Aerodynamic modeling at UCLA

Patricia Keating

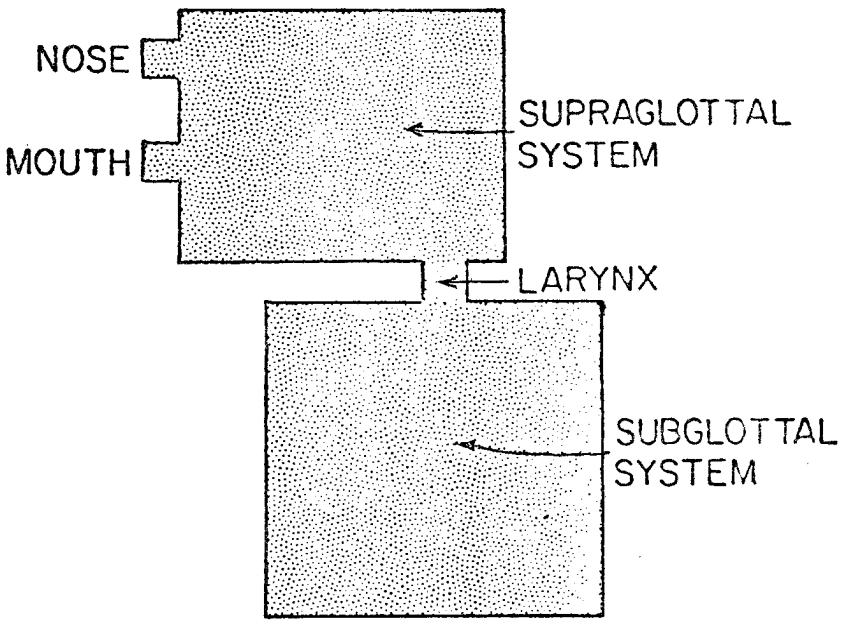
0. Introduction

This report describes the aerodynamic vocal tract simulation currently implemented on the Phonetics Lab's DEC PDP-11/23 computers. The major use of this simulation to date has been in the study of consonants, particularly stop consonant voicing; examples of such work can be seen in other papers in this volume. First I will discuss qualitatively the aspects of speech aerodynamics being modeled, and the kinds of data that such a model will provide. Then I will describe the particular kind of model used in our simulation, outlining some of the basic ideas for phoneticians. Next I will document, in general terms, how the UCLA computer program is used. Finally, I will give an example of work done that will illustrate the use of the model and the kind of results that can be obtained.

1. Speech aerodynamics

Figure 1 shows a schematic of the vocal tract as relevant to a discussion of speech aerodynamics. To a first approximation, the vocal tract consists of two soft-walled cavities, the lungs and the mouth. They are separated from each other by a constriction formed by the vocal cords, and separated from the atmosphere by constrictions at the velum and/or mouth opening. The driving pressure generated by the respiratory system results in airflow from the lungs to the atmosphere via the glottis and one or both of the other openings. Over the course of an utterance, the volumes of both cavities, dimensions of various constrictions, and the mechanical properties of the vocal tract walls may be controlled by a speaker, thereby producing the familiar variations in air pressures and flows which characterize the speech wave.

Figure 1.



The circumstances of air pressures and flows are particularly important in considering vocal cord vibration. A sufficient flow of air through the glottis, from the lungs to the pharynx, is crucial to the occurrence of voicing. According to the myoelastic-aerodynamic theory of phonation (van den Berg 1958), the vocal cords will oscillate when they are suitably adducted and tensed, and when there is a sufficient airflow across them. Such airflow will occur when the pressures above and below the larynx are different: in the usual case, when the pressure in the oral cavity is lower than the pressure in the subglottal system. Air will flow from the region of higher pressure to the region of lower pressure. Calculation of air pressures above and below the larynx, then, will indicate whether voicing could occur for a particular laryngeal state. Since subglottal pressure is relatively constant for most of an utterance, oral pressure is generally the major determinant of that transglottal pressure difference.

