THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 54, No. 3, March 1975
Printed in U.S.A.

Synthesis of Speech From a Dynamic Model of the Vocal Cords and Vocal Tract

By J. L. FLANAGAN, K. ISHIZAKA, and K. L. SHIPLEY

(Manuscript received July 17, 1974)

We describe a computer model of the human vocal cords and vocal tract that is amenable to dynamic control by parameters directly identified in the human physiology. The control format consequently provides an efficient, parsimonious description of speech information. The control parameters represent subglottal lung pressure, vocal-cord tension and rest opening, vocal-tract shape, and nasal coupling. Using these inputs, we synthesize vowel-consonant-vowel syllables to demonstrate the dynamic behavior of the cord/tract model. We show that inherent properties of the model duplicate phenomena observed in human speech; in particular, cord/tract acoustic interaction, cord vibration, and tract-wall radiation during occlusion, and voicing onset-offset behavior. Finally, we describe an approach to deriving the physiological controls automatically from printed text, and we present sentence-length synthesis obtained from a preliminary system.

I. INTRODUCTION

Speech sounds can be synthesized by a variety of means used to construct signal waveforms. Many ingenious methods have been recorded. But speech synthesis generally has the practical purpose of producing intelligible sounds from control data that are as parsimonious as possible. In other words, the control data should represent an efficient, concise coding of the speech information. This motivation applies as much to analysis/synthesis techniques for speech transmission as to computer voice-response systems which strive for efficient vocabulary storage and high versatility in message fabrication.

Because speech is a human-generated signal, it is unlikely that a synthesis method can achieve the ultimate parsimony of input control without considerable attention to the parameters a human overtly manipulates in speaking. That is, one increases the information "built into" the synthesizer when its design exploits fundamental properties of the human speech mechanism.

We therefore have chosen an approach to synthesis with which we can identify overtly the significant physiological parameters important

in speech production. Major system components obviously are the mechanism of voiced-sound generation and the mechanism for intelligibly modulating sound timbre, that is, the vocal cords and the vocal tract. Our approach, unlike that found conventionally in the speech literature, is not to make a linear separation of the sound source and vocal tract. More than this, we believe that source/tract interaction actually contributes built-in natural behavior that is significant in synthesis. This natural interaction is missing in approaches that assume linear separation of source and tract (unless provided at additional expense and coding effort in the input data).

The initial results stemming from this approach to synthesis are described below.

II. ACOUSTIC MODEL OF VOCAL CORDS AND VOCAL TRACT

We view the acoustic system of the human vocal cords and vocal tract as shown at the top of Fig. 1. The lungs are an air reservoir, maintained at subglottal air pressure P_s by contraction of the rib-cage muscles. The subglottal pressure is applied via the bronchi and trachea passages to the variable-area orifice controlled by the vocal cords.

We model the cords as an acoustic-mechanical oscillator, wherein a single vocal cord is described by two masses, each having an associated stiffness and loss, which are "internally" coupled by a third stiffness. In previous work,[1–4] we established the philosophy leading to this description and gave a quantitative analysis of the vocal cord model.

Oscillation of the vocal cord model results in the glottal volume velocity U_g. This quantity typically has an impulsive waveform and it is the excitatory source for voiced sounds.

The vocal tract proper is a nonuniform tube, about 17 cm long in man, extending from the cords to the mouth. Its cross-sectional area varies from zero to upwards of 20 cm^2. The nasal tract is an ancillary tube about 60 cm^3 in total volume and coupled to the vocal tract by the trap-door action of the velum. Sound is radiated from the system as a result of the volume velocities at the mouth U_m and nostril U_n, and from vibration of the yielding sidewalls of the vocal tract.

