COMMUNICATION SCIENCES

AND

ENGINEERING

VII. SPEECH COMMUNICATION

Academic and Research Staff

Prof. K. N. Stevens
Prof. M. Halle
Prof. W. L. Henke
Prof. A. V. Oppenheim
Dr. S. Berkowitz
Dr. Margaret Bullowa
Dr. J. L. Fidelholtz
Dr. A. W. F. Huggins
Dr. A. R. Kessler
Dr. Emily F. Kirstein
Dr. D. H. Klatt
Dr. Paula Menyuk
Dr. J. S. Perkell
Dr. R. A. Stefanski
Dr. Jacqueline Vaissiere
Mary M. Klatt
R. M. Mersereau
Estill Putney

Graduate Students

G. A. Akers
T. Baer
D. S. K. Chan
R. E. Crochiere
D. E. Dudgeon
Ursula Goldstein
S. Maeda
B. Mezrich
H. Pines
M. R. Portnoff
L. Tung
V. W. Zue

A. COMPUTATIONAL METHODS FOR SENSITIVITY ANALYSIS OF DIGITAL FILTERS

National Science Foundation (Grant GK-31353)

R. E. Crochiere

1. Introduction

In this report, we define multiparameter coefficient sensitivity for digital filters, and present an efficient method of computing these sensitivities for arbitrary digital filter structures. We also establish a connection between the group delay and particular coefficient sensitivities in an arbitrary digital filter structure by using a property of homogeneous functions. A practical method for computing the group delay from the sensitivities is then given.

2. First-Order Sensitivity Definitions

The multiparameter sensitivity of a transfer function or a system function, T, of a digital filter with respect to its coefficients c_i $(i = 1, \dots n)$ can be defined in several ways. We consider three basic definitions that have been used classically in filter theory[1-4] and give general relationships among them.

The first definition is referred to as the "basic sensitivity" or simply the "sensitivity" of the filter. It is the vector of partial derivatives (gradient) of the transfer function T with respect to its coefficients,

$$\underline{S}[T] \triangleq \nabla T = \left(\frac{\partial T}{\partial c_1}, \dots \frac{\partial T}{\partial c_n}\right)^t, \qquad (1)$$

where

$$T \triangleq T(c_1, c_2, \dots c_n).$$

A single element of this vector can be expressed as

$$S_i[T] = \lim_{\Delta c_i \to 0} \frac{\Delta T}{\Delta c_i} = \frac{\partial T}{\partial c_i}. \tag{2}$$

The first-order (incremental) variation of T, ΔT, with respect to incremental changes in the coefficients, Δc_i, can be determined by the sensitivities as

$$\Delta T \approx \sum_{i=1}^{n} S_i[T]\, \Delta c_i = \underline{S}^t[T]\, \Delta \underline{c}, \tag{3}$$

where

$\underline{S}^t[T]$ = the transpose of the sensitivity vector

$\underline{\Delta c}$ = the vector of incremental changes in the coefficients

$= (\Delta c_1, \Delta c_2, \ldots \Delta c_n)^t.$

A second definition that has been widely used in analog filter theory is referred to as the "relative sensitivity." A single element of the relative sensitivity vector is defined as

$$S_i^r[T] \triangleq \lim_{\Delta c_i \to 0} \frac{\Delta T/T}{\Delta c_i/c_i} = \frac{c_i}{T} \frac{\partial T}{\partial c_i}$$

$$\triangleq \frac{\partial \ln T}{\partial \ln c_i}, \tag{4}$$

and the sensitivity vector is given as

$$\underline{S}^r[T] \triangleq \left(\frac{\partial \ln T}{\partial \ln c_1}, \frac{\partial \ln T}{\partial \ln c_2}, \ldots \frac{\partial \ln T}{\partial \ln c_n}\right)^t. \tag{5}$$

This definition is popular because it is closely associated with percentage changes of the system function with respect to percentage changes in the coefficients (element values). It has been particularly useful in tolerance studies. In the case of digital filters this definition would seem to be most appropriate for filters with floating-point coefficients.

We see from (4) that the relative sensitivities can be expressed in terms of the sensitivities as follows.

$$S_i^r[T] = \frac{c_i}{T} S_i[T]. \tag{6}$$

The relative (incremental) change of the system function with respect to relative (incremental) changes in the coefficients can be expressed as

$$\Delta T^{r} \triangleq \frac{\Delta T}{T} \approx \underline{S}^{r^{t}}[T]\,\underline{\Delta c}^{r}, \tag{7}$$

where

$$\Delta c_i^r = \frac{\Delta c_i}{c_i} \tag{8}$$

= the relative change in the coefficients.

The third definition of sensitivity is often referred to as the "semirelevant sensitivity," which is expressed as

$$S_i^s[T] \triangleq \lim_{\Delta c_i \to 0} \frac{\Delta T/T}{\Delta c_i} = \frac{1}{T}\frac{\partial T}{\partial c_i} \triangleq \frac{\partial \ln T}{\partial c_i} \tag{9}$$

or

$$S_i^s[T] = \frac{1}{T} S_i[T] \tag{10}$$

and

$$\underline{S}^s = \nabla(\ln T) = \left(\frac{\partial \ln T}{\partial c_1}, \frac{\partial \ln T}{\partial c_2}, \ldots \frac{\partial \ln T}{\partial c_n}\right)^t. \tag{11}$$

The third definition seems appropriate for digital filters with fixed-point arithmetic, since it relates relative changes in the system function to absolute changes in the coefficients. This relation is expressed as

$$\Delta T^{r} \triangleq \frac{\Delta T}{T} \approx \underline{S}^{s^{t}}[T]\,\underline{\Delta c}. \tag{12}$$

We often wish to determine the sensitivity of the magnitude of the system function $|T|$. These sensitivities can be derived from the basic sensitivities, under the assumption that the coefficients c_i are real.

$$S_i[|T|] = \frac{1}{|T|}\,\mathrm{Re}\left(T^* S_i[T]\right) = \mathrm{Re}\left(\frac{|T|}{T}\,S_i[T]\right), \tag{13}$$

$$S_i^r[|T|] = \mathrm{Re}\left(\frac{c_i}{T} S_i[T]\right), \tag{14}$$

$$S_i^s[|T|] = \mathrm{Re}\left(\frac{1}{T} S_i[T]\right). \tag{15}$$

3. Efficient Computation of Network Sensitivities

From Eqs. 6, 10, 13, 14, and 15 it can be seen that all of the first-order sensitivity functions of interest can be computed once the basic sensitivities defined in (1) and (2) are known. We shall now describe a method by which all of the basic sensitivities of a digital network at a given frequency can be determined exactly with only two complete analyses of the network at that frequency.