Oral pressure depends on how much air is flowing through the glottis into the oral cavity, how much air is flowing through any oral constriction (or the nose) out of the oral cavity, and how much the walls of the oral cavity 'give' in the face of rising oral pressure, counteracting such a rise. Therefore, for example, a larger glottal opening, a smaller oral opening, and stiffer oral cavity walls will all tend to contribute to a higher oral pressure, as contemplation of Figure 1 should make clear. During a vowel, with its large oral opening, oral pressure is essentially the same as atmospheric pressure (taken as a baseline of 0 pressure), so the pressure drop across the larynx (the difference between subglottal and oral pressures) is equal to the subglottal pressure. During a stop, the vocal tract is suddenly and fully occluded, preventing any escape of air and causing pressure behind the occlusion to build rapidly. In this case, then, the pressure drop across the larynx may be small or nonexistent. Consonant release allows the built-up air to escape, and oral pressure drops accordingly, giving a transglottal pressure drop.

2. The analog circuit model

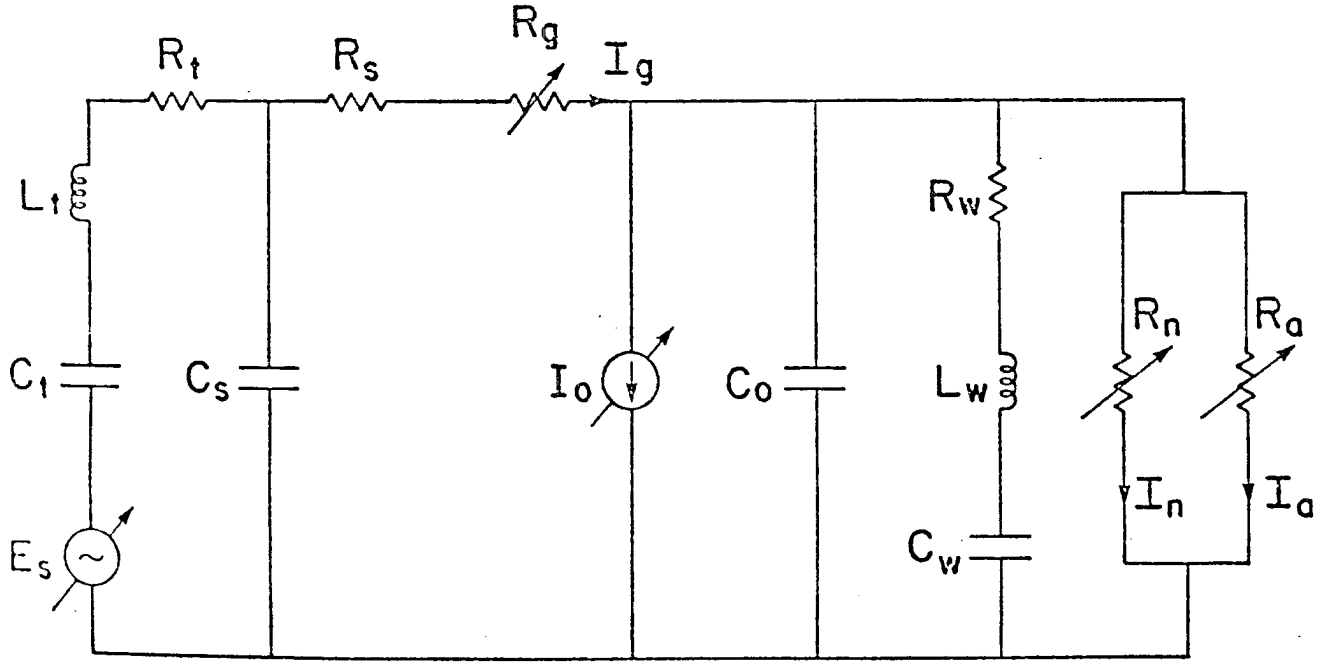
What does it mean to construct a model of speech aerodynamics? In modeling, an idealized representation of a system is constructed to further our understanding of that system. This technique is quite general, and is used in many areas of linguistics besides phonetics. In our case, the simulation is of a physical system, and is numerical, and therefore can be rather precise. One advantage of modeling as a research enterprise is that it forces us to be explicit about our assumptions about the system and about our account of it. Another advantage is that it encourages and focuses the search for new patterns in data.

