

Enlargement of the supraglottal cavity and its relation to stop consonant voicing

John R. Westbury

Dental Research Center, University of North Carolina, Chapel Hill, North Carolina 27514

(Received 1 July 1982; accepted for publication 6 December 1982)

Measurements were made of sagittal plane movements of the larynx, soft palate, and portions of the tongue, from a high-speed cinefluorographic film of utterances produced by one adult male speaker of American English. These measures were then used to approximate the temporal variations in supraglottal cavity volume during the closures of voiced and voiceless stop consonants. All data were subsequently related to a synchronous acoustic recording of the utterances. Instances of /p,t,k/ were always accompanied by silent closures, and sometimes accompanied by decreases in supraglottal volume. In contrast, instances of /b,d,g/ were always accompanied both by significant intervals of vocal fold vibration during closure, and relatively large increases in supraglottal volume. However, the magnitudes of volume increments during the voiced stops, and the means by which those increments were achieved, differed considerably across place of articulation and phonetic environment. These results are discussed in the context of a well-known model of the breath-stream control mechanism, and their relevance for a general theory of speech motor control is considered.

PACS numbers: 43.70.Bk, 43.70.Ve

INTRODUCTION

According to the myoelastic-aerodynamic theory of phonation (van den Berg, 1958), the vocal folds will oscillate when they are properly adducted and tensed, and when a sufficient transglottal pressure gradient (and thereby, air-flow) is present. During a vowel or continuant consonant—when thoracic mechanisms supply a relatively constant driving pressure while the vocal tract is vented to atmosphere—the second condition is virtually always met, so that the presence or absence of voicing depends largely upon the degree of vocal fold approximation and stiffness. But, during a stop consonant, complete closure is made at some point above the glottis, blocking flow of air out of the mouth and causing the pressure gradient across the glottis and airflow through it to decrease. How is it, then, that the vocal folds can continue vibrating through an entire stop closure, as is common for intervocalic /b,d,g/ in American English, even though one of the conditions necessary for their vibration is not obviously met?

An answer to this question can be derived from consideration of a simple model of the articulatory mechanism. As a first approximation, we can suppose that the vocal tract consists of two soft-walled cavities—the lungs and mouth—separated from each other by a constriction formed by the vocal folds, and from the atmosphere by constrictions formed by the soft palate and posterior pharyngeal wall, and/or by the lips or tongue and palate. Essential aspects of this approximation are represented in terms of the equivalent network (adapted from Rothenberg, 1968) shown in Fig. 1, wherein voltage and current can be considered analogous to the acoustic quantities pressure and volume velocity. The elements of the network and their articulatory interpretation are as follows: the voltage source E_s represents the net (inspiratory or expiratory) pressure generated by the respiratory musculature; C_s and C_0 represent the respective acoustic

compliances of air volumes below and above the glottis; R_l, L_l, C_l and R_w, L_w, C_w represent the lumped viscous-mass-compliant character of walls surrounding the subglottal and supraglottal cavities; the resistances R_g, R_n , and R_a represent viscous and turbulent losses generated by flows through potentially time-variable constrictions at the glottis, velopharyngeal orifice, and mouth opening, while R_s represents some nominal loss generated by airflow through the trachea and bronchi; L_g represents the reactive part of the glottal impedance; and finally, the current source I_0 represents muscularly actuated rate of change of volume of the supraglottal cavity.

For this network, we can write a set of variable-coefficient nonlinear differential equations which describe system response to time changes in resistances (when voltage across at least one branch is not equal to zero), or to time changes in the voltage or current sources themselves. We can approximate this set of equations by difference equations which are easily programmed on a digital computer, thereby deriving a method for simulating time functions of volume velocity and pressure which may result from hypothetical changes in net force of the respiratory muscles, volume of the supraglottal

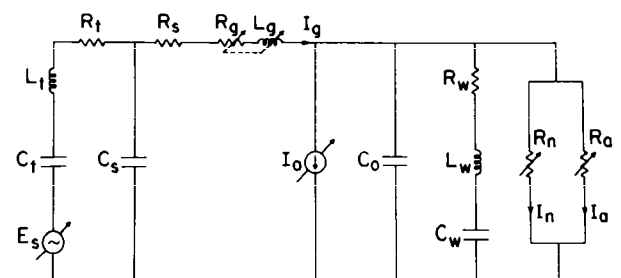


FIG. 1. A circuit representation of the breath-stream control mechanism. Elements of the model are defined in the text.

cavity, or dimensions of articulatory constrictions. Good examples of this approach can be found in work of Flanagan *et al.* (1975) and Müller and Brown (1980).

Given such a model, we can now determine whether and how voicing might continue through the closure interval of a stop consonant. Consider such a segment occurring intervocally, and suppose that the following conditions hold: first, that pressure below the glottis derives entirely from elastic recoil of the stretched tissues which surround the lungs; second, that there are no muscularly induced changes in supraglottal cavity volume over the vowel–consonant–vowel sequence; third, that the tissues surrounding the supraglottal cavity have uniform mechanical properties that approximate those of the cheeks when muscularly tensed (cf. Ishizaka *et al.*, 1975); and fourth, that the degree of muscularly induced vocal fold adduction and stiffness remain constant. If we represent these articulatory conditions by specifying values for appropriate elements in the model (cf. Appendix), we can calculate that pressures above and below the glottis will vary with time as shown in Fig. 2 during a hypothetical labial stop which is held for 80 ms.

Assuming that the vocal folds will continue oscillating as long as the pressure drop across them is greater than 2000 dyn/cm² (cf. Ladefoged, 1964; Lindqvist, 1972; Ishizaka and Matsudaira, 1972; Bäer, 1975), this simulation suggests that voicing might continue for roughly 60 ms following an (intervocalic) occlusion. This interval of voicing is due almost entirely to compliance of tissues surrounding the supraglottal cavity, and closely approximates the not uncommon 65–75-ms closure duration for medial /b/ in American English.

We know, however, that the vocal folds may vibrate throughout a much longer medial closure. A voiced closure as long as 100 ms is certainly not rare. We know too that lengthy closure voicing may occur during stops in other phonetic environments—environments which entail markedly different initial conditions for the respiratory mechanism and/or laryngeal state. For example, voicing may occur dur-

ing stops which are themselves the first (or last) segments in an utterance, when air pressure below the glottis increases from (or decreases to) atmosphere. Moreover, voicing may occur during stops adjacent to other segments which are customarily articulated with an open glottis.

How are these facts possible? Two articulatory adjustments internal to the larynx may prolong voicing under sub-optimal conditions. Tension of the vocal folds may be decreased, thereby increasing their susceptibility to oscillation in low volume velocity flows (cf. Halle and Stevens, 1971); or (average) glottal resistance to flow may be increased by decreasing (average) glottal area (effectively lengthening the relevant time constant of the model). However, there are no physiological data which show that speakers make such adjustments during voiced stops in any phonetic environment. Thus some mechanism(s) external to the larynx must operate to insure a transglottal pressure gradient sufficient for closure voicing during stops articulated under conditions much different from those governing results shown in Fig. 2.

Inspection of the model suggests four such mechanisms. It is possible, first, that the requisite pressure gradient might be maintained if subglottal pressure were increased rapidly and in concert with the supraglottal pressure rise which naturally accompanies vocal tract occlusion. In the model, such a maneuver can be implemented in terms of a positive increase in E_s , and would have the articulatory interpretation of an increase in activity of the expiratory muscles. However, electromyographic evidence for a robust expiratory breath pulse during voiced stops does not exist. Instead, measures of intratracheal pressures (cf. Netsell, 1969; McGlone and Shipp, 1972; Löfqvist, 1975; Westbury and Niimi, 1979) suggest (albeit indirectly) that no great changes occur in the level of activity in the respiratory muscles during the closures of voiced stops. Moreover, thoracic mechanisms seem to function in essentially the same way for voiced and voiceless stops.

A second mechanism which may be used to sustain an adequate transglottal pressure drop during stops is partial lowering of the soft palate and opening of the velopharyngeal port. As early as 1913, Muehold (cited in Stetson, 1950:50) suggested that this mechanism was characteristically used to sustain phonation during German /b,d,g/, though Stetson unceremoniously dismissed the hypothesis as “quite mistaken.” Modern acoustic and cinefluorographic evidence for velopharyngeal opening during some instances of voiced stops, occurring in a variety of environments, has been reported by several authors (cf. Yanagihara and Hyde, 1966; Rothenberg, 1968; Kent and Moll, 1969; Dixit and MacNeilage, 1972). On the other hand, both Lubker (1973) and Westbury (1979) conclude from nasal air-flow data on American speakers of English that velopharyngeal opening is likely not a significant voicing mechanism, at least for stops occurring utterance medially. Nasal flows during medial voiced stops are infrequent (though not unknown), and when measurable, are usually small enough to be attributed to palatal elevation rather than velopharyngeal opening. Furthermore, flows of similar magnitude, independent of their source, are observed as often for voiceless as for voiced stops.

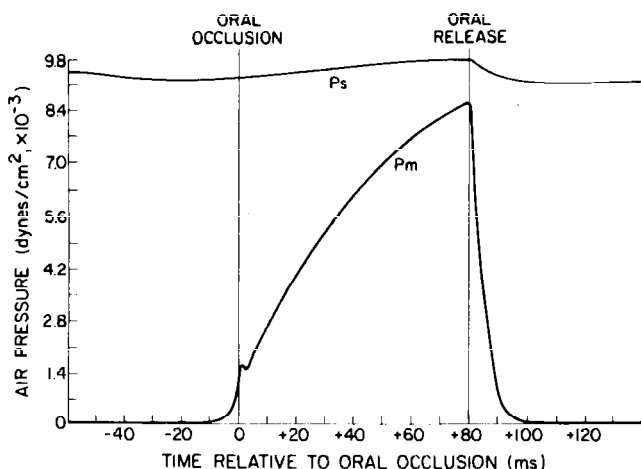


FIG. 2. Synthesized time functions of air pressures below (P_s) and above (P_m) the glottis during an intervocalic labial stop. Walls of the supraglottal cavity are assumed to have mechanical properties similar to those of the cheeks when muscularly tensed.

A third mechanism which will allow on-going transglottal flow during complete upper-tract occlusion is pressure-actuated expansion of the supraglottal cavity (Rothenberg, 1968). The vocal tract walls are not rigid, but compliant, especially at low frequencies, and thus will yield to pressures acting upon them (cf. Fant, 1972; Ishizaka *et al.*, 1975; Fant *et al.*, 1976). It is that fact which is largely responsible for the relatively slow rise in P_m , and thereby, the probable continuation of voicing past occlusion during the stop simulated in Fig. 2.