In a previous report[4] a general matrix formulation for digital networks based on a signal-flow graph representation was presented. The equations describing a digital network can be written in the form

$$\underline{Y}(z) = \underline{H}_c^t \underline{Y}(z) + \underline{H}_d^t \underline{Y}(z)\, z^{-1} + \underline{X}_s(z), \tag{16}$$

where

$\underline{Y}(z)$ = the column vector of the node signal values

$\underline{X}_s(z)$ = the column vector of the source branch values

$\underline{H}_c$ = the matrix of coefficients for branches with simple coefficients

$\underline{H}_d$ = the matrix of coefficients for branches with coefficients and delays.

The matrix of transfer functions in the network can then be expressed as

$$\underline{T} = \left(\underline{I} - \underline{H}_c - \underline{H}_d z^{-1}\right)^{-1}, \tag{17}$$

where $T_{k\ell}$ is defined as the transfer function from node k to node ℓ.

In our previous report it was also shown that with the aid of a relation of Fettweis[5] the sensitivity of a particular transfer function $T_{k\ell}$ with respect to a branch coefficient of a branch directed from some node p to some node q can be expressed as the product of two transfer functions in the network. For a branch with a simple coefficient, c, this expression is

$$S_c[T_{k\ell}] = \frac{\partial T_{k\ell}}{\partial c} = T_{kp} T_{q\ell} \tag{18}$$

and for a branch with a coefficient, c_d, and a delay it is

$$S_{c_d}[T_{k\ell}] = \frac{\partial T_{k\ell}}{\partial c_d} = z^{-1}T_{kp}T_{q\ell}. \tag{19}$$

Therefore, in general, if we know all of the transfer functions from the input node k to the nodes p (p = 1, ... N, where N is the total number of nodes in the network), and all of the transfer functions from the nodes q (q = 1, ... N) to the output node ℓ, then we have sufficient information to compute all of the first-order sensitivities $\underline{S}[T_{k\ell}]$ for the network function $T_{k\ell}$.

The first set of transfer functions T_{kp} can be obtained by solving the following set of simultaneous linear equations

$$\left(\underline{I} - \underline{H}_c^t - \underline{H}_d^t z^{-1}\right)\underline{Y}(z) = \underline{X}_s(z). \tag{20}$$

The elements of the vector of input signals $\underline{X}_s(z)$ are all set to zero except for the one that corresponds to node k. This element is set to unity (the z transform of a unit sample). Using complex arithmetic and assuming that $\underline{H}_d$, $\underline{H}_c$, and z are known, we can solve Eq. 20 for $\underline{Y}(z)$ by a procedure such as Gaussian elimination.[6] The solution $\underline{Y}(z)$ represents the appropriate transfer functions from the input node k to all of the nodes in the network. That is,

$$T_{kp}(z) = Y_p(z). \tag{21}$$

The second set of transfer functions $T_{q\ell}(z)$ can be obtained in a similar manner by analyzing the transpose network.[5,7] The transpose network is obtained by reversing the direction of all branches in the original network and interchanging inputs and outputs. In matrix formulation (16) this corresponds to taking the transpose of the matrices $\underline{H}_c$ and $\underline{H}_d$. It can be shown that the transpose network is interreciprocal with the original network.[5,7] That is,

$$T_{q\ell}(z)\Big|_{\substack{\text{original}\\ \text{network}}} = T_{\ell q}(z)\Big|_{\substack{\text{transpose}\\ \text{network}}} \tag{22}$$

for all q (q = 1, ... N) and ℓ (ℓ = 1, ... N). Therefore, by solving the transpose network with an input at node ℓ, we can obtain all of the transfer functions $T_{q\ell}(z)$ of the original network (q = 1, ... N). This corresponds to solving the set of simultaneous linear equations

$$\left(\underline{I} - \underline{H}_c - \underline{H}_d z^{-1}\right)\underline{Y}(z) = \underline{X}_s(z) \tag{23}$$

for $\underline{Y}(z)$ with all of the elements of $\underline{X}_s(z)$ equal to zero except for the one corresponding to node ℓ, which is set to unity. The desired transfer functions $T_{q\ell}(z)$ can then be obtained with the aid of (22) from the solution of (23) by the relation

$$T_{q\ell}(z) = Y_q(z). \tag{24}$$

Thus, from the solution of two sets of simultaneous linear equations (20) and (23) and with the aid of (18), (19), (21), and (24), all of the basic sensitivities of the system function $T_{k\ell}(z)$ can be determined. This technique of computing network sensitivities is similar in some respects to the adjoint network techniques used by Director and Rohrer[8] in the computation of element sensitivities in analog circuits.

In general, the sets of linear equations (20) and (23) must be solved by using complex arithmetic. For computer solution such programs are often available only for solving sets of linear equations for real numbers. This is the case, for example, in the APL language. In such cases Eqs. 20 and 23 can be reformulated so that the real and imaginary components of $\underline{Y}(z)$ can be solved separately, each requiring the solution of a set of simultaneous linear equations in real numbers. To do this, consider for the sake of convenience the case for $z = e^{j\omega}$. Then (20) becomes

$$\left(\underline{I} - \underline{H}_c^t - \underline{H}_d^t(\cos\omega - j\sin\omega)\right) \underline{Y}(e^{j\omega}) = \underline{X}_s(e^{j\omega}). \tag{25}$$

Premultiplying (25) by $\underline{I} - \underline{H}_d^t\left(\underline{I} - \underline{H}_c^t\right)^{-1}(\cos\omega + j\sin\omega)$ gives

$$\underline{A}\left(\underline{Y}_R(e^{j\omega}) + j\underline{Y}_I(e^{j\omega})\right) = \left(\underline{I} - \underline{H}_d^t\left(\underline{I} - \underline{H}_c^t\right)^{-1}(\cos\omega + j\sin\omega)\right) \underline{X}_s(e^{j\omega}), \tag{26}$$

where

$\underline{Y}_R(e^{j\omega})$ = the real part of $\underline{Y}(e^{j\omega})$,

$\underline{Y}_I(e^{j\omega})$ = the imaginary part of $\underline{Y}(e^{j\omega})$,

$$\underline{A} = \underline{I} - \underline{H}_c^t + \underline{H}_d^t\left(\underline{I} - \underline{H}_c^t\right)^{-1} \underline{H}_d^t - 2\underline{H}_d^t \cos\omega. \tag{27}$$

For a unit sample response $\underline{X}_s(e^{j\omega}) = \underline{X}_R$ is real and the real and imaginary parts of (26) can be separated as

$$\underline{A}\,\underline{Y}_R(e^{j\omega}) = \underline{X}_R - \underline{H}_d^t\left(\underline{I} - \underline{H}_c^t\right)^{-1} \underline{X}_R \cos\omega \tag{28}$$

$$\underline{A}\,\underline{Y}_I(e^{j\omega}) = -\underline{H}_d^t\left(\underline{I} - \underline{H}_c^t\right)^{-1} \underline{X}_R \sin\omega. \tag{29}$$