One way to model speech production would be to build a physical version of a vocal tract, with movable articulators and an air source. This sort of modeling has certainly been tried in the past, but it is less common nowadays than more convenient and more precise numerical simulations of the vocal tract. Such models consist of numerical equations or functions that encode crucial properties of the vocal tract. In some cases these representations are taken directly from the physical system. The aerodynamic model of, for example, Ohala (1976) is of this basic type: airflows and air pressures are calculated directly from things like volume of the vocal tract, and muscular forces.

A less direct kind of model is the electrical analog, which represents the vocal tract as being like a circuit. A circuit is a physical device through which electrical charge can flow; the flow of charge is called current. Various devices or elements may be part of a circuit and will influence this flow in certain known ways. One advantage of a circuit model is that the properties of circuits have been well-studied. If we simply want to know how a circuit would respond to certain inputs, we do not actually need to build it. Rather, we can use circuit theory to devise a set of equations which will describe the behavior of the circuit. Circuit theory also provides mechanical and acoustical correspondances for electrical circuits. Thus a circuit can be designed to represent a non-electrical system, such as a vocal tract; the circuit is then an analog of the vocal tract. The steps in electrical analog modeling are to design a circuit which represents crucial aspects of the vocal tract; formulate the equations describing the behavior of that particular circuit; determine the outputs from those equations for a variety of conditions of interest; interpret those results as phonetic events.

Figure 2 illustrates the circuit used in our work to model vocal tract aerodynamics for low frequency events. This model of the breath-stream control mechanism is derived from work of Rothenberg (1968), who actually built and used a physical circuit. A computer simulation of the circuit model is described by Muller and Brown (1980); essentially the current implementation is described briefly by Westbury (1983). The electrical current moving through the circuit is the analog of volume velocity airflow in the vocal tract. The electrical voltage at any point in the circuit is the analog of air pressure in the vocal tract. The circuit contains five kinds of elements, described below, (plus wire connecting them). These elements, plus their arrangement as a group, represent the crucial aspects of the vocal tract for aerodynamic events.

Figure 2.



The elements labeled E_s and I_o are sources of electrical energy in the circuit: the first introduces voltage, and the second introduces current. I stands for current; the other three I's in the circuit simply label the current coming out of an element. The other three kinds of elements respond to electrical energy, rather than introduce it. All the elements labeled R are resistors, which dissipate energy as heat. The elements labeled C are capacitors, which store energy as electrical energy. The elements labeled L are inductors, which store energy as magnetic energy and in fact introduce magnetic fields into the circuit. Some elements have diagonal arrows through them, which means that their values change over time. Basically, the current starts in the lower left corner, induced by the E_s source, and moves upward through C_t, L_t, and R_t, at which point there is a fork with branches going to both R_s and C_s. Below C_s is a horizontal wire with nothing on it; this is ground or zero where all the charge ends up. Beyond R_s there are several more elements and branches. Every time there is a branching, some current goes one way and some another.