Compliance of the vocal tract walls depends in part upon (and varies inversely with) the level of activity in the underlying musculature. Consequently, it should be possible to affect voicing duration in stops by inhibiting or recruiting certain of the muscles bounding the vocal tract—e.g., the superior, middle, and inferior pharyngeal constrictors, which bound the pharynx posteriorly and laterally; or the buccinator and risorius, which line the cheeks. Gestures such as these determine the extent to which tissues surrounding the cavity yield to increasing supraglottal pressure, and thus the durations of intervals when closure voicing is possible.

In principle, the (lumped) mechanical impedance of tissues surrounding the supraglottal cavity should be considered time variable. Unfortunately, there are no quantitative data which describe either the relation between different levels of activity in muscles underlying the walls of that cavity and mechanical properties of surrounding tissue, or the temporal variation in activity of those muscles. Thus, the viscous-mass-compliant character of the walls is represented in simple fashion in the model by fixed elements.

Calculation of the time change in pressure gradient across the glottis during a hypothetical labial stop bounded by walls with properties different from those assumed in Fig. 2 are represented in Fig. 3. The dotted function shows the decay in that gradient if the walls are considered rigid, while the lower and upper dashed functions show calculated decays when the surrounding tissues are (uniformly) mechani-

cally analogous to the neck wall, and to the cheeks when muscularly lax, respectively (cf. Ishizaka *et al.*, 1975). The solid function shown in this figure represents the time change in difference between sub- and supraglottal pressures shown in Fig. 2, where the vocal tract walls are considered analogous to tense cheek tissue. The total compliance of the supraglottal cavity walls for this last condition agrees closely with Rothenberg's earlier estimate of average supraglottal compliance, for bilabial and retroflex closures (1968). The points in time following occlusion when the pressure gradient falls below the voicing threshold—i.e., less than 2000 dyn/cm²—are indicated by cross marks on each function. In effect, then, this figure shows that we might expect voicing to continue past occlusion for only 7 ms if the vocal tract walls are rigid, roughly 30 ms if they are "tense," slightly more than 60 ms if they are "moderate," and easily a full 80 ms if they are "lax."

Relying solely on data on wall impedance reported by Ishizaka *et al.* (1975), we can assume the dashed ("tense" and "lax") functions in Fig. 3 to represent practical limits on speakers' ability to manipulate the decay of the transglottal pressure, and thereby, the duration of closure voicing via adjustments in tension of the vocal tract walls.¹ The extent to which this mechanism is used in articulating voiced stops is not well-known. Electromyographic recordings of the superior and middle pharyngeal constrictors from five speakers, reported by Minifie *et al.* (1974), showed less activity, and presumably greater compliance of the overlying surface during intervocalic voiced than voiceless stops. Similarly, recordings from the same muscles and palatopharyngeus and palatoglossus from three additional speakers, reported by Bell-Berti (1975), showed relatively lower peak activity, in some instances and for some speakers, during (medial) voiced than voiceless stops bounded on one side by a vowel and on the other by a nasal consonant.

A fourth voice-sustaining mechanism is muscularly actuated enlargement of the supraglottal cavity. Adding air to a closed cavity will obviously cause pressure within the cavity to rise. If at the same time, cavity volume is increased, pressure within may still rise, but at a slower rate. Thus, actively increasing volume of the supraglottal cavity during a voiced stop—by lowering the larynx, raising the soft palate, advancing the tongue root, or drawing the tongue dorsum and blade down toward the mouth floor—can at least partially accommodate the air volume transmitted to it per glottal pulse, and thus prolong the interval during which voicing is possible. In the model, provision for active cavity volume change is represented by the variable current source I_0 . Cinefluorographic data reported by Perkell (1969) and Kent and Moll (1969),² and electromyographic data reported by Bell-Berti (1975), suggest that cavity-enlarging maneuvers are characteristic of /b,d,g/ produced by some speakers of American English, when those sounds occur between vowels, or when clustered with nasals which are themselves bounded by vowels. However, as Lisker (1977:305) has noted, "whether [these reported] cavity-enlarging maneuvers...are sufficient to account for the durations of observed voiced-closure intervals is still not entirely clear." Moreover, the aforementioned studies do not consider whether and

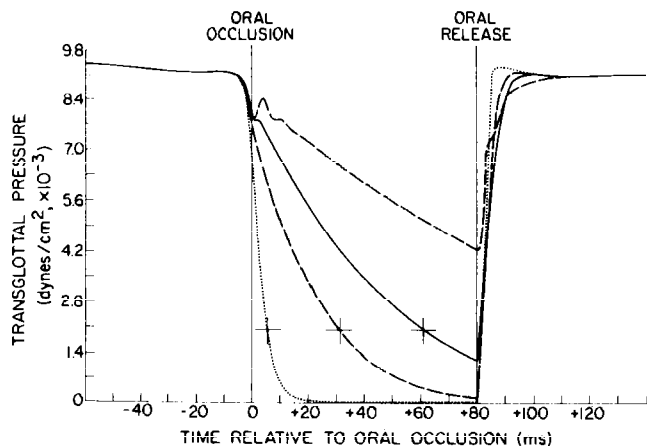


FIG. 3. Synthesized time functions of the transglottal pressure gradient during an intervocalic labial stop bounded by rigid walls (dotted trace), or walls mechanically analogous to the neckwall (lower dashed trace), tensed cheeks (solid trace), or relaxed cheeks (upper dashed trace). Points where the pressure gradient falls below voicing threshold are indicated by crosses.

how cavity enlargement might be used for voiced stops occurring other than utterance-medially—in environments which, by their nature, entail aerodynamic conditions perhaps less conducive to closure voicing than those underlying the medial stop simulation illustrated in Fig. 2. A purpose of this study is to describe new data concerning the nature of active control of supraglottal volume during the production of American-English stops, in several phonetic environments. These data are then discussed in relation to conclusions derived from the model illustrated in Fig. 1, concerning the magnitude of cavity-enlarging maneuvers necessary to maintain voicing during closure.

I. METHOD

An experiment was devised to examine the positioning and extent and rate of movements of certain of the articulators, viewed in a sagittal plane, relative to the incidence and duration of voicing during the production of stop consonants. A cinefluorographic record of the articulation of stops in several phonetic environments was obtained using equipment described by Kent and Moll (1969), operated at a film speed of 100 frames per second, so that continuous speech gestures were sampled at 10-ms intervals. At the same time, an acoustic record of the speech sample was recorded on magnetic tape. The acoustic and cinefluorographic records were synchronized by means of a time code recorded on a second channel of the tape recorder and between sprocket holes of each film frame.

During the course of the experiment, the subject—a normal English-speaking adult male—was seated upright with the head firmly secured by a cephalostat. Immediately prior to filming, the lips and under surface of the tongue were liberally coated with a barium paste solution (Rugar). The paste was subsequently swallowed to aid definition of the soft tissue boundaries of the oral and pharyngeal cavities.

The speech sample for the experiment consisted of isolated nonsense disyllables which were repeated at roughly 1.5-s intervals. The portion of the utterance inventory relevant to this report is shown in Table I. The data described herein are derived from one example of each disyllable. This inventory was designed to satisfy a number of considerations. First, it provided examples of each of the six American English voiced and voiceless stops occurring singly—in utterance-initial, medial, and final positions, always preceded and/or followed by the vowel /a/. Second, it provided examples of each of the three nasal consonants /m,n,ŋ/ occurring singly and bounded on both sides by /a/. Third, it

TABLE I. Test utterances, repeated with primary stress on initial syllables and secondary stress on final ones.

batab	pabpat	panap
gapag	padtat	kamak
dakad	pagkat	tarjat
padap	papbat	pampat
kabak	patdat	papmat
tagat	pakgat	pambat
		pabmat

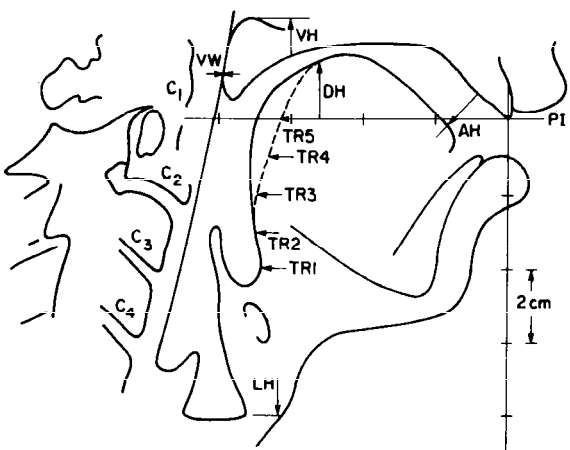


FIG. 4. A sample tracing of one cinefluorographic frame, showing the articulatory configuration preceding closure for medial /g/, and illustrating the measurements made from each frame.

provided examples of the four possible nasal-stop and stop-nasal combinations of /m,p,b/, occurring intervocally and bounded by /a/. And finally, it provided examples of medial stop clusters with mixed voicing which are possible in American English—specifically, /pb/, /bp/, /td/, /dt/, /kg/, and /gk/, all preceded and followed by /a/. In stop clusters such as these, voicing (but not closure) is frequently discontinuous—virtually always so for those whose initial members are nominally voiced and whose final members are nominally voiceless, and sometimes so, depending upon speaker identity, for those whose initial members are voiceless and final members are voiced (cf. Westbury, 1979).

Life-size outlines of the vocal tract, from the superior margin of the vocal folds upward, were hand traced for all individual frames in the cinefluorographic record during orally occluded intervals of nasal and stop consonants. In most utterances, tracings were also made of cavity configurations for one or two frames before and after closure and release of these segments. Many major hard-structure landmarks (e.g., dental inlays in the upper and lower molars, outlines of the maxilla, mandible, hyoid bone body, and cervical vertebrae) were included in each tracing. A sample tracing of one frame may be seen in Fig. 4.

The principal object of synchronous recording of lateral-view cinefluorograms and speech was to monitor changes in positions of speech articulators over time, relative to the incidence and duration of voicing during stop consonant production. Thus the positions of quasi-independent structures such as the soft palate, tongue, and larynx were measured (by hand) relative to a common reference, the maxilla. Inferior-superior measurements of the positions of these articulators were made (to the nearest half-millimeter) along vectors perpendicular to a primary reference line indicated as PI in Fig. 4 and defined by the inferior edges of the central maxillary incisors and a distinctive dental filling in a posterior maxillary bicuspid. Anterior-posterior positions of articulators were measured along vectors parallel to PI, and relative to the maxillary incisors themselves. Thus measurements were made within a simple x-y coordinate system with an origin at the intersection of the upper incisors and PI.