Equations 28 and 29 can be solved for $\underline{Y}_R(e^{j\omega})$ and $\underline{Y}_I(e^{j\omega})$ as linear sets of simultaneous equations in real variables. The matrix inversions in (27)-(29) also require only real variables. It may seem, at first, that this approach requires much more computation, since it is necessary to perform a matrix inversion and several extra matrix multiplications. On the other hand, when we are interested in computing the sensitivities for a large set of frequencies (as is usually the case) the matrix inversions and matrix multiplications all have to be computed only once for the entire set of frequencies. Thus solutions to (28) and (29) can be computed quite efficiently. Similarly, we can obtain the solution to (23) by solving two sets of real simultaneous linear equations which are related to (28) and (29) by the transposes of $\underline{A}$, $\underline{H}_c$, and $\underline{H}_d$.

$$\underline{A}^t\underline{Y}_R(e^{j\omega}) = \underline{X}_R - \underline{H}_d(\underline{I}-\underline{H}_c)^{-1}\ \underline{X}_R \cos\omega \tag{30}$$

$$\underline{A}^t\underline{Y}_I(e^{j\omega}) = -\underline{H}_d(\underline{I}-\underline{H}_c)^{-1}\ \underline{X}_R \sin\omega. \tag{31}$$

An alternative method for determining the sensitivities is to perform the complex matrix inversion in (17) and pick out the necessary transfer functions for (18) and (19) from the inverted matrix $\underline{T}$. This approach is less desirable, however, since it generally involves more computation than the method of solving linear simultaneous equations and it is also prone to greater computational error.[6]

For nonrecursive structures the unit sample responses of the network are all of finite duration. In this case we can solve the simultaneous equations (20) and (23) in the time domain by placing a unit sample in the appropriate inputs of the original network and of the transpose network and iterating the networks to obtain the appropriate unit sample responses. Then, $T_{kp}(z)$ and $T_{q\ell}(z)$ can be obtained by taking the z transforms of these unit sample responses. For special classes of frequency points these transforms can be performed very efficiently with the aid of the FFT algorithm, frequency-warping techniques,[9] and the chirp z transform algorithm.

4. Homogeneity, Sensitivity, and Group Delay in Digital Filters

Interesting properties of first-order sensitivities of digital filters can be determined with the aid of a sensitivity relation associated with homogeneous functions. A function $F(c_1, c_2, \ldots c_n)$ is defined as homogeneous with respect to its parameters c_i $(i = 1, 2, \ldots n)$ if

$$F(\lambda c_1, \lambda c_2, \ldots \lambda c_n) = \lambda^M F(c_1, c_2, \ldots c_n), \tag{32}$$

where λ is an arbitrary number and M is defined as the order of homogeneity. The derivative of this function with respect to λ can be determined by means of the chain rule

$$\frac{\partial F(\lambda c_1, \lambda c_2, \ldots \lambda c_n)}{\partial \lambda} = \sum_{i=1}^{n} \frac{\partial F(\lambda c_1, \lambda c_2, \ldots \lambda c_n)}{\partial \lambda c_i} \frac{\partial \lambda c_i}{\partial \lambda}$$

$$= M\lambda^{M-1} F(c_1, c_2, \ldots c_n). \tag{33}$$

By recognizing that the derivative with respect to λ of λc_i is c_i and letting $\lambda = 1$, we get Euler's relation for homogeneous functions,

$$\sum_{i=1}^{n} c_i \frac{\partial F}{\partial c_i} = MF, \tag{34}$$

where

$$F = F(c_1, c_2, \ldots c_n).$$

We divide by F, and see that the sum of the relative sensitivities of a homogeneous function must equal the constant M, the order of homogeneity.

$$\sum_{i=1}^{n} S_i^r[F] = \sum_{i=1}^{n} \frac{c_i}{F} \frac{\partial F}{\partial c_i} = M. \tag{35}$$

For analog circuits, it can be shown that the impedances and transfer functions are homogeneous functions with respect to their element values.[1, 10] This property has led to interesting sensitivity relations for analog networks.[1] In digital networks, the transfer functions, in general, are not homogeneous with respect to their coefficients. Some special classes of digital structures, however, do exhibit such homogeneity. In particular, two classes of structures are the direct and the direct-canonic forms. For these two classes of structures the transfer function can be expressed as

$$T = \frac{\sum\limits_{i=0}^{n} a_i z^{-i}}{\sum\limits_{i=0}^{m} b_i z^{-i}}, \qquad b_0 = 1. \tag{36}$$

From (36) three forms of homogeneity can be observed. T is homogeneous of degree 1 with respect to the coefficients a_i. That is,

$$T(\lambda a_0, \lambda a_1, \dots \lambda a_n) = \lambda^1 T(a_0, a_1, \dots a_n). \tag{37}$$

T is homogeneous of degree -1 with respect to the coefficients b_i (the unity scale factor b_0 must be included as a coefficient)

$$T(\lambda b_0, \lambda b_1, \dots \lambda b_m) = \lambda^{-1} T(b_0, b_1, \dots b_n). \tag{38}$$

T is homogeneous of degree zero with respect to all of the coefficients a_i and b_i

$$T(\lambda a_0, \dots \lambda a_n, \lambda b_0, \dots \lambda b_n) = \lambda^0 T(a_0, \dots a_n, b_0, \dots b_m). \tag{39}$$

From these homogeneous properties with the aid of (35) we can state that, for direct form and direct-canonic form structures, the sum of the relative sensitivities of the coefficients a_i is equal to 1, the sum of the relative sensitivities of the coefficients b_i is equal to -1, and the sum of the relative sensitivities with respect to all of the coefficients (a_0 and b_0 must be included) is equal to zero.

Another class of structures that exhibit homogeneity with respect to the coefficients is derived by continued fraction expansions.[11] These structures will not be discussed here.

A more general application of homogeneity for digital structures can be observed when the complex frequency z is included as a parameter. In this case it can be seen from (17) that the transfer functions of a digital network are homogeneous of degree zero with respect to the coefficients of the delay branches ($\underline{H}_d$) and the complex frequency z

$$T(\lambda c_{d1}, \lambda c_{d2}, \dots \lambda c_{dm}, \lambda z) = \lambda^0 T(c_{d1}, c_{d2}, \dots c_{dm}, z). \tag{40}$$

Then by (34) we can write

$$z \frac{\partial T}{\partial z} + \sum_{i=1}^{m} c_{di} \frac{\partial T}{\partial c_{di}} = 0, \tag{41}$$

where the sum over i is the sum over all branch coefficients for branches with unit delays. This expression relates the frequency sensitivity of the system function to specific coefficient sensitivities of the system function. It can be used as a practical method for computing the group delay of the filter. To demonstrate this method, let

$$T = |T|\ e^{j\theta}, \tag{42}$$

where θ is the phase. We then can write

$$\frac{\partial T}{\partial \omega} = \frac{\partial T}{\partial e^{j\omega}} \frac{\partial e^{j\omega}}{\partial \omega} = je^{j\omega} \frac{\partial T}{\partial e^{j\omega}} = jz \left.\frac{\partial T}{\partial z}\right|_{z=e^{j\omega}} \tag{43}$$

and also

$$\frac{1}{T}\frac{\partial T}{\partial \omega} = \frac{\partial \ln T}{\partial \omega} = \frac{\partial \ln |T|}{\partial \omega} + j\frac{\partial \theta}{\partial \omega}. \tag{44}$$