As a reference point, R_g in the top middle represents the resistance to flow presented by the glottis. Basically the glottis is acting like a valve whose opening size can be varied. Everything to the left of R_g represents the subglottal system, and everything to the right of it represents the supraglottal system. The voltage (pressure) source E_s represents the respiratory muscular force, which causes inhalation before an utterance, and counters a drop in subglottal pressure later in an utterance. A set of a capacitor, an inductor, and a resistor in a row, as with C_t, L_t, R_t, and C_w, L_w, R_w, represents the properties of vocal tract walls: in the first case for the trachea and other subglottal cavities, and in the second case for the supraglottal cavity. The capacitor represents the stiffness of the walls; this then includes in the subglottal case what we normally describe as "elastic recoil" and in the supraglottal case, "tenseness". The inductor represents the mass of the walls, and the resistor represents mechanical heat loss inside the walls. The other two capacitors, C_s and C_o, represent the volume of air inside the subglottal and oral cavities, respectively. The volume of the oral cavity is one of the ways in which place of articulation will be reflected. R_s represents other subglottal losses (i.e. it is essentially a fudge factor). The other two resistors, R_n and R_a, are, like R_g, valves: R_n to the nasal cavity and R_a to the atmosphere. If the velum is completely closed so that it totally blocks the flow of air, then it is as if R_n were not in the circuit. Currently, R_n is <u>not</u> represented on our implementation. R_a, representing the oral constriction, is given as three dimensions which depend on place of articulation and vary over time as the constriction is formed or released. The other time varying element, the current source I_o, represents in a single, undifferentiated way, various possibilities for active expansion (or contraction) of the oral cavity, e.g. advancement of the tongue root or jaw movement.

To summarize the elements of the circuit model:

Element	Definition
E_s	Voltage source: respiratory muscle force
C_t	Capacitance (compliance) of tracheal walls; in part depends on surface area of walls
L_t	Inductance (mass) of tracheal walls

R_t Mechanical resistance of tracheal walls

C_s Capacitance (compliance) of air in subglottal system; in part depends on volume of cavity

R_s Other subglottal mechanical resistance

R_g Total glottal resistance; in part depends on size of glottal opening

L_g Reactive component of glottal impedance

I_o Supralaryngeal current source: change in cavity size

C_o Capacitance (compliance) of air in oral cavity; in part depends on volume of cavity

C_w Capacitance (compliance) of oral tract walls; in part depends on surface area of walls

R_w Mechanical resistance of oral tract walls

L_w Inductance (mass) of oral tract walls

R_n Total nasal constriction resistance; no nasal options currently implemented

R_a Total oral constriction resistance; in part depends on size of oral opening

Equally important in using the model are the voltages across the capacitors. Voltage is the electrical analog of pressure: here, either the pressure exerted by air within a volume (voltages across C_s and C_o) or by stretched walls (voltages across C_t and C_w). The two subglottal voltages C_s and C_t are involved in representing elastic recoil. The supraglottal voltage C_o is equal to oral pressure.

The arrangement of the elements in the circuit is dictated by the physical system the circuit is modeling in ways that are well understood by engineers, if not linguists. Given this circuit diagram, an engineer can derive a set of differential equations which describe its behavior. These equations can then be approximated by difference equations which compute changes in pressures and flows for each small unit of time. Such equations are suitable for programming on a small laboratory computer.

A caveat is in order about how voicing is represented in this model. It does not directly represent the vocal cords, so voicing cannot be seen directly as vocal cord vibration. Recall that the conditions on vocal cord vibration include position and tension of the cords, and airflow through them. The position can be fairly well represented via the glottal dimensions. The glottis is modeled as a three-dimensional space, essentially a valve-like opening, which does not vary during voicing. The cross-sectional area of the glottal slit approximates the average glottal area during a vowel, determined over the full duration of a

glottal period. This average area is used in simulations where the glottis is meant to be in a position that would allow voicing, and is used as the final value of any glottal adduction gesture. (The area we use is .04 cm^2; this value is justified by the fact that oral flow during a vowel is about 150 cm^3/sec while the pressure drop across the glottis is about 10 cm aq.) The tension of the vocal cords cannot be represented, since there are no vocal cords. Instead, tension is reflected in our estimation of how much airflow would be required to cause vibration: the tenser the cords, the greater the volume velocity required. We consider this aerodynamic condition in terms of the pressure drop across the larynx. There is some concensus in the literature that a pressure drop of 2 cm aq is necessary to sustain vibration (Lindqvist 1972, Ladefoged 1964, Ishizaka and Matsudaira 1972, Baer 1975), though initiating voicing may require a pressure drop twice as large (Baer 1975). Whenever our model is used to look at voicing, we can only look at pressure across the vocal cords. Typically we assume 'normal' tension of the cords, and the glottal dimensions just described, and determine when the pressure drop is great enough to allow voicing.