Specific dimensions in the sagittal viewing plane were measured as follows: larynx height (LH), the vertical distance between PI and the anterior end of the laryngeal ventricle; velum height (VH), the vertical distance between PI and a tangent to the superior-most aspect of the dorsal surface of the soft palate; velopharyngeal width (VW), where possible, the distance between the posterior pharyngeal wall and dorsal surface of the soft palate, along a line 12 mm above and parallel to PI; horizontal position of the tongue root (TR 1–5), the distances between the posterior surface of the tongue midline and the maxillary incisors, along vectors parallel to PI and spaced at 1-cm intervals down from that reference, with measurement TR 1 made 4 cm below the reference, and TR 5 along PI itself; dorsal tongue height (DH), displacement of the tongue midline away from PI along a vector intersecting the reference at a 90° angle 52-mm posterior to the maxillary incisors; and anterior tongue height (AH), displacement of the tongue blade/tip away from PI along a vector intersecting the reference at a 45° angle 15-mm posterior to the maxillary incisors.

The original time-by-displacement representations of articulatory positions were smoothed by convolution with a nine-point triangular window, and curves were then hand fitted to the smoothed coordinates. Graphic representations of movements of various articulators are illustrated in Figs. 5–12. In most instances, the vertical lines which intersect movement records shown in those figures correspond to the first frames in the cinefluorographic record showing clear (midline) contact between the tongue and palate, or lips. No attempt was made in data analysis to interpolate the moments of consonantal closure (or release) between film frames. Consequently, actual occlusion (and release) occurred sometime during the 10-ms interval immediately preceding the points in time when those events are indicated.

II. RESULTS

The results of this study are presented in three sections. The first describes qualitatively the temporal control of voicing during stop consonants and nasals occurring singly and in clusters. The second section describes in greater detail the position and movement of individual articulators during the closures of such segments, particularly insofar as they relate to the segmental properties voicing and place and manner of articulation. The third section then considers the cumulative effect of several articulatory movements relative to the temporal control of volume of the supraglottal cavity, and thereby, to aerodynamic conditions which may affect the incidence and duration of voicing during intervals of complete upper-tract occlusion.

A. Voicing control during consonants

All utterances were digitized at a sampling frequency of 10 kHz and displayed in brief sections (of roughly 100 ms) on a computer-driven oscilloscope. Visual inspection of the digitized acoustic waveforms showed, as expected, that voicing was largely absent during the closures of all single voiceless stops, though one or two glottal pulses could sometimes be observed following (judged) occlusion in such segments

occurring utterance-medially and finally. (Voice onset followed release of initial and medial voiceless stops by 30–60 ms.) In contrast, voicing was continuous through the (oral) closures of single medial nasals and voiced stops, and medial clusters containing both /m/ and /b/. Voicing was discontinuous through the closures of medial clusters containing both /m/ and /p/, and medial clusters containing both voiced and voiceless stops. Among the latter, those with voiced first members showed voicing for approximately the initial half of their closure, and voicelessness thereafter, while those with voiced second members were initially voiceless and then voiced prior to second-member release. Voicing was also discontinuous during the closures of single voiced stops occurring utterance-initially and finally. During initial voiced stops, voice onset followed (oral and velopharyngeal) occlusion, but always preceded release by 60–90 ms, invariably continuing unbroken into the following vowel. During final voiced stops, voicing continued unbroken from the preceding vowel through the moment of occlusion, with offset occurring roughly 60 ms after that point in time, and in advance of release.

B. Movements of articulators

1. The larynx

The position of the larynx in the neck, relative to the maxilla, varied over a vertical range of 17 mm, from a high of 73 mm below PI (approximately 70 ms following closure in utterance-final /k/) to a low of 90 mm below that reference (slightly before oral release in utterance-initial /d/). In general, the larynx tended to be relatively low during /a/—83–87 mm below PI—and usually somewhat higher during adjacent consonants. Maximum vertical movement of the larynx within the closure interval of any single consonant or medial cluster was roughly 6.5 mm, observed over a 130-ms interval during medial /pb/.

Representative larynx movements during single medial consonants are shown in Fig. 5, while movements during utterance-initial /d/ and /t/, utterance-final /g/ and /k/, and the medial clusters /dt/ and /td/ are shown in Fig. 6. In most cases, articulatory movements during segments having the same place of articulation are aligned at the first moment of observed (oral) closure. However, movements during utterance-initial stops are aligned at the moment of complete upper-tract occlusion, when the velopharyngeal port was also closed. Oral closure customarily preceded velopharyngeal closure in all initial stops produced by the speaker in this study.

In these data, larynx position seems to be affected by two factors. First, it can be seen that the larynx was generally highest in the neck during velar consonants, lowest during labials, and intermediate during alveolars. This effect of place of articulation held among single consonants occurring medially and finally, and among medial consonant clusters, but not among single initial stops. (Though not illustrated in Fig. 6, larynx position and movements during initial /b,g/ and /p,k/ were very much like those shown for initial /d/ and /t/, respectively.) It is tempting to attribute the relatively greater laryngeal elevation during most velar conson-

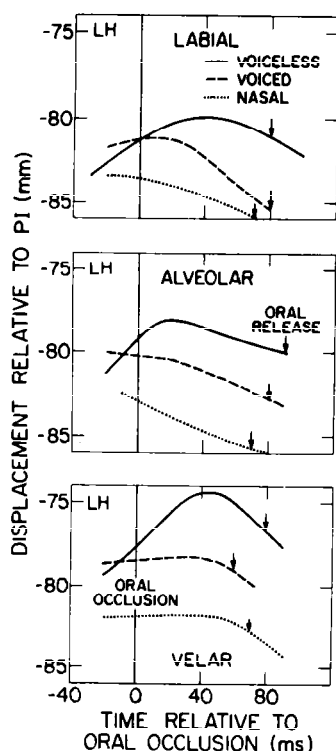


FIG. 5. Vertical movements of the larynx during intervocalic nasals and voiced and voiceless stops.

ants to anatomical connections between the larynx, hyoid bone, and tongue mass, and the functioning of muscle groups which serve to approximate the tongue dorsum with the palate. However, larynx position and movements clearly cannot be *explained* in terms of position and movements of the dorsum. The distance between the two structures is not fixed. Note, for example, that the larynx was lower during medial /g/ than medial /k/, despite the fact that general positioning of the tongue body and its vertical displacement were virtually the same for the two stops. In principle, then, their relative positions may vary. Moreover, their movements may be quite independent. During utterance-initial /g/, the larynx was observed to move downward, away from the maxilla, at the same time that the tongue dorsum moved upward.

A second factor which affects larynx position is consonantal voicing. Consistent with observations reported by Kent and Moll (1969), Perkell (1969), and Ewan and Krones (1974), the larynx was relatively lower during single medial voiced than voiceless stops, and even lower, as noted by Riordan (1980), during comparable nasals. Moreover, the larynx was invariably lower during comparable voiced than voiceless stops occurring utterance-initially and finally. Too, the larynx tended to be lower during voiced than voiceless portions of comparable stop clusters (though not during contrasting portions of clusters containing both a nasal and stop). Considered together, these facts show a clear trend for the larynx to be relatively low during consonants which are voiced and somewhat higher during those that are voiceless.

The direction of larynx movement does not seem to be systematically related to consonantal voicing, manner, or place of articulation. In one respect, this fact is surprising.

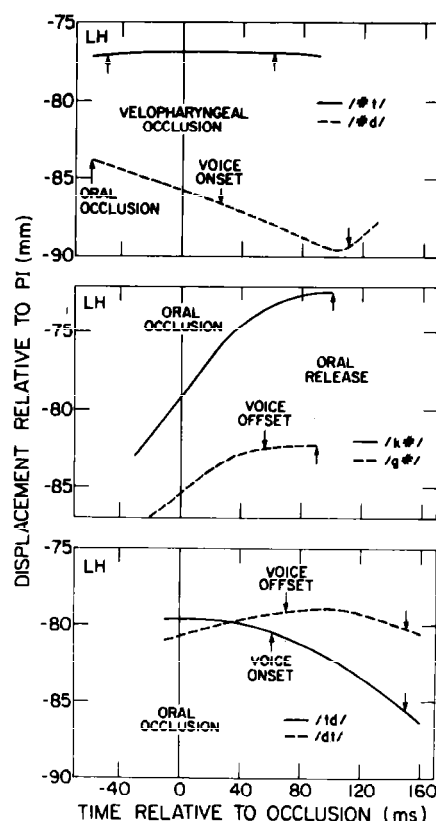


FIG. 6. Vertical movements of the larynx during utterance initial /d,t/, utterance-final /g,k/, and the intervocalic clusters /td,dt/.

All else being constant, larynx depression—because it enlarges cavity volume above the glottis—will facilitate voicing when it occurs during the closure interval of a stop. For this reason, we might expect the larynx to move downward, in a voice-sustaining fashion, during voiced stops. Certainly, the larynx moved steadily downward during the closures of some of the voiced stops observed in the present study (e.g., during utterance-initial, /b,d,g/, medial /b,d/, stop-cluster-final /b,d,g/, and cluster-initial /b/ in /bm/). However, the larynx moved very little during the closures of other voiced stops (e.g., medial /g/, utterance-final /b,d/, cluster-initial /b/ of /bp/, and cluster-final /b/ of /mb/), and even upward in the neck in still other instances (e.g., utterance-final /g/, and cluster-initial /d,g/). Moreover, the larynx sometimes (though less frequently) moved downward during nasals and voiceless stops (e.g., during stop-cluster-final, /p,t,k/ and medial /m,n/). Thus voice-sustaining movements of the larynx were observed during some instances of voiced stops, but such movements were neither characteristic of nor unique to the closures of those segments in all phonetic environments.

