants and high compliance associated with voiced consonants and vowels. Compliant states of the vocal tract are associated with gross changes in the activity of at least the pharyngeal constrictors (Minifie, Abbs, Tarlow, & Kwaterski, 1974; Perlman, Luschei, & DuMond, 1989) and possibly other portions of the walls of the vocal tract (intraoral cavity). The specification of low compliance (resulting in high vocal tract pressures) would be associated with increased activity in laryngeal muscles to assist in the devoicing gesture, and high compliance (resulting in low vocal tract pressures) would be associated with a relaxation of the muscle activity in the pharyngeal and oral cavities to allow cavity expansion for voiced stops and continuants (Bell-Berti & Hirose, 1973; Westbury, 1983; Perkell, 1969). Certain tense vowels may result from an intermediate level of compliance (between high and low) such that voicing is maintained but overall compliance is slightly higher than for lax vowels. It is important to note that modification in compliance is a process that produces a relatively slow change in the state of the vocal tract, with relaxation (high compliance) a slower process than constriction (low compliance). Together, specification of the compliant state of the vocal tract and selection of specific muscular actions is one means by which the vocal tract states may be specified neurally.

It should be noted, however, that the coding of speech motor actions is viewed primarily as a static process in which characteristic states of the vocal tract are identified prior to their actual implementation. Considering some dynamic properties of the speech motor control system provides some insight into the manner in which different sounds may acquire further acoustic and kinematic distinction. For example, lip closing movement associated with the voiceless bilabial stop 'p' is generally but not invariably associated with a higher velocity than the voiced bilabial 'b' or 'm' (Chen, 1970; Gracco, 1990; Summers, 1987; Sussman, MacNeilage, & Hanson, 1973). Lip and jaw closing movements are initiated earlier relative to vowel onset for voiceless 'p' than for voiced 'b' or 'm' (Gracco, 1990) resulting in shorter vowel durations. One possible explanation is that voiceless sounds are produced at a higher rate or frequency than their voiced counterparts, reflecting a different underlying frequency specification. Movement frequency is one dimension along which different speech sounds can be generally categorized. This hypothetical frequency modulation can be integrated with another dynamic property of the control system. Not only are closing movements generally faster for a voiceless than for a voiced consonant, but the preceding opening movement has also been observed to be faster (Gracco, 1990; Summers, 1987). It appears that not only may sounds be coded as a function of the frequency of individual

vocal tract adjustments but that the functional requirements for specific sounds may be distributed across movement cycles rather than focused on a single movement phase. This observation suggests the operation of a look-ahead mechanism (Henke, 1966) similar to or identical with the mechanism underlying anticipatory coarticulation which predictively adjusts vocal tract actions. Speech motor control is a dynamic neuromotor process in which overall vocal tract compliance, the location of primary valving or shaping synergies, and frequency-modulated motor commands are specified by the immediate and future acoustic/aerodynamic requirements.

Invariance, Redundancy, and Precision

Before presenting some of the specific processes of the speech motor control system that are used to modulate overall vocal tract organization, two important and related issues should be addressed: invariance and precision. The search for invariance has a long and generally unsuccessful history in investigations of speech production with the obvious conclusion that invariance is not a directly observable event (alternatively, the appropriate metric has not been identified). From the perspective of speech as a motor control system, a more fundamental issue is the precision with which any quantity, variable, or vocal tract configuration is regulated. The presence of substantial acoustic, kinematic, electromyographic, and aerodynamic variability suggests that the speech motor control process operates at less than maximal precision (or within rather broad tolerance limits). The achievement of characteristic vocal tract configurations or individual articulatory actions is accomplished by a synthesis of general activation of most vocal tract structures (setting of overall vocal tract compliance) and focused activation of the relevant muscular synergies. This is consistent with neurophysiological evidence demonstrated in the studies of Kots (1975), in which voluntary movement is seen as a synthesis of diffuse excitation (pretuning), a more fixed and discrete increase in motoneuron excitability (tuning), and the final 'triggering' process. Similarly, brain potentials prior to the onset of muscle activity display rather diffuse activation over multiple cortical areas for discrete finger and toe movements (Boschert, Hink, & Deecke, 1983; Deecke, Scheid, & Kornhuber, 1969) and involve larger regions for production of speech (Larsen, Skinhøj, & Lassen, 1978; Curry, Peters, & Weinberg, 1978). One plausible perspective is that the nervous system modulates the focus of primary activation but that this process is not punctate. That is, activation and deactivation of cortical and perhaps subcortical cells involves diffuse and long term changes resulting in distributed tonic and phasic muscle activity. Specification of vocal tract configurations for specific sounds may