Other limitations of this model follow from various simplifications. The treatment of the subglottal system is sketchy. As noted above, various ways of changing the size of the oral cavity are not distinguished. Different parts of the tract are not distinguished, for example, the stiffness of the walls is assumed to be uniform, though obviously this is not the case. Furthermore, the model is valid only for low frequencies (large time intervals).

3. Use of the computer program

The FORTRAN computer program is called VTM, for Vocal Tract Model. Presently it runs on the Phonetics Lab's PDP-11/23 computer under a time-sharing system, without graphic capabilities. However, it produces output files that can be transported to the Phonetics Lab's speech system for displaying and printing.

The following is a complete list of all the input variables under control by the user. Many of them are discussed further below.

User likely to vary from run to run:

CO	volume of oral cavity, with constants
CW, RW, LW	oral wall properties, including area and stiffness
GWID, GLEN	constant glottal dimensions
OWID, OLEN	constant oral dimensions
VCS, VCT	elastic recoil factors in subglottal pressure

User unlikely to vary from run to run:

RHO	density of air
VISCOS	viscosity of air
C	speed of sound
CT, RT, LT	subglottal wall properties, including area and stiffness
CS	volume of subglottal cavity, with constants
RS	other subglottal losses
TINCR	time interval for calculations
NPPE	"sampling rate" for output of calculation results
GTURB	glottal turbulence factor (angle of entry)
OTURB	oral turbulence factor (angle of entry)

Functions over time:

GHEI	distance between vocal cords
OHEI	distance between oral articulators
ES	respiratory muscle force
IO	active change in volume of oral cavity

VTM provides default values, suitable for a medial labial stop, for all of the input variables that are constants. Some of these values are given below. None of them have to be changed in using VTM, but any or all of them may be. In contrast, the functions over time, for ES (respiratory muscle force), GHEI (distance between vocal cords), OHEI (distance between oral articulators), and IO (active expansion of oral cavity), do not have default values. Values at any number of points in time are specified by the user, and VTM linearly interpolates between those values.

Following are descriptions of and values for each of the input variables likely to be changed in VTM.

The variables CO, CW, RW, and COILW together encode the size of the oral tract, the surface area of the tract walls, and the stiffness (or tenseness) of the walls. Values appropriate to typical choices are given on the next page. GLEN represents the dimension of the glottis parallel to the flow of air, i.e., the vertical dimension. The glottal area in the horizontal plane is represented as a rectangle with one fixed and one changing dimension. The fixed dimension, GWID, is the larger of the two dimensions perpendicular to flow. The changing dimension, GHEI, is discussed below. The glottal area is calculated as the product of GWID and GHEI.

OLEN and OWID (and OHEI) are the equivalent oral constriction dimensions; again, OHEI is variable and described below. OLEN is the dimension parallel to flow, and OWID is the larger perpendicular dimension -- for a labial, the width across the lip opening. OHEI is the vertical opening dimension. Again, the area of the constriction is the product of OWID and OHEI. Values have been estimated for John Westbury's vocal tract but are schematic. Good values for OLEN (at the moment before release, from X-rays) are .2 cm for labials, .3 cm for alveolars, and .7 cm for velars. OWID has a small enough effect on outputs that we don't bother to change it.

VCS and VCT do not have such easy physical interpretations; they are variables in the subglottal system representing pressures that together control subglottal pressure via elastic recoil. Roughly, VCS is related to the volume of the subglottal cavity and determines the initial subglottal pressure for a simulation; VCT is related to the surface area of the stretched subglottal walls and determines the change in subglottal pressure, which usually rises when VCT is negative. The default values are for high and very slightly falling subglottal pressure, as for utterance-medial material. Subglottal pressures can be scaled down by varying both VCS and VCT. At the beginning of an utterance, subglottal pressure should rise -- this is done by letting VCS = 0, leaving VCT at default, and varying ES as described below. At the end of an utterance, subglottal pressure should fall -- this is done by leaving VCS and VCT at default, but varying ES as described below. Variation of ES is also described below for changes associated with stress.

