947

Journal of Phonetics (1995) **23**, 255–269

Incorporating aerodynamic and laryngeal components into task dynamics

R. S. McGowan and E. L. Saltzman

Haskins Laboratories, 270 Crown Street, New Haven, CT 06511-6995, U.S.A.

(Received 31st August 1994, and in revised form 11th October, 1994)

Aerodynamic and laryngeal components were added to the task-dynamic model to make it a more complete model of speech production. New tract variables included were subglottal pressure, transglottal pressure, and a quantity giving fundamental frequency in sonorant conditions. New articulators included total force on the lungs, generalized laryngeal tension, and supralaryngeal volume. Simulations of /VCV/ sequences, where C was either a voiced or a voiceless aspirated stop, demonstrated the usefulness of these additions.

1. Introduction

Task dynamics has been used successfully to model the coordinated movements of the upper vocal tract articulators in attaining spatial goals in a model vocal tract (Saltzman, 1986; Saltzman and Kelso, 1987; Saltzman and Munhall, 1989). The particular model vocal tract employed with task dynamics was originally developed by Mermelstein (1973), and later incorporated into the Haskins Laboratories articulatory synthesizer (Rubin, Baer and Mermelstein, 1981). In task dynamic simulations, linguistic goals are specified as the formation and release of constrictions in the vocal tract that are produced by movements of the lips, tongue tip, tongue body, velum, and larynx. These constriction tasks are described by a set of so-called *tract variables* that are associated with corresponding sets of *articulators* lip protrusion and lip aperture with the upper lip, lower lip, and jaw articulators; tongue-tip constriction location and tongue-tip constriction degree with the tongue tip, tongue body, and jaw articulators; tongue-body constriction location and tongue-body constriction degree with the tongue body and jaw articulators; velum height with the raising and lowering of the velum; and glottal width with the abduction and adduction of the vocal folds. The glottal width tract variable has been the only laryngeal tract variable employed prior to the work described in this report.

Constriction tasks are performed by recruiting various articulators, so that, for instance, a bilabial closure target is attained by recruiting the upper lip, the lower lip, and the jaw. The tract variables' dynamics are described by a set of independent, linear, second-order differential equations. For example, the changes of lip aperture that are used in attaining a bilabial closure are modeled as the motions of a linear mass-spring system with a rest, or target, position less than or

0095-4470/95/010255 + 15 $08.00/0

equal to zero. These dynamics are activated for finite time intervals, known as *gestural activation intervals*. In performing a gesture, the dynamics of the articulators are obtained by a geometric transformation from tract-variable accelerations to articulator accelerations, resulting in a coupled set of nonlinear differential equations at the articulator level. Thus, for a bilabial closure, three coupled, nonlinear differential equations define the movements of the upper lip, lower lip, and the jaw.

The task-dynamic model of speech production provides a means of modeling the coordinated activities of articulators that are used in making gestures during speech, and has been incorporated into an *articulatory phonology* (Browman & Goldstein, 1990) model in which the basis for phonetic contrasts is provided by the gestures themselves. However, the task-dynamic description is incomplete without a more thorough account of aerodynamic and laryngeal phenomena (e.g., Mattingly, 1990). Such phenomena, largely pertinent to acoustic sources, include the production of stops and fricatives, variations in speech intensity or effort, laryngeal state during both open and closed tract conditions, intonation, and stress contrasts. Within the framework of task-dynamics, it is therefore important to define a goal-directed dynamics for certain aerodynamic and laryngeal quantities as well as for the supralaryngeal articulators. The new tract variables should include, roughly, subglottal pressure/intensity, transglottal pressure/voicing, and laryngeal tension/fundamental frequency. To control the new tract variables, new articulators beyond glottal abduction/adduction should include, at a minimum, a laryngeal tension parameter, compressive force on the lungs, and supralaryngeal volume.

Fig. 1 shows a schematic view of the laryngeal and supralaryngeal vocal tract from the low-frequency aerodynamic perspective [Fig. 1(a)] and its representation as a circuit diagram [Fig. 1(b)]. Although the circuit we have adopted is a simplified version of circuits used by Rothenberg (1968), Müller and Brown (1980), and Westbury (1983), it still manages to capture the essential features of the control of aerodynamic quantities. The aerodynamic model was constrained to be simple since, otherwise, the computation time could grow very quickly with the addition of new aerodynamic components. Also, with the simpler system it was easier to diagnose problems in this new implementation of task dynamics. However, some of the passive responses to changes, such as subglottal pressure variations due to the finite impedance of the sugblottal system, cannot be seen in such a simplified model.

In order to incorporate these simplified aerodynamics into the task dynamic model, the relations between the original articulators and tract variables were retained, and new sets of tract variables and articulators were defined. The new tract variables were, roughly, subglottal pressure (P_s), transglottal pressure (P_t) and fundamental frequency (F_0); the new articulators were lung force (F_l), generalized laryngeal tension (T_g), and the vocal tract volume (V_{uv}) between the glottis and the most constricted part of the supralaryngeal or upper vocal tract. The choice of a particular set of tract variables is dependent on what the modeler considers to be the goals of the speaker. Articulators have a specialized meaning in regards to task-dynamics: they are the effector system variables that are functionally related to the chosen tract variables in whatever articulatory model is chosen to instantiate the task dynamics.