By combining (41), (43), and (44) and using the definition of relative sensitivity (6), we get

$$\frac{\partial \ln |T|}{\partial \omega} + j\frac{d\theta}{\partial \omega} = \frac{-j}{T}\sum_{i=1}^{m} c_{di}\frac{\partial T}{\partial c_{di}} = -j\sum_{i=1}^{m} S_i^r. \tag{45}$$

The first term on the right side of (45) is pure real and the second term, which represents -j times the group delay, is pure imaginary. Therefore, by separation of the real and imaginary parts of (45) the group delay can be expressed in terms of the relative sensitivities as

$$\tau \triangleq \text{group delay of } T \triangleq -\frac{\partial \theta}{\partial \omega}$$

$$= \text{Re}\sum_{i=1}^{m} S_i^r = \sum_{i=1}^{m} \text{Re } S_i^r. \tag{46}$$

Thus the group delay is simply the sum of the real parts of the relative sensitivities of the system function with respect to all of the coefficients in the branches with coefficients and delays. From (45) a convenient expression for the frequency derivative of the log magnitude of the system function is also obtained in terms of these sensitivities.

$$\frac{\partial \ln |T|}{\partial \omega} = \sum_{i=1}^{m} \text{Im } S_i^r. \tag{47}$$

By using the methods of computing the sensitivities discussed in this report, a practical method for computing exactly the group delay and the derivative with respect to frequency of the log magnitude response of the system function may be represented.

References

1. K. Géher, Theory of Network Tolerances (Akadémiai Kiadó, Budapest, 1971).
2. A. J. Goldstein and F. F. Kuo, "Multiparameter Sensitivity," IEEE Trans., Vol. CT-8, No. 2, pp. 177-178, June 1961.
3. S. R. Parker, "Sensitivity: Old Questions, Some New Answers," IEEE Trans., Vol. CT-18, No. 1, pp. 27-34, January 1971.
4. R. E. Crochiere, "Some Network Properties of Digital Filters," Quarterly Progress Report No. 107, Research Laboratory of Electronics, M.I.T., October 15, 1972, pp. 103-112.
5. A. Fettweis, "A General Theorem for Signal-Flow Networks, With Applications," Arch. Elekt. Übertrag. 25, 557-561 (1971); see also Digest of Technical Papers, London, 1971, IEEE International Symposium on Electrical Network Theory, pp. 3-4.
6. G. E. Forsythe and C. B. Moler, Computer Solution of Linear Algebraic Systems (Prentice-Hall, Inc., Englewood Cliffs, N. J., 1967).
7. L. B. Jackson, "On the Interaction of Roundoff Noise and Dynamic Range in Digital Filters," Bell System Tech. J. 49, 159-184 (1970).
8. S. W. Director and R. A. Rohrer, "The Generalized Adjoint Network and Network Sensitivities," IEEE Trans., Vol. CT-16, No. 3, pp. 318-323, August 1969; see also "Automated Network Design – The Frequency Domain Case," Vol. CT-16, No. 3, pp. 330-337, August 1969.
9. A. V. Oppenheim and D. H. Johnson, "Discrete Representation of Signals," Proc. IEEE 60, 681-691 (1972).
10. C. Belove, "Sensitivity Sums of Homogeneous Functions," IEEE Trans., Vol. CT-11, No. 1, p. 171, March 1964.
11. S. K. Mitra and R. J. Sherwood, "Canonic Realizations of Digital Filters Using the Continued Fraction Expansion," IEEE Trans., Vol. AU-20, No. 3, pp. 185-194, August 1972.

B. VOICE-ONSET TIME, FRICATION AND ASPIRATION IN WORD-INITIAL CONSONANT CLUSTERS

National Institutes of Health (Grant 5 RO1 NS04332-10)
U. S. Navy – Office of Naval Research (Contract N00014-67-A-0204-0069)

D. H. Klatt

This report presents further results of a spectrographic study of prestressed word-initial consonant clusters in English.[1] The timing of voice onset relative to plosive release (VOT) has been studied previously in word-initial position in prestressed, pre-unstressed and utterance-final environments,[2, 3] but the detailed acoustic characteristics of plosive consonants have not been systematically investigated in prestressed consonant clusters.

The acoustic characteristics of English consonant clusters such as VOT are of interest in practical applications such as the generation of synthetic speech by rule and the automatic recognition of speech. The acoustic patterns also reveal articulatory recoding strategies in consonant clusters that are of interest to investigators concerned with the control of articulatory models of speech production. The present data are particularly relevant to studies of the coordination of laryngeal and supralaryngeal gestures.

1. Experiment

A list of monosyllabic words was constructed to include 5 examples for each of 25 different word-initial clusters. Words beginning with single plosive consonants were also recorded. Four monosyllabic words involving the vowels /i ɛ a^y u/ were selected for each consonant and cluster, and a fifth word was generated by adding a second syllable to the end of one of the 4 monosyllabic words. For example, the [str] words were "street, stress, strike, strewn, and stressful."

The word list was randomized and recorded at a moderate speaking rate in an anechoic chamber by 3 adult male speakers. All words were spoken in the frame sentence "Say x instead" in order to produce speaking rates more nearly in line with conversational speech, and to avoid effects of prepausal lengthening in utterance-final position.

Broadband spectrograms were made of all of the phrases. Vertical lines were drawn on the spectrograms at the time of release of any plosive consonant and at voice onset time following plosive release. The voice onset time (VOT) is defined as the onset time of normal voicing relative to plosive release.

If a voiced stop is preceded by a phonetic segment that is also voiced (as is the case in our frame sentence), voicing normally continues during closure. In Fig. VII-1a voicing energy can be observed at low frequencies during the early part of the closure interval.

Indications of weak voicing may extend throughout closure and continue during the frication burst, or the vocal cords may discontinue vibration as the mouth pressure increases.

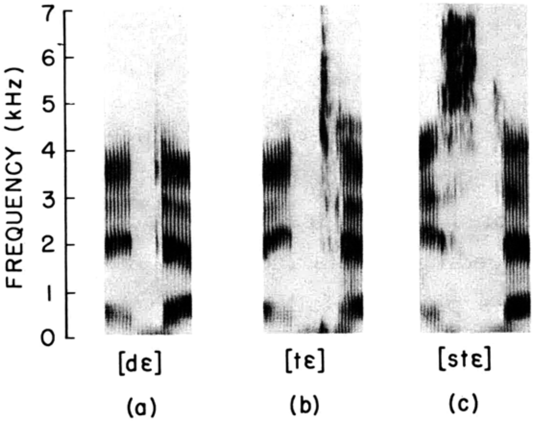

Fig. VII-1.