The following are possible values for the constant representing the volume of air in the oral cavity, CO, depending on place of articulation:

labial	alveolar	velar
7.16E-5	5.82E-5	4.48E-5

The following are possible values for the constants describing the stiffness of the vocal tract walls, depending on place of articulation. They are ordered with respect to increasing stiffness.

WALLS LIKE LAX CHEEKS

labial	alveolar	velar
RW=6.4	RW=7.27	RW=8.0
COILW=1.68E-02	COILW=1.909E-02	COILW=2.1E-02
CW=1.4973E-03	CW=1.3018E-03	CW=1.1834E-03

DEFAULT CASE -- WALLS LIKE TENSE CHEEKS

labial	alveolar	velar
RW=8.48	RW=9.64	RW=10.6
COILW=1.2E-02	COILW=1.364E-02	COILW=1.5E-02
CW=5.6256E-04	CW=4.9505E-04	CW=4.5E-04

WALLS LIKE NECK WALL

labial	alveolar	velar
RW=18.56	RW=21.09	RW=23.2
COILW=1.92E-02	COILW=2.182E-02	COILW=2.4E-02
CW=2.5458E-04	CW=2.2403E-04	CW=2.0367E-04

There are four variables whose values are functions, not single numbers. These variables, GHEI, OHEI, ES, and IO, do not have default values, and they are all given values at the same time and in the same way. In principle, any variable in VTM can be time-varying, and for a specific application a user may want to do the minor amount of programming required to make some variable a time function, but in general the arrangement described here gives a reasonable balance between accuracy and convenience.

GHEI is the function for changing glottal opening. We represent the mean value over a vibratory cycle as .022 cm. A reasonable value for a spread glottis is .18. For an opening or closing gesture, the user can have VTM interpolate between these. OHEI is the function for changing oral opening. A typical value for an oral opening is .2 or .3; closure for a stop is 0.0. Transition durations for an oral opening or closing gesture must be specified with the OHEI function. 20 msec may be used for a quick trial run, but something like 50-55-70 for labials-alveolars-velars is more natural.

ES represents changes in subglottal pressure through muscular force. When it is zero, subglottal pressure is determined solely by elastic recoil. ES should rise from about -10 cm aq to 0 for initial stops and fall back from 0 to -9 cm aq or so for final stops. A rise or fall like this takes 200 msec in English; we model it linearly. Changes in ES to represent stress look like the desired change in subglottal pressure: e.g. for a rise of about 4 cm aq, ES rises about 4 cm aq with the peak in ES about 20 msec earlier than the desired peak in P_s.

IO represents active expansion or contraction of the volume of the vocal tract, such as jaw lowering, pharynx expansion, or larynx lowering. Its unit corresponds to volume changes (in cc's). 40 cc/sec is a large change typical as a peak value for a [b].

When the RUN command is given, VTM performs all calculations, puts an output file into the LP queue, and asks for new inputs.

Envisioned for the future is a version of VTM that works in terms of physiological, rather than circuit, variables, so that the user of the program need not know anything about circuit elements or about the model to use the program. It may even be desirable to group together variables into such higher-level phonetic inputs as "place of articulation", using default values as sketched above, so that undergraduate students can do simple exercises.