While the dynamics of the original tract variables and articulators were not affected by these additions, the original articulators certainly had an effect on the

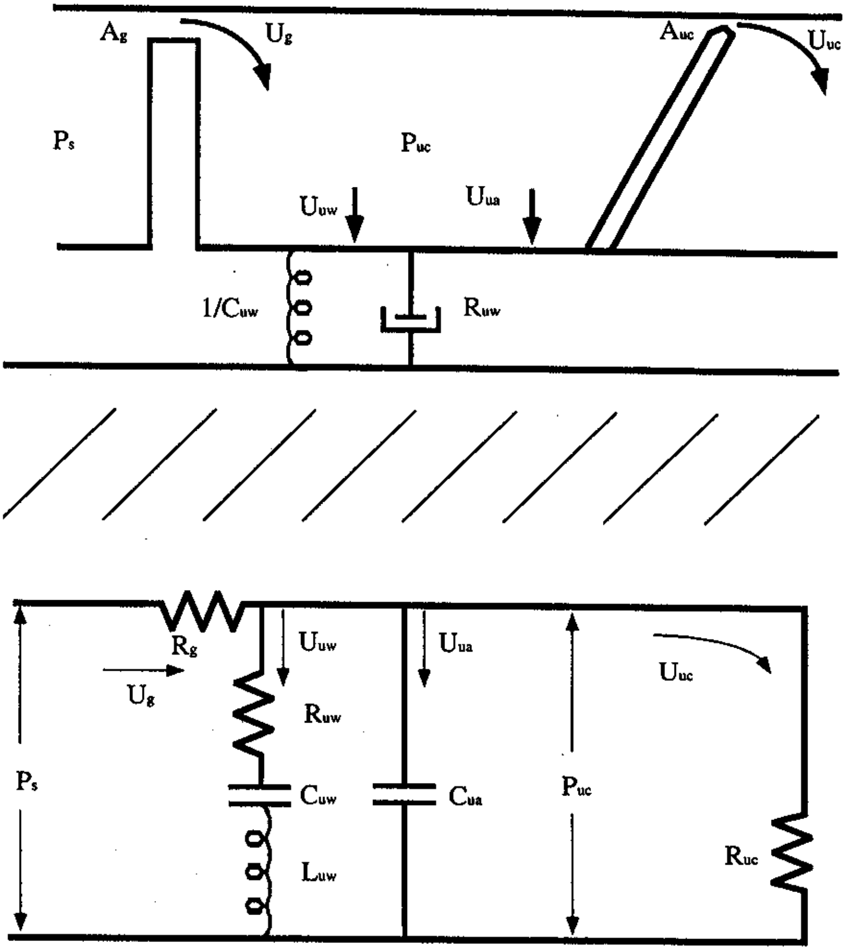

Figure 1. Low-frequency aerodynamic model. P denotes pressure, U volume velocity, A area, C capacitance, R resistance, and L inductance. The subscript s refers to subglottal quantities; the subscript g refers to glottal elements and flow quantities; the subscript uw refers to supraglottal vocal tract wall elements and flow quantities; the subscript ua to elements and flow quantities associated with the volume of air in the supralaryngeal cavity behind a supralaryngeal constriction; and the subscript uc to supralaryngeal constriction elements and flow quantities.

new tract variables. For instance, because transglottal pressure is affected by tongue-tip constriction degree and, hence, articulators such as jaw angle, transglottal pressure can be increased during a voiced alveolar fricative by decreasing the angle between the jaw and hard palate. Although not used in current simulations, future applications of task dynamics with aerodynamic variables should allow the original supralaryngeal articulators to be recruited to help attain the targets of the new tract variables.

Figs 1(a) and (b) show that there are variables evolving in the flow system that are neither articulators nor tract variables, but whose evolution must be accounted for, because they are part of the equations that map articulator states into tract variable states. Thus, for example, the volume velocity[1] of air due to the compliance of the upper vocal tract walls, U_{uw}, is neither an articulator nor a tract variable, yet it must be included in the equations that define the evolution of transglottal pressure. The mathematics for including these intermediate flow variables in the task-dynamic simulations is shown in the Appendix. Essentially, a term must be added to the articulatory evolution equation that accounts for the dependency of the transformation from articulatory to tract variables on these intermediate flow quantities. Also, the evolution of these intermediate flow variables must be solved simultaneously with the evolution of the articulators.

2. The aerodynamic and laryngeal tract variables

2.1. *Subglottal pressure* (P_s)

The respiratory system during speech provides either a relatively constant subglottal pressure (Hixon, 1987; Ohala, 1990, p. 25) or a subglottal pressure trajectory that resembles exponential decay over the course of several syllables (e.g., Gelfer, Harris, Callier and Baer, 1983). The decay in subglottal pressure contributes to a corresponding declination of fundamental frequency, although laryngeal tension may also contribute to declination (see Liberman and Pierrehumbert [1984] on fundamental frequency declination). These statements refer to long-term, time-averaged variations of subglottal pressure trajectories over several syllables, without regard to the short-term variations due to changes from phoneme to phoneme in the glottal and supraglottal loads.

There are two sets of variables that are important determinants of subglottal pressure. One set includes: the passive elastic forces of the lung tissue; the host of connective tissue and (active and passive) muscular forces, such as those produced by the interior intercostals and diaphragm; and the forces due to the volume displacement of the air in the lungs or to the capacitance of the air itself (Rothenberg, 1968; Hixon, 1987). This first set can be collapsed into a single articulator defined as the total compressive force applied to the lungs (Ohala, 1990). The second set of variables that contribute to subglottal pressure are all the load

---

[1] Note that the word "volume" is used in three ways in this work. The first is the common usage of *volume* to define spatial extent in three dimensions (cm^3). In this sense, upper vocal tract volume, V_{uv}, defines the amount of three-dimensional space in the "container" defined by the mouth, that is, in the spatial region from the larynx to the lips. Since air is compressible, there can be different amounts of air in a given upper vocal tract volume; additionally, since the walls of the mouth are distensible, the same amount of air can fill different upper vocal tract volumes. It is easiest to define what is denoted by the phrase "amount of air" after defining the second way that volume is used in our discussions, namely, in the term *volume velocity*. Volume velocity of air, U, can be defined, approximately for low speed flow, as the mass flux of air (g/sec) through a surface element divided by the density of air (g/cm^3). Thus, volume velocity has the units of cm^3/sec. For modeling the aerodynamics of the vocal tract, the surface elements are generally planes whose normals are parallel to the tract's midline, so that the glottal opening and supralaryngeal constriction opening are such surfaces. *Volume displacement* of air, Q, is the time integral of volume velocity, where zero volume displacement corresponds to air at rest at atmospheric conditions. Therefore, Q has the units of volume, cm^3. The volume displacement of air associated with the supralaryngeal volume, V_{uv}, is the time integral of the volume velocity through the glottis, U_g, minus the time integral of volume velocity through the supraglottal constriction, U_{uc}. As alluded to above in terms of "amount of air", the volume displacement in the supralaryngeal tract has two components: one due to the compressibility of air, denoted Q_{ua}, and the other due to the distensibility of the tract walls, denoted Q_{uw}.