Broadband spectrograms of the word-initial plosives, excised from utterances containing (a) a voiced plosive, (b) a voiceless aspirated plosive, and (c) a voiceless unaspirated plosive.

Since the presence or absence of prevoicing is not phonemic in English[2] and is frequently difficult to establish spectrographically because of the limited dynamic range of a spectrogram, prevoicing is ignored. Voicing onset is defined to follow the frication burst and coincide with the onset of normal voicing where a number of formants are visibly excited by voicing striations.

When possible, the VOT interval in voiceless plosives has been divided into two phases: (i) the initial phase, during which the burst of frication noise is generated at the expanding constriction, and (ii) the terminal phase during which the sound output is primarily aspiration noise generated at the glottis. The expected spectral characteristics of frication burst spectra for different places of articulation[4] were used to distinguish the burst phase from the aspiration phase.

All transitions from voiceless to voiced segments were studied in the cluster data and we found that a substantial interval of aspiration noise could be seen following voiceless fricative consonants. The aspiration intervals were delineated by placing vertical lines on the spectrograms at the termination of visible frication noise in the fricative and at the onset of normal vocal-cord vibrations.

2. Results

The data are in the form of average durations obtained from 15 samples of each consonant cluster (5 words from each of 3 speakers). In some cases the identity of the following vowel had an effect and these data are discussed separately. The presence of a second syllable in some of the words of the corpus had no measurable effect on the parameters studied in this experiment. Interspeaker and intraspeaker variability were of the same order of magnitude as reported by Lisker and Abramson.[2,3] Our data suggest that individual VOT measurements for /p t k/ are normally distributed with a standard deviation of approximately 11 ms which was computed from individual means for each speaker and each phonetic environment. The standard deviation for /b d g/ (excluding prevoiced examples) was approximately 5 ms.

a. Voice Onset Time

Measured voice onset times in clusters involving plosive consonants are presented in Table VII-1. The VOT for voiced plosives averaged ~14 ms before a vowel and 20 ms before a sonorant consonant. This interval is occupied by the burst of frication noise.

The VOT for voiceless plosives was ~60 ms preceding vowels and 75 ms preceding sonorant consonants. The VOT values for voiceless plosives followed by vowels are slightly larger than average VOT obtained by Lisker and Abramson from sentences that were read by several speakers,[3] but our present study yields average VOT values that

Table VII-1. Voice onset time (in ms) for selected consonant clusters. Each entry indicates the average VOT as obtained from 15 tokens.

b	7	br	12	bl	9		
d	14	dr	23				
g	23	gr	32	gl	23		
p	47	pr	60	pl	61		
t	64	tr	93			tw	104
k	70	kr	85	kl	77	kw	93
sp	11	spr	18	spl	16		
st	23	str	37				
sk	30	skr	35			skw	39

Table VII-2. Voice onset time (in ms) for single-syllable words involving voiceless plosives not preceded by /s/ as a function of the following vowel. Each entry is the average of 30 tokens.

Vowel	VOT
/i/	78
/ɛ/	71
$/a^y/$	73
/u/	90

are smaller than have been obtained from words spoken in isolation.[2]

When a voiceless plosive is preceded by the consonant /s/ in the same word, Table VII-1 indicates that the VOT is considerably reduced. The VOT is ~21 ms before vowels and ~ 30 ms before sonorant consonants. These values are within a few milliseconds of the data for the analogous voiced consonants. Nevertheless, a brief interval of aspiration was frequently seen in these nominally voiceless consonants, whereas no such interval was found in the voiced-stop data. Spectrograms of portions of the words "desk," "test," and "step" are shown in Fig. VII-1 to illustrate some of these differences.

The data in Table VII-1 indicate that the VOT is always approximately 10 ms less than average for labial plosives and approximately 10 ms greater than average for velar plosives. Lisker and Abramson have found a similar relationship across several languages.[2]

The VOT in word-initial voiceless plosives followed by a vowel or a sonorant consonant was found to be greater if the syllable nucleus is a high vowel. Table VII-2 indicates that the VOT is 20% longer preceding /i u/ than preceding /ε a^y/. This difference is significant at the 0.01 level. Lisker and Abramson, however, report observing no vowel-conditioned effects in a similar study.[3]

b. Burst Durations

The duration of the burst of frication noise is presented in Table VII-3 for all clusters involving plosive consonants. By definition the burst duration is equal to the VOT in voiced-stop environments. Data for /p/ are not included because it was difficult to distinguish the burst phase from the following interval of aspiration. Burst duration

Table VII-3. Duration of the frication burst (in ms) for selected consonant clusters. Each entry is the average of 15 tokens. Note that burst duration is equal to the VOT for voiced stops.

b	7	br	12	bl	9		
d	14	dr	23				
g	23	gr	32	gl	23		
p	–	pr	–	pl	–		
t	24	tr	31			tw	22
k	37	kr	41	kl	23	kw	20
sp	–	spr	–	spl	–		
st	15	str	21				
sk	19	skr	16			skw	13

measurements were found to be more difficult to make and showed more variation than VOT measurements because of the temporal overlap of many frication and aspiration spectra.

The burst durations of /b d g/ average to be approximately 9, 19, and 26 ms, respectively. Burst durations are somewhat longer for voiceless than for voiced plosives. Frication bursts were also more intense in voiceless plosives, but this additional acoustic cue to the voicing distinction in plosives could not be accurately quantified from the spectrographic representation.

Comparison of burst duration (Table VII-3) and voice onset time (Table VII-1) for each of the voiceless plosives shows that both measurements increase from /p/ to /t/ to /k/. The difference between voice onset time and burst duration is the length of the aspiration interval. The data indicate that the aspiration duration is approximately the same for labial, dental, and velar plosives with identical phonetic environments.

Burst durations in the clusters [kw, kl, tw, skw] appear to be shorter than in the environment where the plosive is followed by a vowel. It may be, however, that the frication burst is simply less intense and not as visible in these clusters because of the low first- and second-formant frequencies in /w/ and /l/.

c. Aspiration Following Voiceless Fricatives

Aspiration noise is generated at the glottis whenever (i) the vocal cords are sufficiently spread to inhibit voicing (or to create a nonvibrating glottal chink), (ii) this glottal aperture is smaller than the minimum cross-sectional area of the supraglottal vocal tract, and (iii) the subglottal pressure[5] is greater than a few centimeters of H_2O. In *principle*, these conditions may occur in any transition between phonetic segments that differ in the voicing feature. For example, in a transition between /s/ and a following vowel, the tongue-tip release gesture must be coordinated with an adducting motion of the vocal cords so that there is a rapid reduction in mouth pressure at the same time that the vocal cords attain a position appropriate for the initiation of vibration. If vocal-cord adduction is earlier than normal, simultaneous frication and voicing may occur (or frication production may be inhibited prematurely). If vocal-cord adduction is later than normal, an interval of aspiration generation will be interposed between the end of the fricative and the onset of voicing. The data presented in Table VII-4 indicate that an aspiration interval can be measured following the cessation of visible frication energy in an /s/, and this interval becomes as much as 30-50 ms if the /s/ is followed by a sonorant consonant. An example is shown in Fig. VII-2. The 12 ms average aspiration interval when /s/ is followed by a vowel may not be an accurate indication of aspiration; it may simply reflect the fact that the /s/ frication noise intensity became less than the lower threshold of the spectrograph paper. Nevertheless, it is clear from changes in the spectral composition of the noise that a prolonged interval of true aspiration is present in the /s/-sonorant transitions.