4. An example: voicing in utterance-medial stops

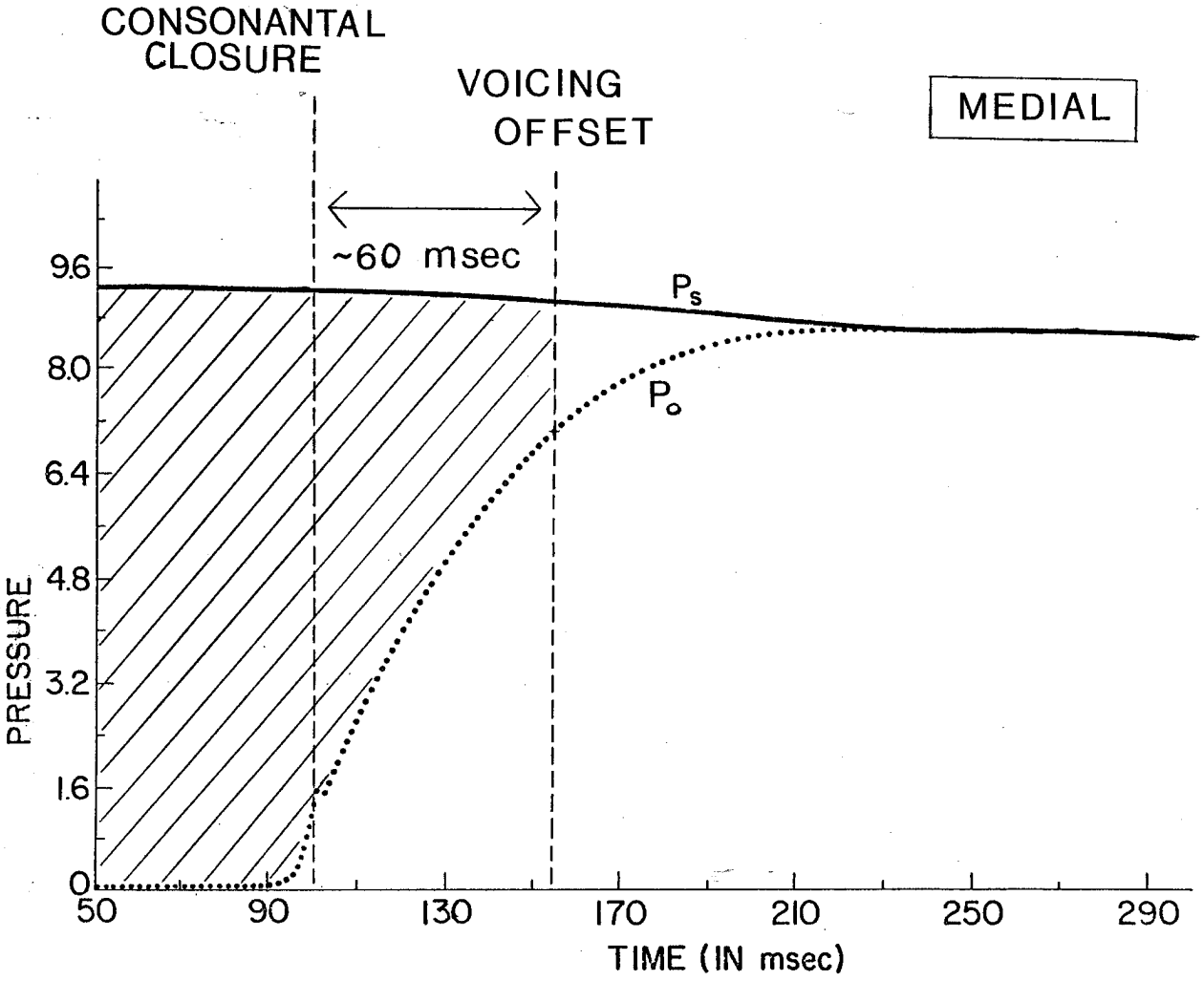
Using the model, it is possible to calculate such things as how P_o will increase in time following the moment of occlusion, as long as sufficient detail about the articulatory states corresponding to its elements is specified. Consider, for example, how P_o will change during a labial stop which occurs utterance medially, between identical vowels. Suppose that all but one of the time-variable elements in the model subject to voluntary control are fixed. Specifically, pressure below the glottis (initially perhaps as much as 10 cm H_2O above atmosphere) derives entirely from elastic recoil of the stretched tissues surrounding the lungs (therefore E_s=0); there are no muscularly induced changes in supraglottal volume (therefore I_o=0), or in the mechanical properties of tissues surrounding the lungs and mouth (therefore RLC_t and RLC_w are constants); and the vocal folds are appropriately adducted and tensed for voicing (glottal area is constant). Only cross-sectional area of the mouth opening (A_a) is allowed to vary, as it must, first to produce a constriction at the lips and then to release it.