elements, both subglottal and supraglottal, that are provided by the subglottal tissue and air (except lung tissue and air in the lungs), glottis, supraglottal constrictions, supraglottal air compliance, and the vocal tract walls (see Rothenberg, 1968, p. 21).

Thus, supraglottal, glottal, and subglottal elements all affect the subglottal pressure. For the purposes of our simulations, we assumed that only the long-term, or overall, trends in subglottal pressure are controlled, i.e. the pressure decay that has been observed experimentally. Also, we adopted the simplifying assumption that total lung force (F_l) was the only variable that contributed to the control of subglottal pressure. Thus, in the absence of an explicitly modeled subglottal system, the specification of total lung force is equivalent to the specification of subglottal pressure. The ramifications of collapsing the active and passive components of force on the lungs into their sum, F_l, will be discussed below.

The subglottal pressure (P_s) dynamics is modeled as the exponentially decaying dynamics of a simple first-order system:

$$\dot{P}_s + \alpha(P_s - P_{s\infty}) = 0 \tag{1}$$

where P_s is the long-term target or asymptote, and $1/\alpha$ is the system time constant. Without a subglottal system there is a one-to-one relation between subglottal pressure (P_s) and total force on the lungs (F_l). Therefore, the articulator F_l has the same dynamics as the tract variable P_s:

$$\dot{F}_l + \alpha(F_l - F_{l\infty}) = 0 \tag{2}$$

Note, however, that the total lung force depends on both the active force of the muscular tissue impinging on the thoracic volume, and on the current thoracic volume (usually expressed as a percent of vital capacity) through its effect on the passive capacitance of the tissues and air volume (Hixon, 1987, p. 27). That is, different *active* force-on-the-lungs trajectories are required to produce the same subglottal pressure trajectory when the lungs are deflating at different rates, i.e., for different percent vital capacity trajectories. In the modeling done here, the total force on the lungs has not been decomposed into its active and passive components, so that articulatory compensation in attaining the same subglottal pressure trajectories for different percent vital capacity trajectories is not modeled explicitly. That such compensation is possible, however, will be illustrated after the following relation between the time-varying subglottal pressure, $P_s(t)$, and the time rate of change of air volume exiting through the glottis, $U_g(t)$, is stated:

$$U_g(t) = \sqrt{P_s(t)/K(t)} \tag{3}$$

where K(t) is a time varying parameter that depends on the effects of the currently active vocal tract gestures on the tract's resistance to air flow. The parameter K is greater the more closed the vocal tract in both the glottal and supraglottal regions.

Equating volume flow from the glottis with volume flow from the lungs over the long term, equation (3) allows a rationalization of experimental results showing identical subglottal pressure trajectories for different percent-vital-capacity or lung volume trajectories. For example, equation (3) shows that utterances with different average K values can display the same subglottal pressure trajectories but be accompanied by different lung volume trajectories (time integrals of glottal volume velocity). For the same exponential P_s trajectory, the utterance associated with smaller average K (a more open vocal tract) will show a steeper descent in lung volume. Such patterns are displayed by subject RC in Gelfer, Harris and Baer (1987, p. 426), who produced the same subglottal pressure trajectories for repetitions of /fa/ and /ma/, yet showed greater lung volume decreases for repetitions of /fa/. The decline in lung volume in the case of /fa/ was faster than that for /ma/ because the overall resistance coefficient, K, was smaller in the case of /fa/. However, there must have been compensatory differences in the active muscle forces exerted on the thoracic cavity for repetitions of /fa/ to produce the same subglottal pressure trajectory as that for /ma/, since the /fa/ and /ma/ sequences were produced with different percent-vital-capacity trajectories.[2] A different yet related pattern was displayed by subject LB in Gelfer *et al.* (1987, p. 426), who showed identical lung volume trajectories as well as identical subglottal pressure trajectories across the /fa/ and /ma/ utterances. From the perspective of our present analyses, this subject seems to have produced the two utterance types with vocal tract shapes whose average overall resistances (and, hence, average K values) were roughly identical.

2.2. *Transglottal pressure* (P_t)

It is necessary to define an aerodynamic tract variable that is activated only during voiced obstruent production. In the case of voiced stop consonants in English, for example, the transglottal pressure decrease is slower than one would expect for a fixed-shape, rigid-wall-tube model of the supralaryngeal vocal tract. This slowing of the rate of transglottal pressure decrease can have the effect of allowing voicing to continue throughout the stop. We have chosen to use transglottal pressure (P_t) as a tract variable in order to allow active maintenance of voicing in these situations. In the current model, P_t trajectories are shaped according the following second-order dynamics:

$$\ddot{P}_t + \beta\dot{P}_t + \omega^2(P_t - P_{t0}) = 0, \quad (4)$$

where P_{t0} = target, ω = natural frequency, and β = damping. Whereas subglottal pressure "gestures" were defined over a relatively long time interval (i.e., at the time scale of multisyllable/word sequences, breath groups, etc.), the P_t gestures are activated over time intervals that are more locally defined (i.e., at the phonemic or syllabic time scale of voiced obstruents). Also, the transglottal pressure trajectories are described using a second-order equation because these trajectories change curvature through their evolution. A first-order differential equation was sufficient for subglottal pressure because the observed trajectories appear exponential in the long term.