involve characteristic patterns of activation and inhibition in all vocal tract muscles with only slightly greater focus on critical articulators involved in the more dominant or sound-critical movements. In some cases muscles may be partially activated just because of the proximity of their motoneurons to other activated motoneurons. One conclusion is that the neural processes underlying speech motor control are broadly specified and that the functional speech production goals (and the requisite perceptual properties) are only categorically invariant. As suggested by the apparent quantal nature of speech (Stevens, 1972), as long as the articulatory patterns are within a certain range (have not made a category change), the corresponding phonetic properties will be perceived, with kinematic variations producing very little perceptual effect. Perhaps speech perception and production should be appropriately represented as stochastic processes based on probability statements implemented through an adequate but imprecise control system. Strict determinism, invariance, and precision are most likely relegated to man-made machines working under rigid tolerance limits or simplified specifications, not to complex biological systems.

Sensorimotor Control Processes

Similar to the temporal organization for speech, spatial interactions are evident that reflect multiarticulate manipulations to achieve characteristic vocal tract states. The clearest examples of cooperative and functionally-relevant spatial interactions are observed when one articulator, such as the lip or jaw, is disturbed during speaking. Following the application of a dynamic perturbation impeding the articulatory movement, a compensatory adjustment is observed in the articulator being perturbed as well as other functionally-related, spatially-distant articulators (Abbs & Gracco, 1984; Folkins & Abbs, 1975; Gracco & Abbs, 1988; Kelso et al., 1984; Shaiman, 1989), reflecting the presence of afferent-dependent mechanisms in the control of speech movements. The distributed compensatory response to external perturbations is a direct reflection of the overall functional organization of the speech motor control process and is comparable to other sensorimotor actions observed for other motor behaviors such as postural adjustments (Marsden, Merton, & Morton, 1981; Nashner & Cordo, 1981; Nashner, Woollacott, & Tuma, 1979), eye-head interactions (Bizzi, Kalli, & Tagliasco, 1971; Morasso, Bizzi, & Dichgans, 1973), wrist-thumb actions (Traub, Rothwell, & Marsden, 1980), and thumb-finger coordination (Cole, Gracco, & Abbs, 1984). Changing the size of the oral cavity with the placement of a block between the teeth similarly results in compensatory changes in articulatory actions resulting in perceptually-acceptable vowel sounds (Lindblom, Lubker, & Gay, 1979;

Fowler & Turvey, 1980). It appears that the speech motor control system is designed to achieve functional behaviors through interaction of ascending sensory signals with descending motor commands.

Human and nonhuman studies have shown that sensory receptors located throughout the vocal tract are sufficient to provide a range of dynamic and static information which can be used to signal position and speed of physiological structures on a movement-to-movement basis (see Munger & Halata, 1983; Dubner, Sessle, & Storey, 1978; Kubota, Nakamura, & Schumacher, 1980; and Landgren & Olsson, 1982, for reviews). Studies utilizing perturbation of speech motor output indicate that the rich supply of orofacial somatic sensory afferents have the requisite properties to interact with central motor operations to yield the flexible speech motor patterns associated with oral communication (Abbs & Gracco, 1984; Gracco & Abbs, 1985; Gracco & Abbs, 1988; Kelso et al., 1984). Because of the constantly changing peripheral conditions during speaking, the absolute position of vocal tract structures can vary widely depending on the surrounding phonetic environment. The speech motor control system apparently adjusts for these movement-to-movement variations by incorporating somatic sensory information from the various muscle and mechanoreceptors located throughout the vocal tract. Considerations outlined elsewhere (Gracco & Abbs, 1987; Gracco, 1987) suggest that the speech motor control system appears to use somatic sensory information in two distinct ways; in a comparative manner to feed back information on the attainment of a speech goal and to predictively parameterize or adjust upcoming control actions. Structurally, there is strong evidence for the interaction of sensory information from receptors located within the vocal tract with speech motor output at many if not all levels of the neuraxis (see Gracco, 1987, and Gracco & Abbs, 1987, for a summary of the vocal tract representation in multiple cortical and subcortical sensory and motor regions). Furthermore, brain stem organization, evidenced by reflex studies, demonstrates a range of complex interactions in which sensory input from one structure such as the jaw or face is potentially able to modify motor output from lip and tongue as well as jaw muscles (Bratzlavsky, 1976; Dubner et al., 1978; Smith, Moore, Weber, McFarland, & Moon, 1985; Weber & Smith, 1987). It appears that there are multiple synaptic interactions possible throughout the neural system controlling the vocal tract, with the specific interaction dependent on how the system is actively configured.