2. The tongue root

The pharynx is characteristically narrow during the vowel /a/, measuring only 8–10 mm along TR 2 from its posterior wall to the tongue root midline. During some consonants, however, the pharynx may be considerably wider. Approximately 40 ms prior to release of utterance-initial /d/, for example, pharynx width (along TR 2) measured slightly more than 26 mm. Since the posterior pharyngeal

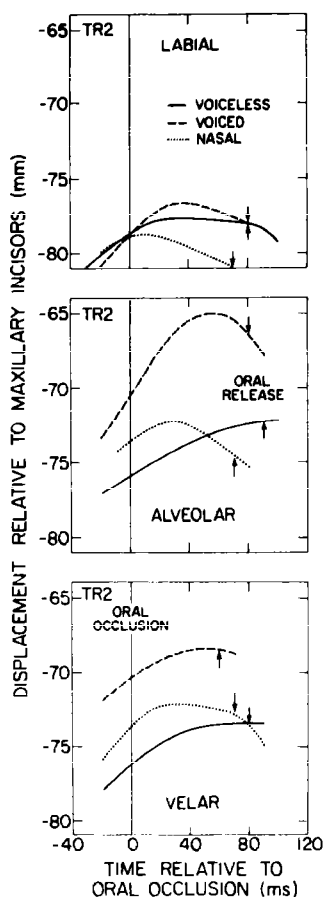


FIG. 7. Horizontal movements of a lower portion of the tongue root during intervocalic nasals and voiced and voiceless stops. Displacement toward the top of each panel corresponds to anterior movement of the tongue root.

wall is largely stationary, changes in pharynx width can be attributed to anterior-posterior movements of the tongue root. In general, such movements into, during, and away from consonantal occlusion were smallest for all places of articulation at measurement levels TR 5 and 4, moderate at level TR 3, and greatest at levels TR 2 and 1. Representative movements of the tongue root (measured along TR 2) during single medial consonants are shown in Fig. 7, while movements during the four utterance-initial stops /p,b,t,d/ and the medial clusters /pm,bm/ are shown in Fig. 8.

Horizontal position of the tongue root, like vertical position of the larynx, appears to be affected both by consonantal place of articulation and voicing. Note foremost from Fig. 7, for example, that the tongue root was most advanced (toward the maxillary incisors) during alveolar consonants, least advanced during labials, and intermediate during velars. The relatively more anterior tongue root position during lingual than labial consonants has also been noted by Perkell (1969), and is probably related to the fact that articulation of a lingual consonant—particularly when following /a/—requires that some portion of the tongue (other than the root) be moved upward and/or forward to form a constriction with the palate. Given that tongue volume must remain constant, the tongue root must shift forward beginning some time before the first moment of occlusion, sometimes continue in that direction for much of the consonant duration, and then reverse direction slightly before or con-

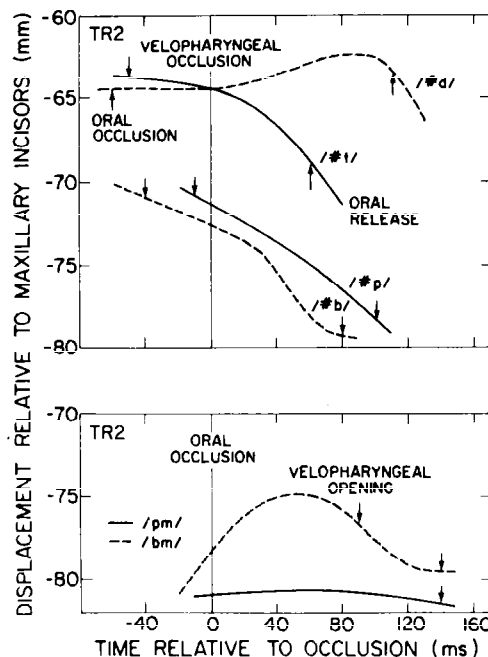


FIG. 8. Horizontal movements of a lower portion of the tongue root during utterance-initial /p,b,t,d/ and the intervocalic clusters /pm,bm/.

comitant with release. During labial consonants, on the other hand, the principal articulatory constriction is formed by the lips. Consequently, there is no need to actively move the tongue about in the mouth, independently of other ongoing movements of the mandible and larynx. The tongue body is therefore freer during labial than lingual consonants to maintain the narrow pharyngeal configuration characteristic of /a/.

Data illustrated in Fig. 7 also show that tongue root position depends upon consonantal voicing. In virtually all instances, the tongue root was most advanced at the first moment of occlusion, and/or advanced more thereafter, for voiced than voiceless stops (cf. Kent and Moll, 1969; Perkell, 1969). Note also from Fig. 7 that tongue root positions during homorganic nasals and voiceless stops were more alike than during nasals and corresponding voiced stops. Together, these facts suggest that a relatively anterior tongue root position is specific to voicing in stop consonants.

The direction of tongue root movement is not systematically affected by consonantal place or manner of articulation, but does appear to depend upon voicing. In virtually every voiced stop, including those occurring singly and clustered with other consonants, the tongue root moved generally forward, in a cavity-enlarging and thereby voice-sustaining fashion, during the portion of the closure when the vocal folds were vibrating. The only exceptions were during /b/ occurring utterance-initially, and cluster-finally in medial /mb/. The tongue root also moved generally forward during the closures of some voiceless stops—notably, medial /t,k/, and utterance-final and stop-cluster-initial /p,t,k/. However, the tongue root remained generally still, or moved rearward during nasals and the remaining voiceless stops. Thus, cavity-enlarging movements of the tongue root were generally characteristic of (though not unique to) voiced stops in all environments.

3. The soft palate

The soft palate is customarily low in the nasopharynx and the velopharyngeal orifice is continuously open during single nasal consonants. Coupling of the oral and nasal cavities occurs as much as 50 ms prior to (oral) closure for these sounds, and usually persists well after their release. In contrast, the soft palate is relatively higher in the nasopharynx and generally in contact with the posterior pharyngeal wall during the closures of non-nasal stop consonants. No instances were observed in which the velopharyngeal port was not occluded during stop closure intervals when the vocal folds were vibrating. Thus incomplete velopharyngeal closure does not seem to have been used as a voicing-maintenance mechanism by the speaker in this study. Representative movements of the soft palate during single medial consonants are shown in Fig. 9, and are accompanied below by measures of velopharyngeal width for /m,n,ŋ/.

Neither the position of the soft palate nor its direction of movement were affected by consonantal voicing or place of articulation among single stops. In particular, the dorsal surface of the palate was equally high at closure and release, and elevated equally during closure, for voiced and voiceless cognates in all environments. This result is consistent with reports of Moll (1962) and Ushijima and Sawashima (1972), but in contrast with reports of Lubker *et al.* (1970) and Künzel (1979), who note greater palatal elevation during voiceless stops, and Bell-Berti (1975), who published indirect (electromyographic) evidence for greater palatal elevation during voiced stops.

Position and movements of the soft palate were also unaffected by place of articulation among medial consonant clusters. However, palatal position (but not movement) was affected by consonantal voicing in stop clusters. The soft palate was consistently higher by 2–3 mm, for example, throughout the closures of voiced–voiceless than comparable voiceless–voiced sequences of homorganic stops.

An interesting voicing-related contrast in palatal position and movement was observed during the medial clusters /bm/ and /pm/. In the former cluster, as illustrated by dashed functions in Fig. 10, the soft palate was well-elevated at first-member closure and the velopharyngeal orifice apparently closed. During the initial portion of that cluster, the

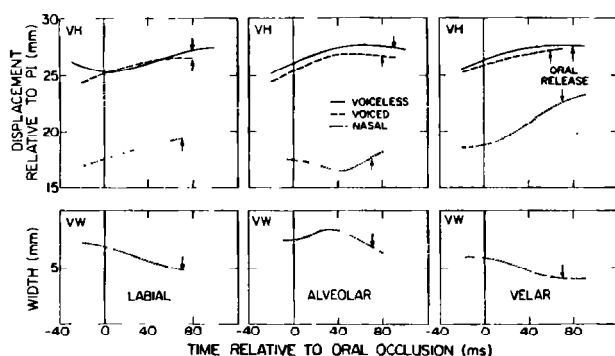


FIG. 9. Vertical movements of the soft palate during intervocalic nasals and voiced and voiceless stops. Changes in width of the velopharyngeal port are shown for nasals.

palate moved slowly but steadily downward while maintaining contact with the posterior pharyngeal wall. Then, approximately 50 ms prior to second-member release, the palate began lowering more rapidly and the velopharyngeal port began to open. In marked contrast, as can be seen via the solid function in Fig. 10, the port was open throughout the oral closure for medial /pm/. Velopharyngeal opening was first observed at least 40 ms prior to this cluster's first-member closure, during the latter third of the vowel which preceded it, and was sustained for virtually the full duration of the vowel which followed.

It is clear from /bm/ that "early" opening of the velopharyngeal port is not necessary during a homorganic stop-nasal cluster. That is, the port may be initially closed and then opened somewhat later, "in time" for the cluster-final nasal. We are left to wonder then why early opening occurs in /pm/. In at least one respect, the fact that it does so is surprising. Certainly in some phonetic environments, voicelessness during a stop closure seems to depend crucially on neutralization of the transglottal pressure gradient—so much so that if intraoral pressure is mechanically vented to atmosphere during the closure interval, voicing will resume (cf. Vencov, 1968; Putnam and Shipp, 1975). Clearly, velopharyngeal opening precludes an increase in pressure above the glottis, and for that reason, is unexpected during a stop closure which is underlyingly voiceless. It is well-known, however, that a stop consonant may be voiceless when accompanied by abduction or adduction of the vocal folds. Glottalized stops are the customary allophones of American English /p,t,k/ which occur as the initial components of certain medial consonant clusters (cf. Fujimura and Sawashima, 1971; Zue and Laferriere, 1979). Undoubtedly, the initial portion of the present instance of /pm/ was voiceless by virtue of this mechanism. Aerodynamic data recorded for the same speaker considered here but in other experiments indicate that instances of /p,t,k/ which occur as the first members of medial consonant clusters are produced with both oral and glottal closures (Westbury, 1979; Westbury and Niimi, 1979). The fact that the cluster-initial /p/ was glottalized "explains" why velopharyngeal opening was pos-

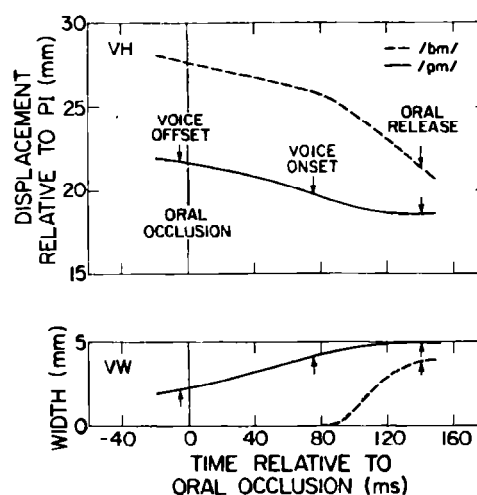


FIG. 10. Vertical movements of the soft palate, and changes in width of the velopharyngeal port, during the intervocalic clusters, /pm,bm/.

sible during its articulation. Voicelessness due to glottalization does not depend upon any increase in pressure above the glottis, and is therefore not incompatible with simultaneous velopharyngeal opening.