At present, simulations are restricted to cases in which only one supralaryngeal constriction exists at any given point in time. In such situations, transglottal pressure and intraoral pressure would have been effectively equivalent choices for tract variables during voiced obstruent productions. Here we define intraoral pressure as the pressure in the cavity that is immediately upstream from the front cavity created by the uppermost constriction, i.e, the pressure in the back cavity in the case of a single supralaryngeal constriction. Since this back cavity is also immediately

---

[2] Note that we have not modeled explicitly the compensatory differences in the active muscle forces that are required to produce a given total lung force (and, hence, subglottal pressure) at different percent-vital-capacity thoracic volumes. We chose not to include these processes in our simulations because, given the current computational model, they would have no effect on either the supraglottal aerodynamics or acoustics as long as subglottal pressure trajectories remain unchanged.

downstream from the tracheal tube, the functional equivalence between transglottal and intraoral pressures is due to the fact that subglottal pressure had already been chosen as a tract variable on independent grounds and, in this context, transglottal pressure is simply subglottal pressure minus intraoral pressure. However, in cases where there is more than one supralaryngeal constriction at a particular point in time (e.g., due to coarticulation), the identification and selection of the appropriate cavity for defining intraoral pressure becomes more ambiguous, and it is simpler to work with transglottal pressure. Finally, there may be instances other than voiced obstruents when transglottal pressure (or intraoral pressure) is controlled using changes in upper vocal tract volume, such as clicks, injectives, and ejectives.

There are several possible mechanisms that could account for slowing the decrease in transglottal pressure during voiced obstruent production. Rothenberg (1968) enumerates these as: passive expansion of the supraglottal cavity behind the constriction due to the compliance of the vocal tract walls; active enlargement of the same cavity, as would occur with larynx lowering; and incomplete velopharyngeal closure. There is evidence that the larynx lowers and that the vocal tract walls expand, either through increases in wall compliance or through changes (increases or decreases) in muscular activity, during voiced stop production (Kent and Moll, 1969; Perkell, 1969; Bell-Berti, 1975; Westbury, 1983).

In addition to supralaryngeal volume and wall compliance, there are other variables that affect transglottal pressure during obstruent production. In the case of a constricted supraglottal vocal tract, these variables include the total force on the lungs, glottal width, and supraglottal constriction area. There appears to be little evidence that total lung force is used to control transglottal pressure in a constricted vocal tract situation. It would be counter-productive, in fact, to increase total lung force (and, hence, subglottal pressure) to maintain a minimum transglottal pressure, since air would accumulate faster in a constricted supralaryngeal tract with such an increase, and intraoral pressure would therefore increase without a concomitant increase in transglottal pressure. Thus, we did not use force on the lungs as an articulator to actively control transglottal pressure in our task dynamic simulations. Further, since there is evidence that glottal width is not used to maintain transglottal pressure during constricted supralaryngeal vocal tract situations (Lisker and Baer, 1984), glottal width was also not used as an articulator to actively control transglottal pressure.

We were left, therefore, with only a single articulator with which to actively control the aerodynamic tract variable P_t in our simulations of single supralaryngeal constrictions. This "articulator" corresponds to the volume of the supralaryngeal back cavity (V_{uv}), and accounts for all volume changes not due to changes in the current constriction location. Thus, V_{uv} lumps together several means of changing this volume, such as lowering the larynx, raising the velum, and expanding the vocal tract walls, in much the same way that total force on the lungs (F_l) lumps many physiologically independent components.[3] In future simulations, the supralaryngeal articulators will be allowed to participate in the active control of transglottal pressure, through their effects on upper vocal tract constriction location(s). That is,

---

[3] While the wall impedance elements are assumed constant in Fig. 1, the supralaryngeal volume articulator includes the effects of volume change due to increasing wall compliance. Since the acoustic consequences of the different means of volume change are not the same, the performance of the model may diverge from human speaker data during simulations and production, respectively, of stop releases.

a component of the active control of P_t will be the supraglottal constriction location(s) of the currently active supralaryngeal gesture set.

Finally, our informal observations of simulations show that small changes in supralaryngeal constriction area during frication can have marked effects on the transglottal pressure evolution during the production of (voiced) fricatives. This means that motions of the original supralaryngeal vocal tract articulators also affect transglottal pressure through their effects on the constriction-degree tract variables as well as on the constriction-location tract variables. Thus, in future simulations that allow the original model's supralaryngeal articulators (e.g., lips, tongue tip, tongue body) to participate in the active control of transglottal pressure, these articulators will affect P_t through their effects on both the degree and location of the currently active supraglottal constriction. That is, the supraglottal constriction degree(s) and location(s) of the currently active supralaryngeal gesture set will become components of the active control of P_t.

2.3. *Delta virtual F_0 (ΔF_{0v})*

A quantity related to fundamental frequency (F_0) was needed as the third new tract variable. There are many articulators that control fundamental frequency because of the complexity of the larynx and its relation to other structures. However, experience gained from the two-mass model of laryngeal vibration and phonation was useful in this regard (Ishizaka and Flanagan, 1972) and, in the current task-dynamic model, the complicated geometric and tissue dynamic parameters of the larynx are represented by a single generalized glottal tension parameter (T_g) that is the analog of the Q-factor of Ishizaka and Flanagan (1972). Other important factors in fundamental frequency determination are transglottal pressure (and, hence, the force applied to the lungs and supralaryngeal tract volume), and supralaryngeal vocal tract load (e.g., Titze, 1989; Ishizaka and Flanagan, 1972).