Speech motor actions involve the activation or inactivation of various muscles of the vocal tract which are adjusted based on the peripheral conditions and the specific phonetic requirements. An important question related to the neural representation for speech is the character of the underlying activation process for different articulatory actions. A number

of recent studies, evaluating the kinematic characteristics of different articulators, are consistent with a single sensorimotor process to generate a variety of articulatory actions. One method for evaluating the similarity in the underlying representation for multiple speech sounds and their associated movement dynamics is to compare the geometric (normalized) form of velocity profiles. A change in velocity profile shape accompanying experimental manipulation of phonetic context suggests a change in the movement dynamics, and by inference a change in the underlying neural representation. Conversely, a demonstration of trajectory invariance or scalar equivalence for a variety of movements suggests that different movements can be produced from the same underlying dynamics (Atkeson & Hollerbach, 1985; Hollerbach & Flash, 1982). That is, in order to produce movement variations appropriate to peripheral conditions and task requirements, it may be necessary only to scale the parameters of a single underlying dynamic relation; a much simpler task and, by inference, a simpler neural process. For movements of the vocal folds, tongue, lips, and jaw during speech it has been shown that changes in movement duration, and to a lesser extent movement amplitude, reflect a scaling of a base velocity profile (Gracco, 1990; Munhall, Ostry, & Parush, 1985; Ostry & Cooke, 1987; Ostry, Cooke, & Munhall, 1987; Ostry & Munhall, 1985). A scalar relation across a class of speech sounds involving the same articulators maintained for different initial conditions (different vowel contexts) suggests that the neural representation has been maximized and such a representation might reflect a basic component of speech production. That is, all speech movements may involve a simple scaling of a single characteristic dynamic (force-time) relationship (Kelso & Tuller, 1984) with the kinematic variations reflecting the influence of biomechanical and timing specifications. In addition, specification of control signals in terms of dynamics eliminates the need to specify individual movement trajectories since the path taken by any articulator is a consequence of the dynamics rather than being explicitly specified (see Kelso et al., 1985; Saltzman, 1986; Saltzman & Munhall, 1989). The scaling of individual actions appears to be another characteristic process that eliminates the need to store all possible phonetic variations explicitly. Rather, the control process is a scaling of characteristic motor patterns adjusted for endogenous conditions (speaking rate, emphasis, upcoming functional requirements) and the surrounding phonetic environment (sensorimotor adjustments). The classic central-peripheral, motor program-reflex perspectives have given way to more reasonable and realistic issues including when and how sensory information may be used and how the different representations are coded for the generation of all possible speech movements.