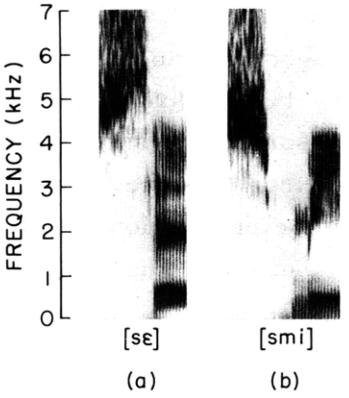

Fig. VII-2.

Broadband spectrograms of the initial consonants excised from utterances containing the words (a) "set" and (b) "smear." An interval of aspiration can be seen following the /s/ in "smear."

Table VII-4. Aspiration duration (in ms) in consonant clusters when /s/ is followed by a sonorant consonant. Each entry is the average of 15 tokens.

s	12
sw	49
sm	38
sn	32
sl	29

3. Discussion

There is a wide variation in the average voice-onset time for plosives in different consonant cluster environments. The average VOT for voiced plosives ranged from 7 ms for /b/ to 32 ms for /gr/. The average VOT for voiceless plosives ranged from 11 ms for /sp/ to 104 ms for /tw/. The considerable overlap in average VOT for plosives normally categorized in English as voiced and voiceless implies that a perceptual decision about the voicing feature for an English plosive cannot be made on the basis of absolute VOT alone.

Overlap in the VOT distributions primarily involves the category that in traditional phonetic theory has been called a "voiceless unaspirated plosive." This allophone is considered an instance of the English voiced phoneme in the phrase "this boy" and an instance of the voiceless phoneme in the phrase "the spy."[6]

In the articulatory domain there are good reasons for maintaining that the so-called voiceless unaspirated plosive is voiced in the first example and voiceless in the second. Evidence that "voiceless unaspirated" is a covering term for two distinct phonetic categories comes from direct observation of the temporal course of glottal adduction[7] and model studies of the conditions under which vocal-cord vibrations begin.[8,9] The vocal cords are adducted approximately 50 ms prior to plosive release even when /b d g/ are not prevoiced,[10] but completion of vocal-cord adduction is delayed until shortly after plosive release in /sp st sk/.[7] The vocal cords are also thought to be approximated and slack in all allophones of /b d g/ in order to facilitate voicing during closure or rapid voicing onset following release, whereas the vocal cords are spread and stiffened in all allophones of /p t k/ in order to prevent vibration during closure.[11] In fact, Halle and Stevens[11] have abandoned the phonetic feature ±voiced in favor of the glottal features ±spread, ±constricted, ±stiff and ±slack in order to express these and other universal phonetic distinctions.

The observation that two distinct voiceless unaspirated plosives occur in English prompted us to look in detail at our acoustic data and data from other investigators to see whether there are acoustic cues in addition to VOT that could signal the voicing distinction directly. If unambiguous acoustic cues are present, a listener would not have to determine the identity of the preceding phonetic segment and word boundary locations before deciding whether a given plosive is phonemically voiced or voiceless.

a. Acoustic Cues for the Voicing Distinction in English Plosives

The sequence of events that takes place during the production of voiced and voiceless plosives has been studied by several investigators.[9,12-14] It would appear from their descriptions and our observations that the following acoustic cues can contribute to a voicing decision.

(i) Voice onset time relative to plosive release

If voicing onset is delayed more than 20-25 ms relative to plosive release, burst and voicing onset are perceived as two separate events.[15] If the VOT is less than 20 ms, the burst and voicing onsets are perceived as occurring simultaneously.

(ii) Presence or absence of aspiration

Aspiration noise has a source spectrum that differs considerably from the spectral characteristics of the voicing source.[5] The spectrum of the source of normal voicing is a line spectrum with greatest energy concentrated at low frequency; that is, higher harmonics fall off in intensity at a rate of ~-12 dB/octave of frequency increase. The spectrum of the aspiration noise source has nearly equal energy at all frequencies between ~0.5 kHz and 4 kHz and intensity falls off gradually outside this range. The two sources produce approximately equal output energy at frequencies in the neighborhood of 2-3 kHz. Therefore, at frequencies near the first formant, the aspiration source output is 12-18 dB less than the voicing source. The spread glottis also increases the bandwidth of the first formant during aspiration production, with the net result that the first formant is very weakly excited by aspiration. The apparent disappearance of the first formant during aspiration has been called the first-formant cutback by Liberman et al.[16]

Aspiration is an aperiodic signal. Perceptual experiments have shown that the auditory system is sensitive to the presence or absence of periodicity.[17] The spectral difference in the region of the first formant and the presence of periodicity indicate the transition from aspiration to voicing.

(iii) Rapid spectral change following voicing onset

The formant transitions that signal place of articulation are nearly completed before voicing onset if the VOT is greater than approximately 25-35 ms. The presence or absence of a rapid spectrum change following voicing onset has been shown to have a strong influence on voicing perception.[18]

(iv) Intensity and duration of the burst of frication

If the glottis is spread at plosive release, the intensity[19] and the duration of the plosive burst are larger than in the corresponding voiced plosive. Increases in both intensity and duration contribute to the perceived loudness of the burst.[20]

(v) Duration of the preceding segment

In English, a voiced segment appearing before a voiceless plosive has approximately 2/3 the duration it would have before a voiced plosive.[21] The difference in duration is less (~20-30 ms) in several other languages.[22] English seems to have expanded upon a universal phonetic tendency to shorten a vowel before a voiceless consonant. A cue of longer duration would seem to allow the unambiguous devoicing of post-stressed voiced fricatives.[23] Neutralization of other cues can be expected in post-stressed plosives if the vowel-duration cue is present.

One possible reason for a universal tendency to encroach on the duration of a vowel if a voiceless segment follows concerns the relative timing of oral closure and glottal opening gestures at vowel termination. A slightly delayed glottal opening gesture is likely to produce a few cycles of vibration during closure, which is a cue to voicing. Since perfect synchrony of glottal and supraglottal activity is impossible, it is proposed that the glottis is normally opened somewhat early in order to avoid generation of a false voicing cue.