The actual value of F_0 at any given time in the simulation is a function of vocal fold tension (T_g) and transglottal pressure (P_t) (Recall that, among other factors, P_t is itself a function of subglottal pressure [P_s] and, in the case of constricted vocal tracts, supralaryngeal volume [V_{uv}]). Thus, F_0 is affected by the evolving state of a host of articulator and aerodynamic variables. In the implementation of task dynamics here, we worked under the assumption that the fundamental-frequency-related tract variable that is controlled is ΔF_{0v}, the deviation of a so-called *virtual* fundamental frequency (F_{0v}) from its ongoing baseline level (F_{0vB}). The value of F_{0vB} at any given time in the simulation is determined by the ongoing subglottal pressure trajectory, a set of assumed constant values of T_g and G_w appropriate for voicing, and an assumed open vocal tract (under which conditions P_t simply equals P_s). Thus, F_{0vB} is meant to provide a default measure of the actual fundamental frequency that would occur during sonorants, when there is a small acoustic load due to supralaryngeal constrictions, and there are no added pitch accents; F_{0v} then, is the fundamental frequency that would occur in sonorant conditions, but with pitch accents. Since the set of assumptions made in the computation of F_{0vB} will obviously be violated during simulated speech sequences, the trajectories for actual and virtual fundamental frequencies will be similar during sonorants, but may be quite dissimilar during obstruents. Whether ΔF_{0v} turns out to be a good choice for a fundamental-frequency-related tract variable remains a topic for further research. For instance, it is conceivable that F_{0vB} can be defined in a manner in which subglottal pressure variations are factored out.[4]

In our simulations, the dynamics of ΔF_{0v} is defined according the following second-order equation of motion:

$$(5)\qquad \Delta\ddot{F}_{0v} + \beta\Delta\dot{F}_{0v} + \omega^2(\Delta F_{0v} - \Delta F_{0vT}) = 0,$$

where $\Delta F_{0v} = (F_{0v} - F_{0vB})$, ΔF_{0vT} = target deviation of F_{0v} from the current baseline level, β = damping, and ω = natural frequency. The ΔF_{0v} gestures are activated over time intervals that are defined in a relatively local manner, i.e., at the time scales of syllables or words. The articulators used to control these gestures are generalized laryngeal tension (T_g) and force on the lungs (F_l).

3. Simulations

Here we report on the simulations of two utterances using the flow variable additions to task dynamics. One, /ədæ/, contains a voiced stop, and the other, /ətæ/, contains an unvoiced, aspirated stop. For both /ədæ/ and /ətæ/, the vocal tract goes from a neutral configuration to an alveolar constriction using the tongue-tip tract variables and then the tongue body and lips are used to go into the final vowel. In order to focus on the consequences of changes in the control of the aerodynamic tract variables during these two sequences, the kinematics of the supralaryngeal articulators were constrained to be the same in both cases. The resulting constriction area is shown as a function of time in Figs 2(a) and 3(a). Due to the brevity of both sequences, subglottal pressure was not actively controlled and, therefore, did not decay exponentially during the simulations.

To synthesize voicing through the alveolar closure for /ədæ/, the transglottal pressure tract variable was activated over an interval of 50 ms to 130 ms into the utterance with a target of 4 cm-H_2O, critically damped dynamics, and a 10 Hz natural frequency. The results, in terms of the transglottal pressure tract variable (P_t), its single supralaryngeal-tract-volume articulator (V_{uv}), and speech acoustics are shown in Figs 2(b), (c) and (d). The dashed line in Fig. 2(b) indicates the trajectory that P_t would have been if it had not been controlled. Note that the change of V_{uv} of nearly 10 cm^3 (here, a volume change from 50 cm^3 to about 60 cm^3) is close to the estimates derived from measurements of articulatory movement for medial alveolar stops (Westbury, 1983).

Because the stop was unvoiced in the utterance /ətæ/, glottal width dynamics was specified but transglottal pressure dynamics was not. The glottis was abducted from 30 ms to 60 ms with a width target of 0.4 cm (corresponding to an open area of 0.63 cm^2), critical damping, and a natural frequency of 10 Hz. The glottis was adducted from 50 ms to 140 ms using the default target area for voicing of 0.07 cm^2, a 0.4 damping ratio, and a natural frequency of 4 Hz. The simulated glottal area trajectory is shown in Fig. 3(b). Because there is no target specified for transglottal

---

[4] In general, the choice of tract variable can change depending upon the biases of the modeler. For example, instead of choosing transglottal pressure as a tract variable, one might choose amplitude of voicing. This tract variable would recruit glottal width, supralaryngeal tract volume, V_{uv}, and generalized laryngeal tension, T_g, as articulators.