Notice, however, an interesting consequence of this view. If we suppose that articulatory gestures may be coarticulated only when they do not seriously compromise the acoustic and physiological properties which distinguish any particular speech sound from all others, then the instance of medial /pm/ illustrated in Fig. 10 suggests that the speaker "knew," prior to actual articulation, that "early" opening of the velopharyngeal port (for the cluster-final nasal) would not jeopardize voicelessness during the preceding stop. This result suggests, then, that the articulatory gestures (and/or acoustic properties) which characterize segmental allophones are specified and "known" to a speaker prior to specification of the precise nature of articulatory transitions between adjacent speech sounds in an utterance.

4. The tongue dorsum and tip

The articulatory constriction for a velar consonant is formed by elevating the tongue dorsum so that it touches the palate. In velar consonants observed in the present study, the dorsum usually appeared to move directly upward in transition from a preceding /a/, characteristically contacting the palate 58–64 mm posterior to the central maxillary incisors. The tongue body then appeared to roll forward following occlusion, so that the area of contact between dorsum and palate was anterior at release (by as much as 8 mm) to its position at closure (cf. also Houde, 1968; Perkell, 1969). Subsequently, the dorsum appeared to move both rearward and downward in transition to a following /a/, largely in concert with the mandible, tongue root, and hyoid bone.

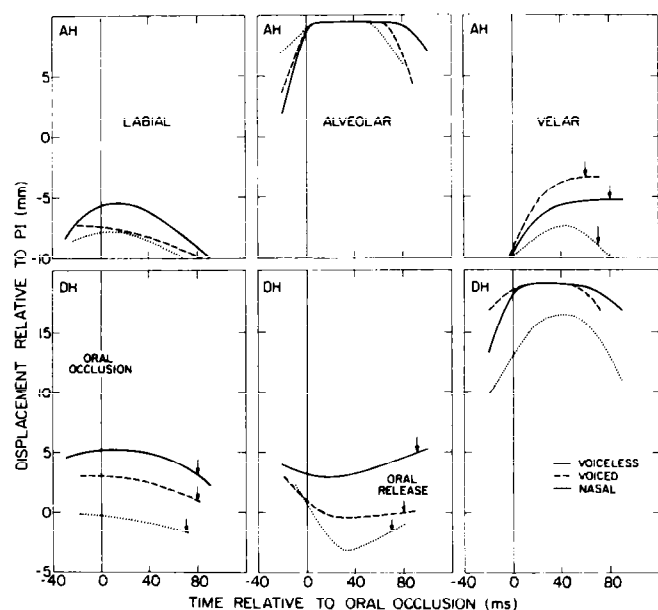


FIG. 11. Oblique movements of the tongue blade, and vertical movements of the tongue dorsum during intervocalic nasals and voiced and voiceless stops.

Similarly, the articulatory constriction for an alveolar consonant is formed by contacting the palate with the tongue tip. In transition from a preceding /a/, the tongue tip approaches the palate obliquely from below, eventually contacting that surface some 8–12 mm posterior to the central maxillary incisors. During closure, the area of contact between the tongue tip and anterior palate remains generally constant. The tongue tip then moves away from the palate in transition to a following /a/ along roughly the same path which it follows in its approach from that vowel.

The general positioning of the tongue dorsum and tip, and the nature of their approach and release movements, were largely unaffected by voicing of velar and alveolar stops, respectively. Representative measurements of position of those portions of the tongue during single medial consonants are shown in Fig. 11. The only notable difference in extent of tongue movements within each of the three places of articulation shown in this figure occurred among the velars, where the dorsum was lower (by 3–4 mm) during /ŋ/ than either /g/ or /k/. This difference is not surprising given that the palate itself was lower during the nasal than oral stops, and thus the dorsum need not have been elevated as high to contact it.

The tongue dorsum and tip are considered partly independent both of each other, and of the mandible and laryngo-hyoid complex (cf. Mermelstein, 1973). Consequently, they are potentially free to move about (and thereby affect cavity volume) during consonants whose constrictions are anterior to them. For that reason, movements of these portions of the tongue may play a role in the voiced-voiceless distinction among stops. In fact, Perkell (1969) has noted previously that the tongue dorsum appears to be depressed somewhat during the closure of medial /d/ but not /t/.

In most instances in the present study, position and movements of the upper tongue surface were unaffected by voicing during anterior stops. During medial /b/ and /p/, as shown in Fig. 11, both the dorsum and tip appeared to move slowly but steadily downward following occlusion. During

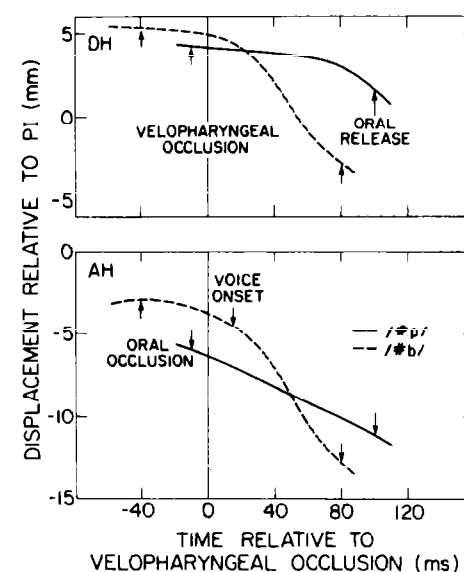


FIG. 12. Vertical movements of the tongue dorsum, and oblique movements of the tongue blade during utterance-initial /p,b/.

both medial /d/ and /t/, on the other hand, the dorsum moved slightly downward for 20–30 ms immediately following occlusion, and then slightly upward as release was approached. However, voicing-related differences in movement of the upper tongue surface were observed during utterance-initial labial and alveolar stops. Movements of the dorsum and tip during initial /p/ and /b/ are illustrated in Fig. 12. Note particularly that both portions of the tongue appeared to move downward in the mouth at a markedly faster rate following occlusion for the voiced stop. (Movements of the dorsum during initial /d/ and /t/ differed in the same way.) A consequence of the markedly faster downward movements during initial /b/ was a rapid increase in volume of the oral cavity during the interval prior to release when voicing was both initiated and sustained.

C. The temporal control of supraglottal volume

The cumulative effect of articulatory movements on volume of the cavity above the glottis is more relevant to the problem of voicing maintenance during consonantal closure than are the direction and extent of movements of any single articulator. Recall from data described previously, for example, that the larynx, tongue dorsum, and tongue tip moved steadily downward during the closure interval of utterance-initial /b/. At the same time, the tongue root moved steadily rearward. When considered alone, the first three movements (which are undoubtedly interdependent both with one another and with movement of the tongue root) cause cavity volume above the glottis to increase, and thereby facilitate voicing during closure. Conversely, rearward movement of the tongue root constricts the supraglottal cavity, and will impede closure voicing. It is important to know whether the net effect of changes in vocal tract dimensions resulting from several ongoing movements such as these is positive (i.e., cavity enlarging, and thereby voice sustaining) or negative (i.e., cavity constricting). For this reason, approximations of the temporal control of supraglottal cavity volume relative to the incidence and duration of voicing during stop closures have been calculated from data described in previous sections.

Changes in cavity volume between successive cinefluorographic frames can be approximated in rough form as the sum of products of movements of individual articulators between those frames and the respective surface areas over which those movements act. The articulator surface areas used in computing such changes were as follows: for the larynx, 2.5 cm²; for the soft palate, 3.0 cm²; for the tongue root, 2.8 cm² at level TR 1, 2.9 cm² at TR 2, 3.0 cm² at TR 3, 3.1 cm² at TR 4, and 3.2 cm² at TR 5; for the tongue dorsum, 10.0 cm²; and for the tongue blade/tip, 7.0 cm². The estimates of relevant surface areas of the larynx, soft palate, and portions of the tongue root are based upon anatomical data described by Goldstein (1980), while estimates of surface areas of the tongue dorsum and blade/tip are based upon measures of dental casts of the speaker in the current study.

Approximations of volume change per time change between successive film frames can be plotted as a function of elapsed time during any segment of interest, yielding a

time function equivalent to the parameter I_0 in the model described in the introduction to this study.² The integral of this function depicts the temporal control of cavity volume during a stop closure, relative to its value at the first moment of consonantal occlusion. An approximation of either $I_0(t)$ or its integral provides a means for determining whether the net effect of cavity-enlarging movements is likely sufficient to initiate and/or sustain voicing during the closure interval of a stop consonant.³

Functions which show the time change in supraglottal cavity volume during all consonants and consonant clusters considered in the current study are illustrated in Figs. 13 and 14. The volume function for any given consonant or cluster is derived from all measured articulatory movements behind the principal consonantal constriction (except for movements of the soft palate during intervals when the velopharyngeal orifice was open).

It is readily apparent from Fig. 13 that cavity volume—insofar as it is determined by positions of the larynx, soft palate, and portions of the tongue—generally increased during the closures of all single voiced stops. In every case but one—e.g., utterance-final /b/—cavity volume was increasing during the interval immediately after occlusion, and continued to increase thereafter, never beginning to decrease more than 20 ms prior to consonantal release (or voicing offset, if the latter occurred first). Cavity volume also increased during some but not all of the closures of single nasals and voiceless stops. However, for any place of articula-

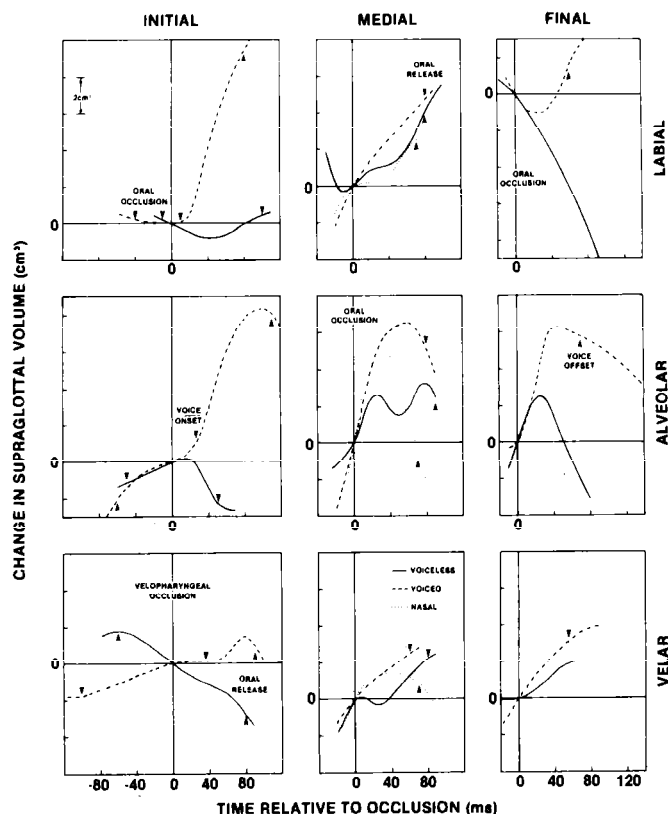


FIG. 13. Changes in supraglottal volume during the closures of intervocalic nasals, and voiced and voiceless stops occurring utterance-initially, intervocalically, and utterance-finally.