If our hypothesis is correct, one would expect that a portion of the formant motions that indicate place of articulation for final voiceless plosives would be truncated. Indeed, second-formant loci are somewhat abbreviated in final voiceless, as opposed to voiced, stops.[24] If the glottis is sufficiently open at closure, a weak burst of frication might also be seen, on occasion, immediately prior to closure. Figure VII-3 illustrates an example of this situation.

(vi) Average fundamental frequency and pitch skip

The average fundamental frequency (F_o) is approximately 5% higher in a vowel placed between voiceless plosives than in a vowel placed between voiced plosives.[25] The fundamental frequency is also lower during the closure interval for an intervocalic voiced plosive.[26] Presumably the vocal cords continue to be somewhat more slack in a voiced consonantal environment and more stiff between voiceless consonants because of coarticulation, and this accounts for the observed differences in fundamental frequency.

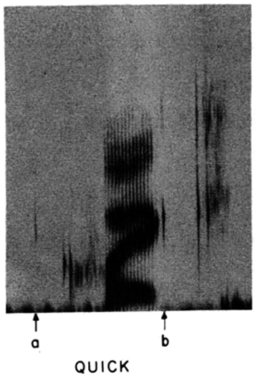

Fig. VII-3.
Broadband spectrogram of the word "quick" as spoken in isolation. A brief burst of frication noise can be seen immediately preceding both stop closures (time points "a" and "b") in this utterance.

Lea[27] has measured the fundamental frequency trajectory in the first 50 ms following voicing onset. He found an average 8% rise in F_o if the preceding plosive was voiced, and an average 7% fall if the preceding consonant was voiceless. Haggard et al.[28] have shown that these rapid upward or downward pitch skips are capable of transmitting the voicing distinction in synthetic plosive-vowel stimuli in which all other cues have been neutralized.

(vii) Prevoicing

Prevoicing of a voiced plosive during closure generates a voice bar on the spectrogram. The spectrum of prevoicing contains only low-frequency harmonics because the first formant is low, sound radiation through the tissues attenuates higher frequencies, and the glottal mode of vibration may differ from normal vowel production.

(viii) Plosive duration

In prestressed position, voiced and voiceless plosives have approximately the same closure duration.[1] In preunstressed position, a voiced plosive becomes approximately 2/3 the duration of the voiceless plosive.[29] Dental plosives, both voiced and voiceless, may be flapped in intervocalic preunstressed position. Their duration is only ~20-30 ms if the tongue tip is positioned to produce a single aerodynamic flap.

b. In Search of Acoustic Invariance

Seven acoustic cues to the voicing feature in English plosives have been described. Some are clearly more important than others. Primary cues whose presence in prestressed CV syllables tends to override conflicting values for the other parameters include perceptual simultaneity of burst and voicing onset, and presence or absence of rapid spectral change.

In other phonetic environments the primary cues may be neutralized. One might speculate that, during the acquisition of language, the presence of the primary cues is absolutely required because these cues are most closely coupled to a set of basic acoustic feature detectors, each of which determines the presence or absence of a given property.[15, 30] As language acquisition progresses, associations are formed between the primary cues and other, secondary cues to voicing which appear in the canonical prestressed CV syllable environment. At a later developmental stage, the secondary cues may acquire greater perceptual importance and contribute to a voicing decision in post-stressed or consonant cluster environments. The importance of the secondary cues to voicing is difficult to assess, however, because it may be that lexical, syntactic, and semantic constraints take over when the primary cues are neutralized, or perhaps we learn more complex decoding strategies with regard to the primary cues.

The first five acoustic cues listed above are useful in distinguishing the voicing feature in prestressed CV syllables. In plosive-sonorant clusters such as /gr/, voice onset times for voiced plosives frequently exceed the simultaneity threshold, so that other cues must take over. In /s/-plosive clusters, both primary cues suggest the presence of a phonemically voiced plosive, but F_o, burst loudness, and possibly aspiration cues have the potential of countermanding this evidence. In preunstressed environments the voicing distinction may be neutralized, but if it is not, the same general relations hold in comparable segmental environments.

Two secondary cues, F_o and burst loudness, appear to satisfy acoustic invariance considerations, in that they tend to indicate the correct phonemic categorization of plosives in all phonetic environments. True acoustic invariance, however, would seem to require that these cues acquire primary status, and this is not consistent with psychophysical data. It must be concluded that true acoustic invariance does not exist. The primary cues can be used directly during acquisition of the voicing distinction in CV environments, but other environments require the mediation of information about the acoustic and phonetic characteristics of segments adjacent to the plosive.

c. Influence of Mouth Pressure on Voice Onset Time

Theoretical calculations indicate that the time of voicing onset following a voiceless interval is a function of 3 primary variables: transglottal pressure, static glottal opening, and vocal-cord tension.[9, 11] Voicing onset is delayed when the transglottal pressure increases more slowly, the static glottal opening is greater or the vocal cords are stiff.

We should be able to explain the changes in VOT in various consonant clusters in terms of these variables. For example, the VOT for both voiced and voiceless plosives increased when the plosive was followed by a sonorant consonant rather than a vowel. There was a 20% increase in VOT of voiceless plosives followed by the high vowels /i, u/ as opposed to the nonhigh vowels /ɛ, a^y/, and there was a 30-50 ms delay in voicing onset in /s/-sonorant clusters. Each of these delayed VOT effects shares a common attribute, a less rapid increase in the cross-sectional area of the oral constriction, which suggests that all may result from the same pattern of active or passive forces. On the surface, passive forces seem to be involved because the just-cited VOT changes tend to increase the variance in the VOT distributions, whereas an active mechanism might be expected to minimize variance-inducing effects. The most likely passive mechanism is related to the longer durations of frication bursts in these phonetic environments. Frication noise is generated only while the mouth pressure behind the oral constriction is significantly above atmospheric pressure. We detail below a feedback mechanism whereby the glottal closing gesture is triggered by a sudden reduction in mouth pressure as sensed at the upper surface of the vocal cords. Thus any mechanism that prolongs the duration of the frication burst in a voiceless plosive can be expected to increase voice onset time as well.

d. Plosive Release Velocity and Burst Duration

The inherent differences in burst durations for labials, dentals, and velars may be explained by observing the time course of the pressure developed across the oral closure following release. A labial release is quite rapid and the generated burst spectrum is weak in intensity. Both factors contribute to the acoustic appearance of a short burst.