Movement Sequencing

A significant characteristic of many motor behaviors such as speech, locomotion, chewing, and typing is the production of sequential movements. Observations that interarticulator timing is not disrupted following perturbation (Gracco & Abbs, 1988), that speech rate can be modulated by changes in sensory input (Gracco & Abbs, 1989), and that perturbation induces minimal changes in speech movement duration (Gracco & Abbs, 1988; Lindblom, Lubker, Gay, Lyberg, Branderal, & Holgren, 1987) are consistent with an underlying oscillatory mechanism for speech. Further, somatic sensory-induced changes in the timing of oral closing action (due to lower lip perturbation) is consistent with an underlying oscillatory process (Gracco & Abbs, 1988; 1989). Qualitative observations of temporal consistency of sequential movements are also consistent with an underlying oscillatory or rhythm generating mechanism. Figure 4 shows 24 superimposed movements of the upper lip, lower lip, and jaw for the sentence "Buy Bobby a Poppy". These repetitions were produced as part of a larger study and were produced at different times during the experiment. The subject produced one repetition per breath and each repetition was produced at a comfortable subject-defined rate. As can be seen, there is a consistency to the repetitions that suggests an underlying periodicity indicative of a rhythmic process.

A few studies, attempting to address the periodicity and apparent rhythmicity of speech, have demonstrated the presence of some form of underlying frequency generating mechanism. Ohala (1975) recorded over 10,000 jaw movements within a 1.5 hour period of oral reading and was able to identify frequencies ranging from 2-6 Hz with significant durational variability. Kelso et al. (1985), using reiterant productions of the syllable 'ba' or 'ma', demonstrated a rather strong periodicity at approximately 5-6 Hz with minimal durational variability. The findings of Kelso et al. (1985) are consistent with an underlying oscillatory process. In contrast, the range of frequencies found by Ohala (1975) may reflect the frequency modulation associated with the sounds of the language, a factor minimized in the Kelso et al. (1985) study. The modulation of frequency, dependent on specific aerodynamic properties of the specific sounds and surrounding articulatory environment, may be a mechanism underlying speech movement sequencing (see also Saltzman & Munhall, 1989, for further discussion of serial dynamics). The fact that the frequency values reported by Kelso et al. (1985) were similar for 'ba' and 'ma' suggest that vowels may be a major factor in determining the local periodicity. However, it is the case that the individual movements or movement cycles are not the same; local frequencies are different depending on the phonetic context.