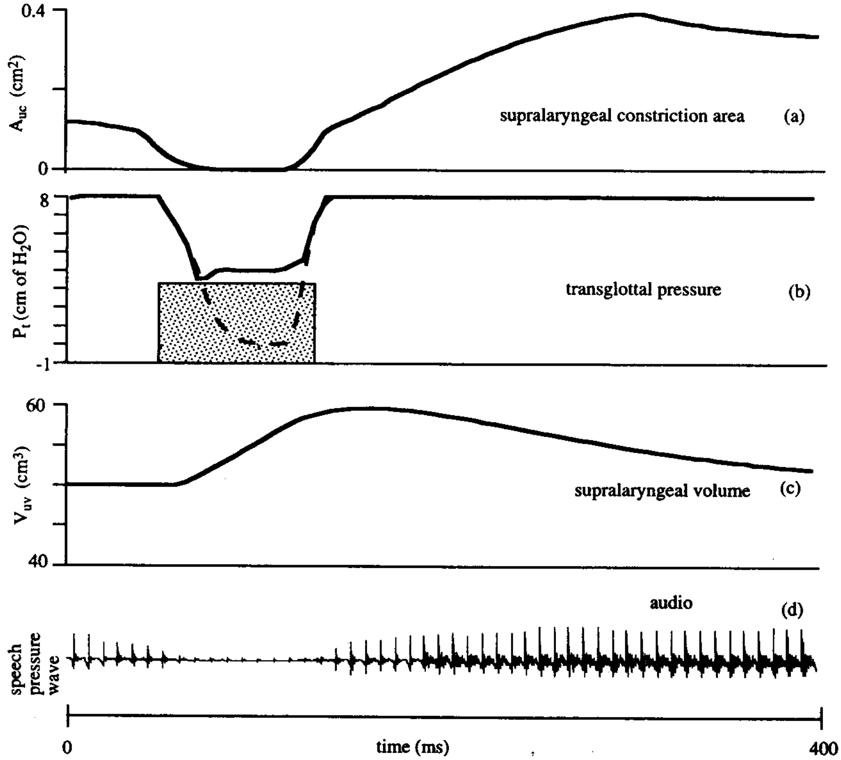

Figure 2. (a) Supralaryngeal constriction area, A_{uc}, for the utterance /ədæ/. (b) Transglottal pressure, P_t, for utterance /ədæ/. The horizontal extent of the box indicates activation interval and the vertical extent indicates target. The solid line is the P_t with activation and the dashed line is what P_t would have been without activation. (c) Supralaryngeal vocal tract volume, V_{uv}, for the utterance /ədæ/. (d) Synthesized speech pressure wave for utterance /ədæ/.

pressure (P_t), this quantity goes to zero during supraglottal closure and glottal abduction [Fig. 3(c)].

It is known that there is a transient initial elevation of fundamental frequency at the release of voiceless stops, and that this elevation is due to heightened tension in the vocal folds (Löfqvist, Baer, McGarr and Story, 1989). There are two alternative accounts for these increases in F_0 and vocal fold tension at voiceless stop releases. The first is that F_0 is actively elevated in order to provide an additional acoustic cue for voicelessness, and that vocal fold tension is used in part to create this elevation. The second account posits that the elevated F_0 is an indirect, passive consequence of "carryover" residual tension in the vocal folds as they relax from a tensed, abducted state for devoicing into an adducted, less tense state for voicing. While we admit that the latter account may be the more elegant one, our preliminary simulation of the utterance /ətæ/ was consistent with the former account, and we modeled devoicing using both an active increase in glottal aperture and an active elevation of

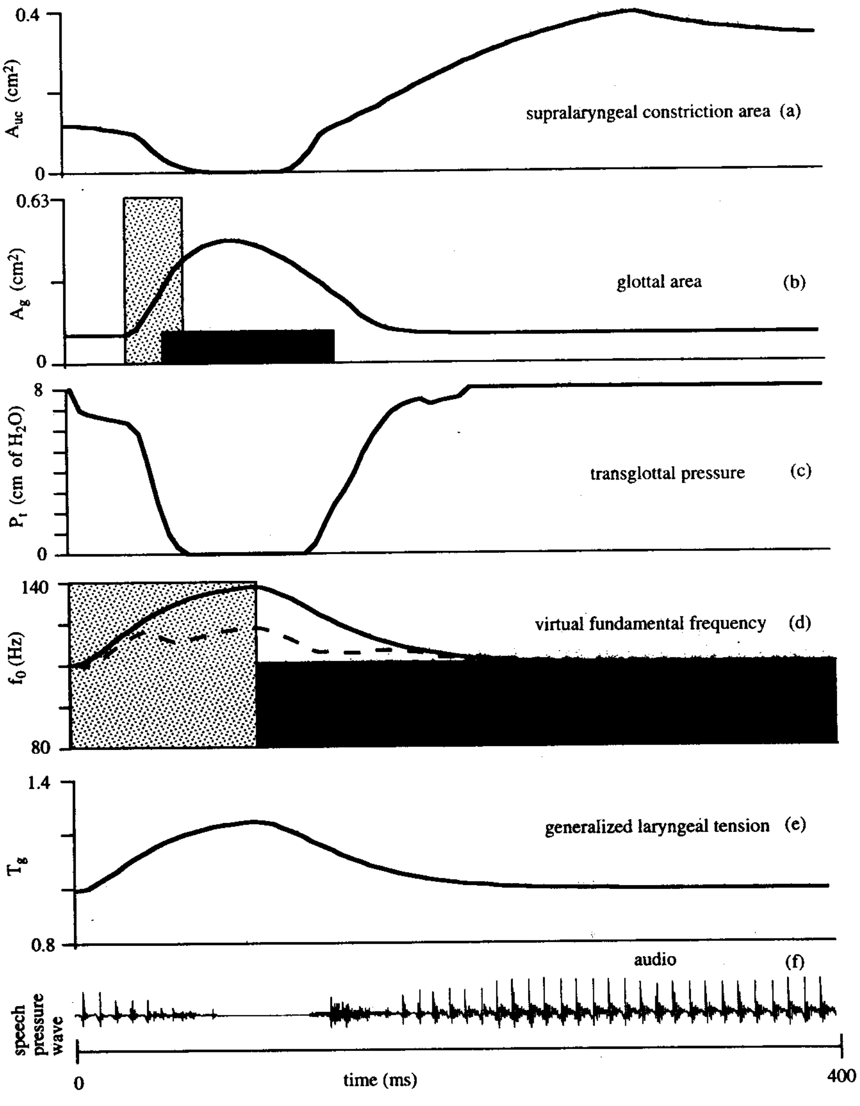

Figure 3. (a) Supralaryngeal constriction area, A_{uc}, for the utterance /ətæ/. (b) Glottal area, A_g, for utterance /ətæ/. The horizontal extents of the boxes indicate activation interval and the vertical extents indicate target. (c) Transglottal pressure, P_t, for utterance /ətæ/. (d) Virtual fundamental frequency, F_{0v}, (solid line) and fundamental frequency, F_0, (dashed line) for utterance /ətæ/. The horizontal extents of the boxes indicate activation interval and the vertical extents indicate target. (e) Generalized laryngeal tension, T_g, for the utterance /ətæ/. (f) Synthesized speech pressure wave for utterance /ətæ/.