Under conditions such as these, pressures above and below the glottis (P_o and P_s, respectively) can be expected to change with time as shown in Figure 3. Note from this figure that the difference between P_s and P_o, though decreasing, is clearly greater than 2 cm aq, the amount thought to be required for voicing maintainence, for the first sixty-odd msec of the 80 msec closure interval. Thus, voicing can be expected during that portion of the intervocalic stop, with offset occurring only late in the closure, within 20 msec of release. The general result, then, is that a labial stop articulation under the aforementioned conditions will naturally be largely voiced.

The relatively lengthy interval of closure voicing for the simulation shown in Figure 3 is due almost entirely to the yielding walls which surround the supraglottal cavity. In effect, their outward motion during the stop closure --

in response to the increasing air pressure they contain -- slows down the decrease in the transglottal pressure drop, and thereby lengthens the interval during closure when voicing is possible. If the walls of the vocal tract were rigid, effective pressure neutralization (and voice offset) would occur within 10 msec of occlusion, as Rothenberg (1968) showed. The value used here for wall stiffness is a more realistic one, and allows voicing to continue for a longer time.

Figure 3.



Of course, a voiceless output could be guaranteed by a glottal spreading gesture during the closure interval. On the other hand, there are several articulatory adjustments which may prolong the voicing interval well beyond the 60 msec or so suggested by Figure 3, producing a fully voiced stop. These include increasing P_s by activating the expiratory muscles; decreasing average area of the glottis and/or tension of the vocal folds; and decreasing P_o by decreasing the level of activity in muscles which underlie the walls of the supraglottal cavity; actively enlarging the volume of that cavity by adjusting positions of the larynx, tongue, and soft palate; or creating a narrow opening between the posterior pharyngeal wall and soft palate (nasal leak). These maneuvers,

occurring singly or in combination, will have their greatest effect on the duration of closure voicing when they occur during the closure interval itself, in concert with the rise in P_o which accompanies vocal tract occlusion. Implementing maneuvers such as these in the model involves specifying how each of the relevant control parameters will vary in time.

Acknowledgements

The implementation of the aerodynamic model described in this report was carried out in the Speech Communication Group at MIT, largely by John Westbury. The results described in Section 4 above are part of research that was a joint effort between me and John, and will be presented more fully in a forthcoming paper. Some of the prose in this report is taken directly from that paper. With support from a grant from NINCDS to Peter Ladefoged, the very primitive computer program used at MIT has been replaced at UCLA with one written by Chris Pettus and Flora Wu, students in our Linguistics & Computer Science major, and me, and documentation has been provided. Thanks are due to the UCLA graduate students who have tried out the program and its documentation.

References

Baer, T. (1975). "Investigation of Phonation Using Excised Larynxes", unpublished doctoral dissertation, MIT.

Ishizaka, K. and M. Matsudaira (1972). "Fluid Mechanical Considerations of Vocal Cord Vibration", Speech Commun. Res. Lab. Mono. No. 8.

Ladefoged, P. (1964). "Comment on 'Evaluation of Methods of Estimating Subglottal Air Pressure'", J. Speech Hear. Res. 7, 291-292.

Muller, E. M., and W. S. Brown, Jr. (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 Press, New York), pp. 317-389.

Ohala, J. J. (1976). "A model of speech aerodynamics", Report of the Phonology Laboratory (Berkeley) 1, 93-107.

Rothenberg, M. (1968). The Breath-Stream Dynamics of Simple-Released-Plosive Production. Bibl. Phonetica 6.

van den Berg, J. (1958). "Myoelastic-Aerodynamic Theory of Voice Production", J. Speech and Hearing Research, Vol.1, No. 3, 227-244.

Westbury, J. R. (1983). "Enlargement of the supraglottal cavity and its relation to stop consonant voicing", J. Acoust. Soc. Am. 73, 1322-36.