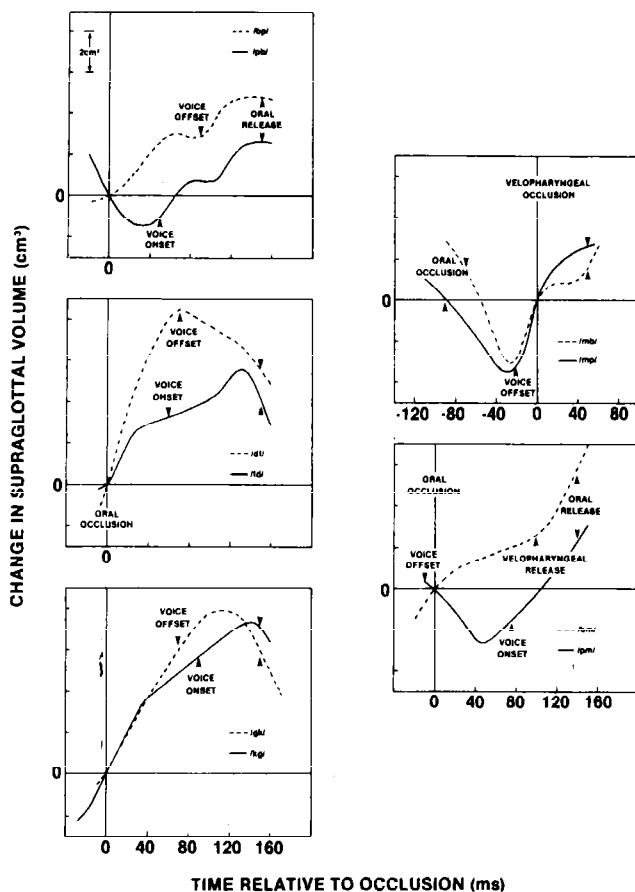


FIG. 14. Changes in supraglottal volume during the closures of the intervocalic clusters /pb, bp, td, dt, kg, gk, mb, mp, bm, pm/.

tion and phonetic environment where cavity volume increased during both (or all three) consonant types, the rate, and/or duration of the volume increase during the voiced stop was greater than during its nasal or voiceless cognate.

Data in Fig. 14 also show that cavity volume generally increased after occlusion in most medial consonant clusters. Among the stop clusters, particularly, it is both interesting and significant to note that the initial rate of volume increase was greater during two of three comparable sequences with voiced than voiceless first members, while the duration of the volume increase was greater for two of three sequences with voiced than voiceless second members. The results for contrasting pairs of mixed medial stop clusters—in which voicing is discontinuous—show then that the temporal control of supraglottal cavity volume seems to be related to the position of clusters' voiced intervals.

Perhaps because they are so few, the data do not reveal any startling effect of phonetic environment on the rate and duration of cavity expansion during the voiced stops. Among the single voiced stops themselves, volume increments are somewhat more robust during initial than medial or final /b,d,g/. It is also apparent that volume increments during mixed medial clusters—in which voiced stops abut their voiceless cognates—tend to be greater, in absolute terms, than during most single stops.

III. DISCUSSION

Although the data presented in the preceding section derive from a small set of utterances produced by one speaker, they show a number of systematic effects. Voiced stops were invariably accompanied by a general increase in volume of the supraglottal cavity, independent of the phonetic environment in which they occurred. In contrast, some voiceless stops were accompanied by a decrease in cavity volume. Cavity volume increased during other instances of /p,t,k/, but the rate and/or duration of such increases were customarily less than during comparable instances of /b,d,g/.

As long as the vocal folds remain suitably adducted and tensed, cavity enlargement will clearly facilitate their vibration during a stop closure. Enlargement may in fact be necessary for closure voicing, especially if the vocal tract walls are stiff, and/or the closure itself is relatively long or articulated close in time to other articulatory gestures which substantially lower either the driving pressure generated by thoracic mechanisms or the magnitude of glottal resistance to flow. Conversely, cavity enlargement or constriction will little affect voicelessness during a stop which is voiceless by virtue of either tight adduction or relatively wide abduction of the vocal folds. The relationship between glottal impedance and glottal area is nonlinear. In particular, muscularly actuated separation of the vocal folds—as is customary for prevocalic /p,t,k/ in American English—will hasten considerably the rise in pressure above the glottis. As a consequence, any transglottal pressure gradient which exists prior to occlusion will decay rapidly during a stop articulated with an open glottis, independent of even large changes in volume of the supraglottal cavity. For these reasons, we can assume—and the data seem to show—that cavity volume is a significant control variable for the speech motor system during stops which are phonetically voiced, but not during those which are both voiceless, and aspirated or glottalized.

Temporal control of supraglottal cavity volume was similar among the voiced stops only in categorical terms, in the sense that volume generally increased during the closure of every instance of /b,d,g/. However, the rates of observed increases and their time functions varied widely across both place of articulation and phonetic environment. For example, maximum rates of volume change within the closure intervals of voiced stops ranged between roughly +40 cm³/s (sustained for virtually the full duration of medial /g/) and something more than +200 cm³/s (shortly after occlusion during medial /d/). Minimum rates of volume change ranged between roughly -130 cm³/s preceding second-member release of medial /td/ and +40 cm³/s (prior to voice offset in medial /dt/). Despite such variation, all instances of /b,d,g/ were again categorically similar in that all evidenced significant intervals of voicing during their closures. Thus the incidence of closure voicing does not seem to depend upon a specific time function of volume change.

As a control parameter in speech, volume velocity flow due to (active) cavity enlargement (I_v) may be represented as a pulse whose own magnitude and duration are independent-

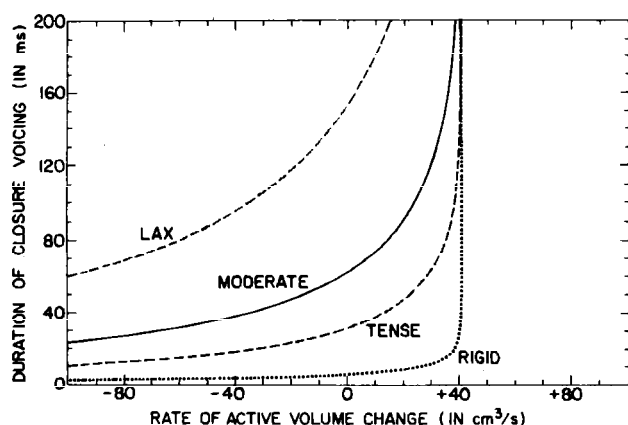


FIG. 15. The relationship between rate of volume change and calculated duration of closure voicing, determined by using the model shown in Fig. 1. All input functions to the model were the same as those used in generating Fig. 2, except for $I_0(t)$, which is here defined to be zero preceding occlusion and following release, but equal to some constant during the intervening interval. The different contours in this figure correspond to different levels of compliance of the vocal tract walls.

ly variable, and whose synchronization with the moments of articulatory closure and release is also variable. The model described in the introduction to this study can be used to estimate what values of I_0 magnitude, for example, will insure a pressure gradient continuously greater than some voicing threshold for the full duration of a stop. Figure 15 illustrates the effect of I_0 magnitude on the estimated duration of voicing during an intervocalic closure having a duration less than or equal to 200 ms, for the class of functions $I_0(t)$ wherein I_0 is zero preceding consonantal occlusion and following release and equal to some constant during the interval bounded by those events.

According to this figure, a medial stop bounded by "lax" walls and having a closure duration less than or equal to 152 ms will be fully voiced even when I_0 is zero during occlusion. (This is because the lax walls allow enough passive expansion to insure a suitable pressure drop over that interval.) For the same condition (i.e., $I_0 = 0$), stops bounded by moderate, tense, or rigid walls will be fully voiced as long as their respective closures do not exceed 62, 32, or 7 ms. Suppose, however, that a closure has a duration of 100 ms. What constant value of I_0 is necessary to maintain voicing for its full duration? If the walls of the supraglottal cavity remain rigid for the full closure interval, voicing will be continuous through it only if I_0 is at least $+40 \text{ cm}^3/\text{s}$, or if active mechanisms increase cavity volume 0.4 cm^3 every 10 ms. For tense and moderate cases, respectively, at least 37 and $25 \text{ cm}^3/\text{s}$ are required, while for the lax case, active mechanisms may produce decrements in cavity volume at a rate of up to $35 \text{ cm}^3/\text{s}$. (This is because the walls yield rapidly enough to absorb both the glottal flow and "negative" flow generated by larynx and tongue raising and palatal lowering.) An interesting conclusion suggested by Fig. 15 is that closure voicing can be sustained virtually indefinitely— independent of variation in wall impedance—for values of I_0 greater than roughly $+40 \text{ cm}^3/\text{s}$, as long as the rate of cavity enlargement is uniform over the full closure interval.

However, the rate of cavity enlargement is rarely uni-

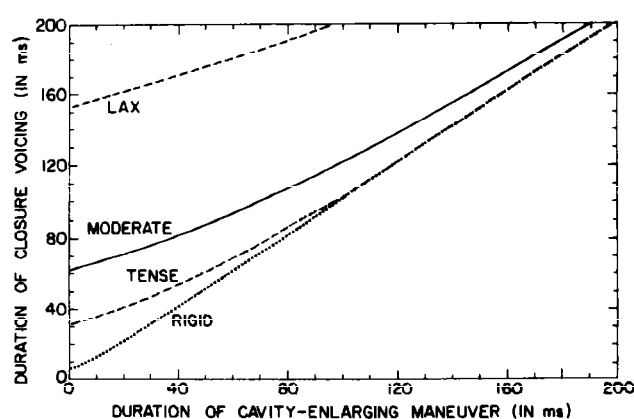


FIG. 16. The relationship between duration of a $50 \text{ cm}^3/\text{s}$ cavity-enlarging maneuver and calculated duration of closure voicing, for different levels of compliance of the vocal tract walls. Preceding occlusion, and coincident with or preceding release, I_0 is zero.

form in any of the volume functions shown in Figs. 13 and 14. Maximum rates of volume change derived from those functions are sometimes considerably in excess of $+40 \text{ cm}^3/\text{s}$. But rates of active volume change are less than that amount—in some instances, even negative—at some point during the closure of virtually every voiced stop. What is the effect of enlarging the cavity volume for only part of a stop closure? The model is used to estimate the effect of I_0 duration on closure voicing duration, for the class of functions $I_0(t)$ wherein I_0 is zero prior to occlusion, increased stepwise to $+50 \text{ cm}^3/\text{s}$ at that point and held constant for some interval, and then reduced stepwise back to zero prior to or coincident with release. In Fig. 16, the duration of closure voicing is plotted as a function of the duration of a $50 \text{ cm}^3/\text{s}$ cavity-enlarging maneuver, for intervocalic closures less than or equal to 200 ms in length. Note, for example, that when I_0 is held constant at $+50 \text{ cm}^3/\text{s}$ for the initial 80 ms of a medial closure, voicing will continue through a stop bounded by rigid, tense, moderate, or lax walls having a duration of 82, 86, 107, or 190 ms, respectively. For voicing over a full 100-ms interval, the duration of a closure-initial enlarging maneuver of that magnitude must fall between 96–100 ms for a rigid or tense-walled stop, 70–100 ms for a moderate stop, and 0–100 ms for a lax stop.