F_0.[5] More specifically, delta-virtual-F_0 (ΔF_{0v}) was activated in the interval from 0 ms to 90 ms using a target of 40 Hz above F_{0vB} (virtual-F_0 baseline), critical damping, and a natural frequency of 8 Hz. It was also activated from 90 ms to 400 ms using a target of 0 Hz above F_{0vB}, critical damping, and a natural frequency of 6 Hz. Since F_{0vB} stayed approximately constant at a value of 120 Hz in both the /ədæ/ and /ətæ/ simulations, this sequence of ΔF_{0v} gestures was associated with F_{0v} targets of approximately 160 Hz and 120 Hz, respectively. Fig. 3(d) displays the resulting virtual F_0 (F_{0v}; solid line) and actual F_0 (dashed line) trajectories, respectively. (Note that, in this figure, only the portions of the F_0 plot during voicing are meaningful.). With the total lung force (F_l) and generalized laryngeal tension (T_g) weighted equally (see Saltzman and Munhall, 1989, for details regarding the articulatory weighting process), the ΔF_{0v} tract variable uses T_g exclusively [Fig. 3(e)], and P_s stayed approximately at its initial value of 8 cm-H_2O throughout. With different articulatory weightings, F_l (and hence, P_s) can be brought into play, but the present simulation indicates how relatively insensitive subglottal pressure is to fundamental frequency changes. The correct weightings of these articulators should be determined experimentally. The output simulated speech acoustics for /ətæ/ is shown in Fig. 3(f).

4. Summary

We have attempted to generalize the task dynamic model of speech production. This model was originally formulated to simulate the coordinated motion patterns of the (mainly) supralaryngeal articulators that largely determine the vocal tract's time-varying acoustic filter characteristics. We have added to this model aerodynamic and laryngeal variables that, in cooperation with other variables, determine the time-varying source properties observed during spoken utterances. The new aerodynamic tract variables were subglottal pressure, transglottal pressure, and delta-virtual-F_0. The articulator list was expanded to include the total force on the lungs, generalized vocal fold tension, and supraglottal vocal tract volume (glottal width was already used as an articulator in previous, non-aerodynamic versions of the model). More realistic articulatory synthesis of speech with the task dynamic model is now possible.

This work has profited from discussions with Cathe Browman and Louis Goldstein. This work was supported by the NIH through grants DC-01247 and DC-00121 to Haskins Laboratories.

---

[5] At issue is the origin of the increase in laryngeal tension during the laryngeal devoicing gesture. According to the first account (i.e., active determination of F_0), devoicing is accomplished using two tract variables (glottal width [G_W], delta-F_0 [ΔF_{0v}]), and laryngeal tension is increased in service of the tract-variable goal of increasing F_0. According to the second account (i.e., passive determination of F_0), devoicing would entail a single "voicing/devoicing" tract variable that recruited the articulator variables of laryngeal width, vocal fold tension, total lung force, and upper vocal tract volume in a coordinated manner. In particular, laryngeal tension would be increased as the vocal folds are abducted in service of the tract-variable goal of devoicing. We are currently working to develop a set of new simulations consistent with the latter scenario, using a new voicing amplitude tract-variable that would subsume and replace the current transglottal pressure and glottal aperture tract variables. The former, as described in the section Transglottal pressure (P_t), was designed to allow the continuation of voicing during oral closure by enlarging the oral-pharyngeal cavity (i.e., by recruiting the upper-vocal-tract-volume articulator variable), and the latter has been used as a relatively simplistic kinematic specification of voicing/devoicing.

References

Bell-Berti, F. (1975) Control of pharyngeal cavity size for English voiced and voiceless stops. *Journal of the Acoustical Society of America*, **57**, 456–461.

Browman, C. P. & Goldstein, L. G. (1990) Gestural specification using dynamically-defined articulatory structures. *Journal of Phonetics*, **18**, 299–320.

Gelfer, C. E., Harris, K. S., Collier & Baer, T. (1983). Is Declination actively controlled? In I. Titze and R. S. Scherer (Eds). *Vocal Fold Physiology*. Denver: The Denver Center for the Performing Arts.

Gelfer, C. E., Harris, K. S. & Baer, T. (1987) Controlled variables in sentence intonation. In T. Baer, C. Sasaki, & K. Harris (Eds.) *Laryngeal Function in Phonation and Respiration*. Boston: College-Hill.

Hixon, T. J. (1987) *Respiratory Function in Speech and Song*. Boston: College-Hill.

Ishizaka, K. & Flanagan, J. L. (1972) Synthesis of voiced sounds from a two-mass model of the vocal cords. *Bell System Technical Journal*, **51**, 1233–1268.

Kent, R. D. & Moll, K. L. (1969) Vocal-tract characteristics of the stop cognates. *Journal of the Acoustical Society of America*, **46**, 1549–55.

Liberman, M. & Pierrehumbert (1984) Intonational invariance under changes in pitch range and length. In M. Arnoff & R. T. Oehrle (Eds.) *Language Sound Structure*, Cambridge, Mass.: MIT Press.

Lisker, L. & Baer, T. (1984) Laryngeal management at utterance-internal word boundary in American English. *Language and Speech*, **27**, 163–171.

Löfqvist, A., Baer, T., McGarr, N. S. & Seider Story, R. (1989) The cricothyroid muscle in voicing control, *Journal of the Acoustical Society of America*, **85**, 1314–1321.

Mattingly, I. G. (1990) The global character of phonetic gestures. *Journal of Phonetics*, **18**, 445–452.

Mermelstein, P. (1973) Articulatory model for the study of speech production. *Journal of the Acoustical Society of America*, **53**, 1070–1082.

Müller, E. M. & Brown, W. S. (1980) Variations in the supraglottal air pressure waveform and their articulatory consequences. In N. Lass (Ed.), *Speech and Language: Advances in Basic Research and Practice*, **4**, pp. 318–389. New York: Academic Press.