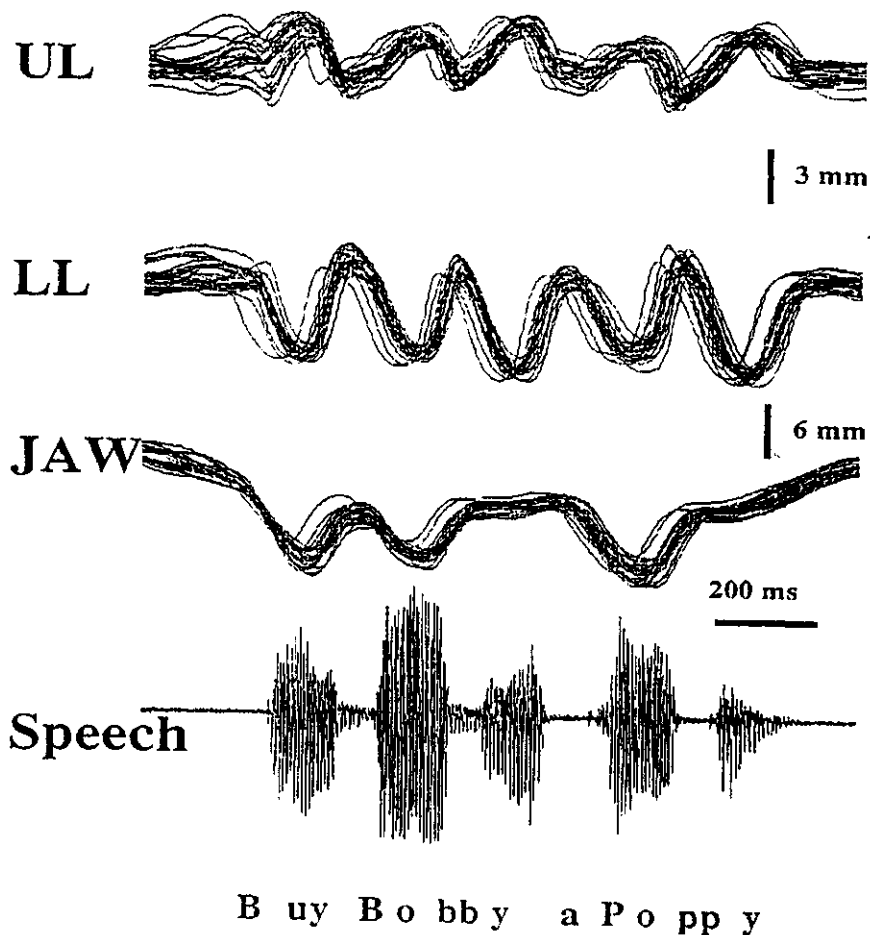


Figure 4. Superimposed Upper Lip (UL), Lower Lip (LL) and Jaw (J) movements associated with 24 repetitions of the sentence "Buy Bobby a Poppy"; the patterns are remarkably similar, displaying little spatiotemporal variation. Only the acoustic signal from a single repetition is shown. The signals are aligned to jaw opening peak velocity for the vowel in 'Buy'.

In addition, speech production involves many of the same muscles as such automatic behaviors as breathing, chewing, sucking, and swallowing. It has been suggested that the mechanisms underlying speech may incorporate, to some degree, the same mechanisms as more automatic motor behaviors but adapted for the specialized function of communication (Evarts, 1982; Gracco & Abbs, 1988; Grillner, 1982; Kelso, Tuller, & Harris,

1983; Lund, Appenteng, & Seguin 1982). Few studies have focused specifically on the similarity of speech with more innate, rhythmic motor behaviors (Ostry & Flanagan, 1989; Moore, Smith, & Ringel, 1988) with mixed interpretations. Recent experiments and theoretical perspectives on the organization of central pattern generators for rhythmic behaviors such as locomotion, respiration, and mastication suggest a more flexible conceptualization of the possible behavioral outputs than has previously been envisioned for the neural control of rhythmic behaviors (see Cohen, Rossignol, and Grillner, 1988, and Getting, 1989, for reviews). For example, *in vitro* results suggest that the central pattern generator for respiration may more appropriately be considered as two separate but interrelated functions; one generating the rhythm and one generating the motor pattern (Feldman, Smith, McCrimmon, Ellenberger, & Speck, 1988). The implication for other rhythmic and quasi-rhythmic behaviors such as speech is that each function can be modulated independently, thus generalizing the concept of a central pattern generator to a wider range of behaviors. Recently, Patla (1988) has suggested that nonlinear conservative oscillators are the most plausible class of biological oscillators to model central pattern generators in that they provide the necessary time-keeping function as well as independent shaping of the output (see also Kelso & Tuller, 1984). The recent demonstration by Moore, Smith, and Ringel (1988) that mandibular muscle actions for speech are fundamentally different than for chewing suggests that the patterning for each behavior is different. That is, speech and chewing may share the same generator but have different patterning or, conversely, rely on different generators and patterns. Conceptually and theoretically, a fundamental frequency oscillator and static nonlinear shaping function can generate a number of complex patterns. Although speculative, some current central pattern generator models have the necessary complexity to be tentatively applied and rigorously tested as to their appropriateness for speech motor control.

Summary

From the present perspective, the speech motor control system is viewed as a biophysical structure with unique configurational characteristics. The structure does not constrain the systems' operation but significantly affects the observable behavior and hence the resulting acoustic manifestations. Consideration of the structural organization and the potential contributions from biomechanical interactions are suggested as potential explanations for some speech motor variability. Sensorimotor mechanisms were implicated as the means by which adjustments in characteristic vocal tract shapes can be dynamically and predictively modified to accommodate the changing peripheral conditions. From the

perspective of the vocal tract as the controlled system, the consistent coordinative timing relationships reflect the functional modification of all the control elements or articulatory structures. Rather than describing sound production as the modulation or assembly of discrete units of action, the current functional perspective suggests that entire vocal tract actions are modulated to regulate acoustic/aerodynamic output parameters. The different parameters are realized by manipulation of the frequency of the forcing function applied uniformly to the control elements of the system. Rather than a parametric forcing in which some parameter such as stiffness is viewed as a regulated variable, it is hypothesized that the system is extrinsically forced by manipulation of the frequency of neural output consistent with the spatial requirements (e.g. movement extent) of the task. The frequency-modulated neuromotor actions are then filtered by the biomechanical environment resulting in intricate kinematic patterns. Speech motor control is viewed as a hierarchically organized control structure in which peripheral somatic sensory information interacts with central motor representations. The control scheme is viewed as hierarchical from the standpoint that the motor adjustments are embedded within a number of levels of organization, reflecting the overall goal of the motor act, communication. Modifications in the control signals reflect the parallel processing of multiple brain regions to scale and sequence changes in overall vocal tract states (Gracco & Abbs, 1987). The organizational characteristics of speech as a motor control system are fundamentally similar to other sequential motor actions and are felt to involve a limited number of general sensorimotor control processes.