If it is more important during voiced stops to control whether (rather than how) the vocal folds oscillate, then all cavity-enlarging maneuvers whose magnitude and duration satisfy the boundary conditions necessary for oscillation can be considered equally well-suited for that behavioral goal. In this regard, then, closure voicing is not unlike other acoustic properties in speech which "can be generated by the articulatory apparatus without...precise positioning or maneuvering of the articulatory structures" (Stevens, 1972:52). Certainly there is good reason to believe the acoustically well-defined phenomenon to be insensitive to variation in *form*—i.e., magnitude and relative timing—of one of the principal articulatory parameters which give rise to it.

It is also apparent that the *means* by which volume increases are effected vary considerably among the different voiced stops. Data describing the positioning and move-

ments of individual articulators show, for example, that the directions of articulatory movements which affect cavity volume can differ among voiced stops produced in the same environment but having different places of articulation. Recall from Fig. 5 that the larynx moved generally downward during single medial /b/ and /d/, enlarging the cavity above, but remained relatively stable in the neck during medial /g/. At the same time, as shown in Fig. 7, the tongue root moved steadily forward during virtually the entire closures of /d/ and /g/, but first slightly forward and then rearward during /b/. The directions of voicing-related movements can also differ among voiced stops having the same place of articulation but occurring in different environments. During /b/ occurring utterance-initially, for example, the larynx, tongue dorsum, and tongue tip moved downward, enlarging the supraglottal cavity, while the tongue root moved rearward. During /b/ occurring intervocalically, the larynx, dorsum, and tip again moved steadily downward, but the root moved initially slightly forward and then rearward. During /b/ occurring utterance-finally, the larynx remained relatively stationary while the tongue dorsum and tip moved slightly upward, and the root moved forward. These examples illustrate that physiological realization of the categorical feature specification [+ voice]—when considered in terms of several quasi-independent articulatory movements (which may or may not be specific to, but are undeniably relevant to control of closure voicing)—is neither place nor context-invariant. Rather, interpretation of that specification by the speech motor system is as variable as the combination of place of articulation and environment will allow.

The observation that “a given feature may be given more than one motor interpretation” (Kent and Minifie, 1977:131) is not new. But, the significance of the explanatory problem posed by such an observation for a general theory of speech motor control cannot be overstated. Such a theory must explain not only why, for a common articulatory or acoustic goal, variability is possible in the movement patterns by which it is achieved, but also, what principles determine the specific movement pattern we observe in any single implementation of that goal. Answers to the first question have been suggested above. Recall that the incidence of closure voicing does not require a specific time function of volume change for its successful realization. Any of a great number of active volume adjustments—including no change at all—can be consonant with ongoing vibration of the vocal folds. Moreover, any given function $I_0(t)$ sufficient for closure voicing may be achieved in many ways. If we consider the larynx, soft palate, tongue root, dorsum, and tip each to be at least partly independent, the number of coordinated combinations of their movements (each having its own direction, rate, and duration) satisfying the general constraint that volume above the glottis increase somewhat in time is quite large. Thus variability in motor behavior at least compatible with closure voicing is possible both because that end is (presumably) categorical in nature, and because the number of degrees of freedom which can be brought to bear on it is relatively large.

There is an intuitive appeal to acoustic and articulatory

goals which permit considerable variation in the means by which they are achieved. Motor behavior directed toward such goals will likely be highly successful, independent of the peripheral state of the articulatory mechanism when such behavior is initiated, or of perturbations of that behavior which may arise in the course of speaking. As a consequence, the control mechanism need not devote significant time and energy to monitoring, moment-to-moment, the relationship between its state and concurrent and future intentions. Motor activity relevant for “quantal” acoustic (or articulatory) goals, as they have been described by Stevens (1972), might therefore be initiated in an open-loop fashion and without regard for context.

Suppose, then, that motor behavior relevant for stop voicing is *not* somehow “tuned” to the contexts in which it occurs. This assumption implies either that such behavior should be the same for all contexts, or that such behavior may vary across contexts, but—implausible as it may seem—in an unpredictable way. The new data reported here—derived from single examples of each of three voiced stops in several phonetic environments—show clearly that the former is not true. However, it is equally clear that these same data cannot show whether the contextual variations in motor behavior relevant for stop voicing are predictable. It would be of great interest to know whether and to what extent voicing-related behavior might vary for the same stop, repeated many times by the same speaker, in the same phonetic environment. For two reasons, data such as these are crucial for a theory which seeks to relate categorical descriptors such as features to their specific motor interpretations. First, such data will show whether the relationship between a feature and its interpretation is indeed deterministic. And second, in the event that it is not, such data may provide insight into optimization criteria which favor some of the possible feature interpretations over others, or alternatively, suggest factors which greatly constrain the set of feature-compatible behaviors.

IV. CONCLUSION

Positions and movements of the larynx, soft palate, and various portions of the tongue have been measured from a high-speed cinefluorographic record of isolated disyllables produced by a single speaker of American English. The movements of each of these articulators may affect the time change of volume of the cavity above the glottis, and may therefore affect the incidence and duration of voicing during stop closure. Summing the “volume contributions” of individual articulatory movements reveals that stops which are both categorically and phonetically [+ voice] always exhibit active volume increases during their closures. That is not the case for stops which are [– voice]. There is good reason to believe that an active increase in supraglottal volume will facilitate voicing during a stop, as long as the glottal state remains constant and essentially like that during a vowel. On the other hand, an active change in supraglottal volume will little affect voicelessness during a stop, if the stop is voiceless by virtue of substantial changes in the degree or nature of approximation of the vocal folds. Thus, *how* supraglottal

cavity volume changes with time is relatively more important for stop voicing than voicelessness.

We note, however, that *when* volume increases occur during voiced closures, how large they are, and how they are generated, vary widely across both place of articulation and phonetic environment. This observation suggests that the precise nature of any active volume increase during any instance of /b,d,g/ is not especially important for closure voicing. The same conclusion may be derived from consideration of a model of breath-stream dynamics. In particular, there are infinitely many ways to satisfy the boundary conditions for voicing during closure. Thus specifying only that a stop is [+ voice] implies very little about how cavity volume will be controlled in time. Perhaps the place of articulation of a stop and where in an utterance it occurs must also be known. That is to say, the motor interpretation of the voicing feature during a stop may depend upon interpretations of its other features, and of features of its environment.

ACKNOWLEDGMENTS

Data described in this paper are taken from one chapter of a doctoral thesis completed at The University of Texas at Austin under the direction of Peter F. MacNeilage. Implementation of the aerodynamic model was completed during a term as an NIH Post-doctoral Trainee at the Research Laboratory of Electronics of MIT, and in collaboration with Patricia A. Keating and Kenneth N. Stevens. I am grateful to K. N. Stevens, J. S. Perkell, R. M. Dalston, and M. Rothenberg for comments on versions of this report. This work was supported at various times by NSF Grant BNS 77-07686, NIDR Grant DE-00853, NIH Grant NS-07040, and NIDR/NIH Grant DE-02668.

APPENDIX

Values for fixed elements in the model are as follows:

$$C_i = 0.1 \text{ cm}^5/\text{dyn}$$

$$L_i = 3.7 \times 10^{-2} \text{ dyn-s}^2/\text{cm}^5$$

$$R_i = 3.3 \text{ dyn-s/cm}^5$$

$$C_s = 3.437 \times 10^{-3} \text{ cm}^5/\text{dyn}$$

$$R_s = 3.0 \text{ dyn-s/cm}^5$$

$$C_w = 5.625 \times 10^{-4} \text{ cm}^5/\text{dyn}$$

$$L_w = 1.2 \times 10^{-2} \text{ dyn-s}^2/\text{cm}^5$$

$$R_w = 8.48 \text{ dyn-s/cm}^5$$

$$C_0 = 8.1 \times 10^{-5} \text{ cm}^5/\text{dyn}$$

The last four values are considered appropriate for a vocal tract bounded by moderately tense walls (cf., Ishizaka *et al.*, 1975), having a length of 16 cm and uniform elliptical cross section with major and minor axes of 4 and 2 cm (cf., Fant *et al.*, 1976). The surface area of such a tube, which determines any of R_w , C_w , or L_w , has been adjusted downward somewhat to 125 cm² to take into account the fact that a large portion of the palate is hard. Values for R_w , C_w , and L_w , for "lax" and "tense" conditions simulated in Figs. 3, 15, and 16, are 6.4 dyn-s/cm⁵, 1.479×10^{-3} cm⁵/dyn, 1.68×10^{-2} dyn-s²/cm⁵, and 18.56 dyn-s/cm⁵,

2.5458×10^{-4} cm⁵/dyn, 1.92×10^{-2} dyn-s²/cm⁵, respectively (again, based on data of Ishizaka *et al.*, 1975).

When the model is initialized to approximate an inter-vocalic stop, as in Figs. 2, 3, 15, or 16, E_s is set to zero, with pressure drops of -1.0×10^4 dyn/cm² and 9.8×10^3 dyn/cm² across C_i and C_s , respectively, and a volume velocity flow of 125 cm³/s through L_i . The glottis is assumed to be a rectangular constriction with constant dimensions of 1.8 cm (length), 0.022 cm (width), and 0.3 cm (thickness, parallel to the direction of laminar flow). Presumably, its constant cross-sectional area (of roughly 0.04 cm²) approximates the *average* area of the glottal slit over a full period when the vocal folds are vibrating during a vowel. No attempt has been made to represent the time change in glottal area which occurs during voicing, since the focus of this model is to consider pressure dynamics over time intervals considerably longer than the glottal period (cf., discussion by Müller and Brown, 1980). It should also be noted that vocal-fold stiffness is not represented directly in the model. Rather, that parameter is included only insofar as it determines the pressure gradient threshold necessary for oscillation.