A velar release involves the entire tongue body and is relatively slow. This is due in part to the mass involved, and in part to the fact that the release vector of the tongue motion is usually not perpendicular to the long dimension of the acoustic tube formed by the vocal tract,[31] and the effective release velocity is only $|v| \cos \theta$. The release vectors of the tongue tip and lips are normally more nearly perpendicular. In the clusters [tr, dr, str], however, burst durations are longer and the tongue tip probably executes a sliding motion in these clusters.

e. Feedback Control of VOT in Voiceless Plosives

An interesting result of this study is the observation that VOT differs for /p t k/ if measured relative to plosive release, but is nearly equal for /p t k/ if measured relative to the end of the burst of frication. The end of the burst corresponds to the time when oral pressure is falling rapidly toward zero. The close correlation between oral pressure fall and the subsequent glottal closing gesture suggests a possible feedback mechanism by which the appropriate timing between supralaryngeal and glottal gestures is learned and maintained. Since completion of the glottal closing gesture requires approximately 40-80 ms,[13] the oral pressure drop would appear to occur too late to trigger glottal adduction in the speech of adults. Children acquiring the voiceless plosives typically use much greater VOT values[32] and oral pressure feedback is not implausible.

A second possible explanation for the differences in VOT for /p t k/ is that the VOT is delayed until the formant transitions that signal place of articulation are nearly over. Then there is no rapid spectral change following voicing onset.

References and Footnotes

1. D. H. Klatt, "Durational Characteristics of Prestressed Word-Initial Consonant Clusters in English," Quarterly Progress Report No. 108, Research Laboratory of Electronics, M.I.T., January 15, 1973, pp. 253-260.

2. L. Lisker and A. S. Abramson, "A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements," Word 20, 384-422 (1964).

3. L. Lisker and A. S. Abramson, "Some Effects of Context on Voice Onset Time in English Stops," Language and Speech 10, 1-28 (1967).

4. K. N. Stevens, "The Quantal Nature of Speech: Evidence from Articulatory-Acoustic Data," in P. B. Denes and E. E. David, Jr. (Eds.), Human Communication: A Unified View (McGraw-Hill Book Company, New York, 1972), pp. 51-66.

5. K. N. Stevens, "Airflow and Turbulence Noise for Stop and Fricative Consonants: Static Considerations," J. Acoust. Soc. Am. 50, 1180-1192 (1971).

6. J. E. Shoup, "The Phonemic Interpretation of Acoustic-Phonetic Data," Ph.D. Thesis, University of Michigan, Ann Arbor, Michigan, 1964.

7. M. Sawashima, A. S. Abramson, F. S. Cooper, and L. Lisker, "Observing Laryngeal Adjustments during Running Speech by Use of a Fiberoptics System," Phonetica 22, 193-201 (1970).

8. K. Ishizaka and M. Matsudaira, "Fluid-Mechanical Considerations of Vocal Cord Vibration," Monograph No. 8, Speech Communication Research Laboratory, Inc., Santa Barbara, California, 1972.

9. K. N. Stevens, "An Analysis of Glottal Activity during Consonant Production" (in preparation for publication).

10. Voicing will occur prior to plosive release if a sufficiently large transglottal pressure is realized during closure. Voicing commonly occurs during closure if the voiced plosive is preceded by a voiced segment. Voicing may begin during the closure if the plosive is preceded by a voiceless segment or in utterance-initial position, but it is more likely that oral closure is completed prior to adduction of the vocal cords and pressure behind the oral constriction thus has had time to approach the subglottal pressure. Then the transglottal pressure is insufficient to initiate voicing until the plosive is released. At release, the vocal cords are in an ideal state to induce voicing and a very short VOT results.

11. M. Halle and K. N. Stevens, "A Note on Laryngeal Features," Quarterly Progress Report No. 101, Research Laboratory of Electronics, M.I.T., April 15, 1971, pp. 198-213.

12. C. G. M. Fant, Acoustic Theory of Speech Production (Mouton and Company, 's-Gravenhage, The Netherlands, 1960), p. 199.

13. M. Rothenberg, "The Breath-Stream Dynamics of Simple-Released Plosive Production," Bibliotheia Phonetica 6 (S. Karger, Basel, 1968).

14. C. W. Kim, "A Theory of Aspiration," Phonetica 21, 107-116 (1970).

15. K. N. Stevens and D. H. Klatt, "The Role of Formant Transitions in the Voiced-Voiceless Distinction for Stops" (in preparation for publication).

16. A. M. Liberman, P. C. Delattre, and F. S. Cooper, "Some Cues for the Distinction between Voiced and Voiceless Stops in Initial Position," Language and Speech 1, 153-167 (1958).

17. I. Pollack, "Periodicity Pitch for Interrupted White Noise – Fact or Artifact," J. Acoust. Soc. Am. 45, 237-238 (1969).

18. K. N. Stevens and D. H. Klatt, "The Role of Formant Transitions in the Voice-Voiceless Distinction for Stops," Quarterly Progress Report No. 101, Research Laboratory of Electronics, M. I. T., April 15, 1971, pp. 188-197.

19. M. Halle, G. W. Hughes, and J. P. A. Radley, "Acoustic Properties of Stop Consonants," J. Acoust. Soc. Am. 29, 107-116 (1957).

20. G. A. Miller, "The Perception of Short Bursts of Noise," J. Acoust. Soc. Am. 20, 160-170 (1948).

21. G. E. Peterson and I. Lehiste, "Duration of Syllabic Nuclei in English," J. Acoust. Soc. Am. 32, 693-703 (1960).

22. M. Chen, "Vowel Length Variation as a Function of the Voicing of the Consonant Environment," Phonetica 22, 129-159 (1970).

23. P. B. Denes, "Effect of Duration on the Perception of Voicing," J. Acoust Soc. Am. 27, 761-764 (1955).

24. D. J. Broad and R. H. Fertig, "Formant Frequency Trajectories in Selected CVC-Syllable Nuclei," J. Acoust. Soc. Am. 47, 1572-1582 (1970).

25. A. S. House and G. Fairbanks, "The Influence of Consonantal Environment upon the Secondary Acoustical Characteristics of Vowels," J. Acoust. Soc. Am. 25, 105-113 (1953).

26. L. A. Chistovich, "Variations of the Fundamental Voice Pitch as a Discriminatory Cue for Consonants," Sov. Phys. – Acoust. 14, 372-378 (1969).

27. W. A. Lea, "Influences of Phonetic Sequences and Stress on Fundamental Frequency Contours of Isolated Words," J. Acoust. Soc. Am. 53, 346 (A) (1973).

28. M. Haggard, S. Ambler, and M. Callow, "Pitch as a Voicing Cue," J. Acoust. Soc. Am. 47, 613-617 (1970).

29. L. Lisker, "Stop Duration and Voicing in English," Status Report SR-19/20, Haskins Laboratories, New York, 1969, pp. 27-35.

30. P. D. Eimas, E. Siqueland, P. Jusczyk, and J. Vigorito, "Speech Perception in Infants," Science 171, 303-304 (1971).

31. R. A. Houde, "A Study of Tongue Body Motions during Selected Speech Sounds," Monograph No. 3, Speech Communication Research Laboratory, Inc., Santa Barbara, California, 1968.

32. D. K. Port and M. S. Preston, "Early Apical Stop Production: A Voice Onset Time Analysis," Status Report SR-29/30, Haskins Laboratories, New York, 1972, pp. 125-149.