Acknowledgements

The author thanks E. Vatikiotis-Bateson and C. Fowler for editorial comments and Y. Manning for word processing. The writing of this paper was supported by NIH grants DC-00121 and DC-00594.

References

- Abbs, J.H., & Gracco, V.L. (1984). Control of complex motor gestures: orofacial muscles responses to load perturbation of the lip during speech. *Journal of Neurophysiology*, 51, 705-723.
- Abbs, J.H., Gracco, V.L., & Cole, K.J. (1984). Control of multimovement coordination: sensorimotor mechanisms in speech motor programming. *Journal of Motor Behavior*, 16, 195-232.
- Atkeson, C.G., & Hollerbach, J.M. (1985). Kinematic features of unrestrained vertical arm movements. *Journal of Neuroscience* 5, 2318-2330.

- Bell-Berti, F., & Hirose, H. (1973). Stop consonant voicing and pharyngeal cavity size. *Journal of the Acoustical Society of America*, 53, 295-315.
- Bernstein, N. (1967). *The co-ordination and regulation of movements*. New York: Pergamon Press.
- Bizzi, E., Kalil, R., & Tagliasco, V. (1971). Eye-head coordination in monkeys: evidence for centrally patterned organization. *Science*, 173, 452-454.
- Boschert, J., Hink, R.F., & Deecke, L. (1983). Finger versus toe movement-related potentials: further evidence for supplementary motor area (SMA) participation prior to voluntary action. *Experimental Brain Research*, 52, 73-80.
- Bratzlavsky, M. (1976). Human brainstem reflexes. In M. Shahani (Ed.), *The motor system: neurophysiology and muscle mechanisms* (pp. 133-154). New York: Elsevier.
- Brown, W.S., & McGlone, R.E. (1979). Supraglottic air pressure during a valsalva maneuver. In J.J. Wolf & D.H. Klatt (Eds.), *Speech communication papers: proceedings of the 97th Meeting of the Acoustical Society of America* (pp. 157-160). New York: The Acoustical Society of America.
- Chen, M. (1970). Vowel length variation as a function of the voicing of the consonant environment. *Phonetica*, 22, 129-159.
- Cohen, A.H., Rossignol, S., & Grillner, S. (1988). *Neural control of rhythmic movements in vertebrates*. New York: Wiley.
- Cole, K.J., Gracco, V.L., & Abbs, J.H. (1984). Autogenic and nonautogenic sensorimotor actions in the control of multiarticulate hand movements. *Experimental Brain Research*, 56, 582-585.
- Curry, S.H., Peters, J.F., & Weinberg, H. (1978). Choice of active electrode site and recording montage as variables affecting CNV amplitude preceding speech. In D.A. Otto (Ed.), *Multidisciplinary perspectives in event-related brain potential research* (pp. 275-279). Washington, DC: US Environmental Protection Agency.
- Deecke, L., Scheid, P., & Kornhuber, H.H. (1969). Distribution of readiness potential, pre-motion positivity, and motor potential of the human cerebral cortex preceding voluntary finger movements. *Experimental Brain Research*, 7, 158-168.
- Dubner, R., Sessle, B.J., & Storey, A.T. (1978). *The neural basis of oral and facial function*. New York: Plenum Press.
- Evarts, E.V. (1982). Analogies between central motor programs for speech and for limb movements. In S. Grillner, B. Lindblom, J. Lubker, & A. Persson (Eds.), *Speech motor control* (pp. 19-41). Oxford: Pergamon Press.