Ohala, J. J. (1990) Respiratory activity in speech. In W. Hardcastle & A. Marchal (Eds.), *Speech Production and Speech Modeling*, pp. 23–53. Dordrecht: Kluwer.

Perkell, J. S. (1969) *Physiology of Speech Production*. Cambridge, Mass.: MIT press.

Rothenberg, M. (1968) *The Breath-Stream Dynamics of Simple-Released-Plosive Production*. Basel: S. Krager.

Rubin, P., Baer, T. & Mermelstein, P. (1981) An articulatory synthesizer for perceptual research. *Journal of the Acoustical Society of America*, **70**, 321–328.

Saltzman, E. (1986) Task-dynamic coordination of the speech articulators: A preliminary model. *Experimental Brain Research*, Series **15**, 129–144.

Saltzman, E. & Kelso, J. A. S. (1987) Skilled actions: A task-dynamic approach. *Psychological Review*, **94**, 84–106.

Saltzman, E. & Munhall, K. (1989) A dynamical approach to gestural patterning in speech production. *Ecological Psychology*, **1**, 333–382.

Titze, I. R. (1989) On the relation between subglottal pressure and fundamental frequency in phonation. *Journal of the Acoustical Society of America*, **85**, 901–906.

Westbury, J. (1983) Enlargement of the supraglottal cavity and its relation to stop consonant voicing. *Journal of the Acoustical Society of America*, **91**, 2903–2910.

Appendix: Formal Mathematical Relations

The transformations between the tract variables and articulatory variables are no longer purely geometric with the inclusion of aerodynamics. In fact, the new tract variables depend upon the evolution of the aerodynamic system, which in turn, depends upon the articulatory state. The relation between the evolution of tract variables and articulatory variables must incorporate the evolution of intermediate aerodynamic variables that are neither tract variables nor articulators.

Let: $\mathbf{A}$ = vector of articulator state (i.e., position and velocity),

$\mathbf{P}$ = vector of tract-variable state,

$\mathbf{X}$ = vector of aerodynamic state (intermediate aerodynamic variables, those aerodynamic variables that are neither tract variables nor articulatory variables)

Task dynamics can be expressed in first order as:

$$\dot{\mathbf{P}} = \mathbf{T}(\mathbf{P}) \tag{A1}$$

The transformation from articulators to tract variables is stated as:

$$\mathbf{P} = \mathbf{H}(\mathbf{X}, \mathbf{A}) \tag{A2}$$

The aerodynamic evolution can be expressed as:

$$\dot{\mathbf{X}} = \mathbf{G}(\mathbf{X}, \mathbf{A}) \tag{A3}$$

Differentiating Equation (A2) with respect to time:

$$\dot{\mathbf{P}} = \frac{\partial \mathbf{H}}{\partial \mathbf{X}} \dot{\mathbf{X}} + \frac{\partial \mathbf{H}}{\partial \mathbf{A}} \dot{\mathbf{A}} \tag{A4}$$

Substituting Equation (A4) into Equation (A1), using Equation (A3):

$$\dot{\mathbf{P}} = \mathbf{T}(\mathbf{P}) = \frac{\partial \mathbf{H}}{\partial \mathbf{X}} \mathbf{G}(\mathbf{X}, \mathbf{A}) + \frac{\partial \mathbf{H}}{\partial \mathbf{A}} \dot{\mathbf{A}} \tag{A5}$$

Rearranging and solving for the time evolution of **A**:

$$\dot{\mathbf{A}} = \left(\frac{\partial \mathbf{H}}{\partial \mathbf{A}}\right)^{*} \left\{\mathbf{T}(\mathbf{P}) - \frac{\partial \mathbf{H}}{\partial \mathbf{X}} \mathbf{G}(\mathbf{X}, \mathbf{A})\right\} \tag{A6}$$

where the * indicates pseudo-inverse.

Finally Equation (A2) can be used to obtain:

$$\dot{\mathbf{A}} = \left(\frac{\partial \mathbf{H}}{\partial \mathbf{A}}\right)^{*} \left\{\mathbf{T}(\mathbf{H}(\mathbf{X}, \mathbf{A})) - \frac{\partial \mathbf{H}}{\partial \mathbf{X}} \mathbf{G}(\mathbf{X}, \mathbf{A})\right\} \tag{A7}$$

Solving Equation (A3) simultaneously with Equation (A7) simulates the task dynamics of vocal tract geometry and aerodynamics. Note that with the aerodynamic variables fixed (i.e., **X** fixed) the evolution Equation (A7) reduces to that derived in Saltzman & Munhall, Appendix A (1989).

Written explicitly, system of Equations (A3) that were used here to describe the evolution of the intermediate flow variables were (see Fig. 1):

$$\begin{aligned} X_1 &= Q_{uw} \\ X_2 &= \dot{Q}_{uw} = U_{uw} \\ X_3 &= Q_{ua} \end{aligned} \tag{A8}$$

where Q denotes the volume displacement of air held by a capacitive element.

$$\begin{aligned} \dot{X}_1 &= X_2 \\ \dot{X}_2 &= \left\{P_{uc} - R_{uv} \cdot X_2 - \frac{X_1}{C_{uw}}\right\} \Big/ L_{uw} \\ \dot{X}_3 &= U_g - U_{uc} - U_{uw} - \dot{V}_{uv} \end{aligned} \tag{A9}$$

where,

$$U_g \cdot R_g = P_s - P_{uc}$$

(A10) $$U_{uc} \cdot R_{uc} = P_{uc}$$

$$P_{uc} = \frac{Q_{ua}}{C_{ua}}$$

The resistive elements R_g and R_{uc} are nonlinear flow resistance due to both vorticity and viscosity. They depend on the areas A_g and A_{uc}, respectively. C_{ua} is the capacitance due to the compliance of air below the supralaryngeal constriction (area A_{uc}) in the supralaryngeal tract, and depends, therefore, on V_{uv}.