The glottal "state"—in particular, its area—remains unchanged over the full 200-ms interval illustrated in Fig. 2. It is assumed that no overt effort is made to adjust position of the vocal folds in a vowel-voiced stop-vowel sequence. The velopharyngeal orifice is considered closed, while mouth opening (at the lips) is represented in terms of a rectangular constriction with two dimensions fixed—0.2 cm in length (parallel to flow) and 2.0 cm in width—and one dimension (height) variable. The occlusions in Figs. 2 and 3 are simulated by "closing" and "opening" such a constriction at uniform rates of 20 cm/s over the 15-ms intervals immediately preceding occlusion and following release.

¹Using methods different from those of Ishizaka *et al.* (1975), Rothenberg (1968) determined that total compliance of the supraglottal cavity walls might be roughly three times greater than the "upper limit" used to derive data summarized in Fig. 3—perhaps as high as 5×10^{-3} cm⁵/dyn. If total compliance may be as large as that, then the duration of medial closure voicing which might result from making the walls "as lax as possible" could approximate 500 ms.

²It is at least difficult, if not impossible, in cinefluorographic data like that described by Perkell (1969), Kent and Moll (1969), or in this study, to differentiate changes in vocal tract dimensions resulting from cavity expansion and cavity enlargement. Following Rothenberg (1968), expansion refers to volume increases resulting from movements of the elastic vocal tract walls due to air pressure they contain. Enlargement, on the other hand, refers to volume increases resulting from muscularly actuated movements of certain articulatory structures. If we suppose the vocal tract walls to be everywhere as compliant as tense cheek tissue, then they will be displaced outward 0.045 mm per 1000 dyn/cm². If those surfaces are mechanically "closer" to the neck wall, they will yield perhaps 0.02 mm per 1000 dyn/cm²; or if they are "closer" to lax cheek tissue, 0.118 mm per 1000 dyn/cm². Assuming further that wall compliance is constant, and that air pressure in the mouth approaches (relatively slowly) a maximum of 10 000 dyn/cm² during a stop consonant, we can expect the walls to move passively outward perhaps as much as 1 mm. But because the actual compliance is unknown, we cannot know the extent to which relatively small movements of surfaces bounding the vocal tract are due to expansion or enlargement. It is certainly plausible, however, to assume movements of more than 1.5 mm to result from significant muscular effort to enlarge the cavity above the glottis. Discussion of this point has particular relevance for data presented in latter sections of this study. It is something of a fiction, though unavoidable, to attribute the volume changes observed dur-

- ing various voiced and voiceless stops to the control parameter I_0 . Undoubtedly, some portion of what is referred to as I_0 is I_c .
- ³The aspect of cavity volume most relevant to voicing maintenance during stops is not its absolute value. That is, voicing during closure will not necessarily be well-served by neuromotor commands to articulators which entail, as their goal, a vocal tract configuration defining some absolute value. The fact that the pharynx is generally larger at the first moment of occlusion for voiced than voiceless stops is of relatively little use in sustaining voicing during the former. Rather, voicing will be better served by commands which entail what we might loosely refer to as a trajectory of volume change, wherein change per unit time, for all moments in time, exceeds some reference value.
- Baer, T. (1975). "Investigation of Phonation Using Excised Larynxes," unpublished doctoral dissertation, MIT.
- Bell-Berti, F. (1975). "Control of Pharyngeal Cavity Size for English Voiced and Voiceless Stops," *J. Acoust. Soc. Am.* **57**, 456-464.
- van den Berg, Jw. (1958). "Myoelastic-Aerodynamic Theory of Voice Production," *J. Speech Hear. Res.* **1**, 227-244.
- Dixit, R. P., and MacNeilage, P. F. (1972). "Coarticulation of Nasality: Evidence from Hindi," *J. Acoust. Soc. Am.* **52**, 131.
- Ewan, W. G., and Krones, R. (1974). "Measuring Larynx Movement Using the Thyroembrometer," *J. Phon.* **2**, 327-335.
- Fant, G. (1972). "Vocal Tract Wall Effects, Losses, and Resonance Bandwidths," *Speech Trans. Lab. Q. Prog. Status Rep.* **2-3**, 28-52.
- Fant, G., Nord, L., and Branderud, P. (1976). "A Note on the Vocal Tract Wall Impedance," *Speech Trans. Lab. Q. Prog. Status Rep.* **4**, 13-20.
- Flanagan, J. L., Ishizaka, K., and Shipley, K. L. (1975). "Synthesis of Speech From a Dynamic Model of the Vocal Cords and Vocal Tract," *Bell Syst. Tech. J.* **54** (3), 485-506.
- Fujimura, O., and Sawashima, M. (1971). "Consonant Sequences and Laryngeal Control," *Ann. Bull. Res. Inst. Logoped. Phoniatri.* (Tokyo) **5**, 1-6.
- Goldstein, U. G. (1980). "An Articulatory Model for the Vocal Tracts of Growing Children," doctoral dissertation, Massachusetts Institute of Technology (unpublished).
- Halle, M., and Stevens, K. N. (1971). "A Note on Laryngeal Features," *MIT Res. Lab. Electron. Q. Prog. Rep.* **101**, 198-213.
- Houde, R. A. (1968). "A Study of Tongue Body Motion During Selected Speech Sounds," *Speech Commun. Res. Lab. Mono. No.* **2**.
- Ishizaka, K., and Matsudaira, M. (1972). "Fluid Mechanical Considerations of Vocal Cord Vibration," *Speech Commun. Res. Lab. Mono. No.* **8**.
- Ishizaka, K., French, J. C., and Flanagan, J. L. (1975). "Direct Determination of Vocal Tract Wall Impedance," *IEEE Trans. Acoust., Speech Signal Process.* **ASSP-23** (4), 370-373.
- Kent, R. D., and Moll, K. L. (1969). "Vocal Tract Characteristics of the Stop Cognates," *J. Acoust. Soc. Am.* **46**, 1549-1555.
- Kent, R. D., and Minifie, F. D. (1977). "Coarticulation in Recent Speech Production Models," *J. Phon.* **5**, 115-133.
- Künzel, H. J. (1979). "Some Observations on Velar Movement in Plosives," *Phonetica* **36**, 384-404.
- Ladefoged, P. (1964). "Comment on 'Evaluation of Methods of Estimating Subglottal Air Pressure,'" *J. Speech Hear. Res.* **7**, 291-292.
- Lindqvist, J. (1972). "Laryngeal Articulation Studied on Swedish Subjects," *Speech Trans. Lab. Q. Prog. Stat. Rep.* **2-3**, 10-27.
- Lisker, L. (1977). "Factors in the Maintenance and Cessation of Voicing," *Phonetica* **34**, 304-306.
- Löfqvist, A. (1975). "A Study of Subglottal Pressure During the Production of Swedish Stops," *J. Phon.* **3**, 175-189.
- Lubker, J. F. (1973). "Transglottal Airflow During Stop Consonant Production," *J. Acoust. Soc. Am.* **53**, 212-215.
- Lubker, J. F., Fritzell, B., and Lindqvist, J. (1970). "Velopharyngeal Function: An Electromyographic Study," *Speech Trans. Lab. Q. Prog. Stat. Rep.* **4**, 9-20.
- McGlone, R. E., and Shipp, T. (1972). "Comparison of Subglottal Air Pressures Associated with /p/ and /b/," *J. Acoust. Soc. Am.* **51**, 664-665.
- Mermelstein, P. (1973). "Articulatory Model for the Study of Speech Production," *J. Acoust. Soc. Am.* **53**, 1070-1083.
- Minifie, F. D., Abbs, J. H., Tarlow, A., and Kwaterski, M. (1974). "EMG Activity Within the Pharynx During Speech Production," *J. Speech Hear. Res.* **17**, 497-504.
- Moll, K. L. (1962). "Velopharyngeal Closure on Vowels," *J. Speech Hear. Res.* **5**, 30-37.
- Müller, E. M., and Brown, Jr., W. S. (1980). "Variations in the Supraglottal Air Pressure Waveform and their Articulatory Interpretation," in *Speech and Language: Advances in Basic Research and Practice*, Vol. 4, edited by N. Lass (Academic, New York), pp. 317-389.
- Netsell, R. (1969). "Subglottal and Intraoral Air Pressures During the Intervocalic Contrast of /t/ and /d/," *Phonetica* **20**, 68-73.
- Perkell, J. S. (1969). *Physiology of Speech Production* (MIT, Cambridge).
- Putnam, A. H. B., and Shipp, T. (1975). "EMG-Aerodynamic Measures for Intervocalic /p/ Production," *J. Acoust. Soc. Am. Suppl.* **1** **57**, S70.
- Riordan, C. J. (1980). "Larynx Height During English Stop Consonants," *J. Phon.* **3**, 353-360.
- Rothenberg, M. (1968). "The Breath-Stream Dynamics of Simple-Released-Plosive Production," *Bibl. Phonetica* **6**.
- Stetson, R. H. (1950). *Motor Phonetics* (North Holland, Amsterdam).
- Stevens, K. N. (1972). "The Quantal Nature of Speech: Evidence from Articulatory-Acoustic Data," in *Human Communication, A Unified View*, edited by P. F. Denes, and E. E. David, Jr. (McGraw-Hill, New York), Chap. 3, pp. 51-66.
- Ushijima, T., and Sawashima, M. (1972). "Fiberscopic Observation of Velar Movements During Speech," *Ann. Bull. Res. Inst. Logoped. Phoniatri.* (Tokyo) **6**, 25-38.
- Vencov, A. V. (1968). "A Mechanism for Production of Voiced and Voiceless Intervocalic Consonants," *Z. Phonet., Sprachwiss. Kommunikationsforsch.* **21**, 140-144.
- Westbury, J. R. (1979). "Aspects of the Temporal Control of Voicing in Consonant Clusters in English," *Texas Linguist. Forum* **14**, 1-304.
- Westbury, J. R., and Niimi, S. (1979). "An Effect of Phonetic Environment on Voicing Control Mechanisms During Stop Consonants," in *Speech Communication Papers* presented at the 97th Meeting of the Acoustical Society of America, edited by J. Wolf and D. Klatt (Acoust. Soc. Am., New York).
- Yanagihara, N., and Hyde, C. (1966). "An Aerodynamic Study of the Articulatory Mechanism in the Production of Bilabial Stop Consonants," *Stud. Phonolog.* **4**, 20-27.
- Zue, V. W., and Laferriere, M. (1979). "Acoustic Study of Medial /t,d/ in American English," *J. Acoust. Soc. Am.* **66**, 1039-1050.