Cross-dimensions of the acoustic system are small compared to sound wavelengths of interest, and hence we confine our analysis to plane-wave propagation in the tract. We therefore represent the acoustic system as the bilateral, time-varying transmission line shown in the lower part of Fig. 1. Formulation of this system follows that given by Flanagan.[5]

As illustrated, the lossy lung volume is "charged" to subglottal lung pressure P_s, which is applied via the trachea-bronchi network, to the glottal (vocal-cord opening) impedance Z_g. This nonlinear glottal

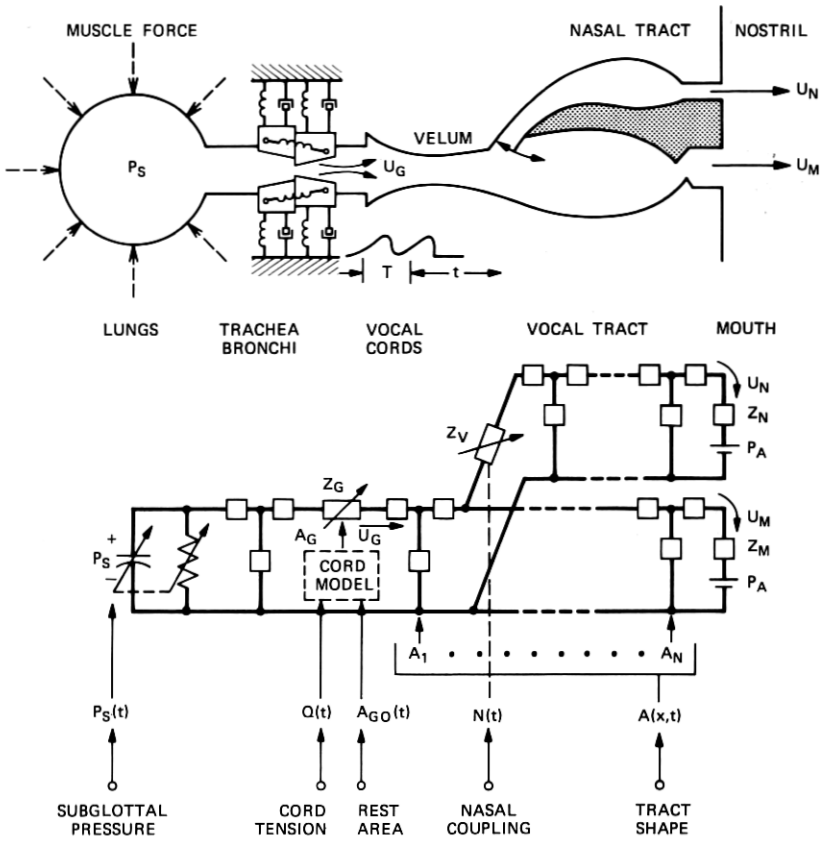

Fig. 1—Schematic diagram of the vocal cord/vocal tract system.

impedance depends upon the glottal flow and area A_g, which in turn depend upon the self-oscillating properties of the vocal cord model described in detail in an earlier paper.[1] The resulting volume flow U_g is the excitation source for the vocal and nasal tracts.

The shape of the vocal tract is defined by its cross-sectional area as a function of distance $A(x)$, and the coupling to the time-invariant nasal cavity is governed by the velar impedance Z_v. Volume velocity at mouth U_m and nostril U_n flow through their respective radiation impedances Z_m and Z_n, both of which are in series with batteries representing the constant atmospheric pressure P_a. (This formulation permits simulation of respiration as well.) The mouth and nostril radiation impedances are those for a circular piston in an infinite baffle.[5]

Parameters of control for the speech synthesis system are the physiologically-based functions shown in Fig. 1. All vary with time.

They are subglottal lung pressure P_s, vocal-cord tension Q, rest (or neutral) area of cord opening A_{go}, nasal coupling N, and cross-sectional area function of the tract shape $A(x)$. We are concerned here only with nonnasal sounds, hence nasal coupling will not figure in the discussion.

Each T-section of the vocal-tract transmission line is represented in Fig. 2.[5] An elemental length Δx of the vocal tube has cross-sectional area A, terminal sound pressures p_1 and p_2, and terminal volume velocities U_1 and U_2. The sidewall has noninfinite mechanical impedance, and vibrates in response to the enclosed sound pressure with displacement ξ. This displacement radiates a per-unit-length sound pressure p_{wall}. Relations between terminal values of pressure and volume

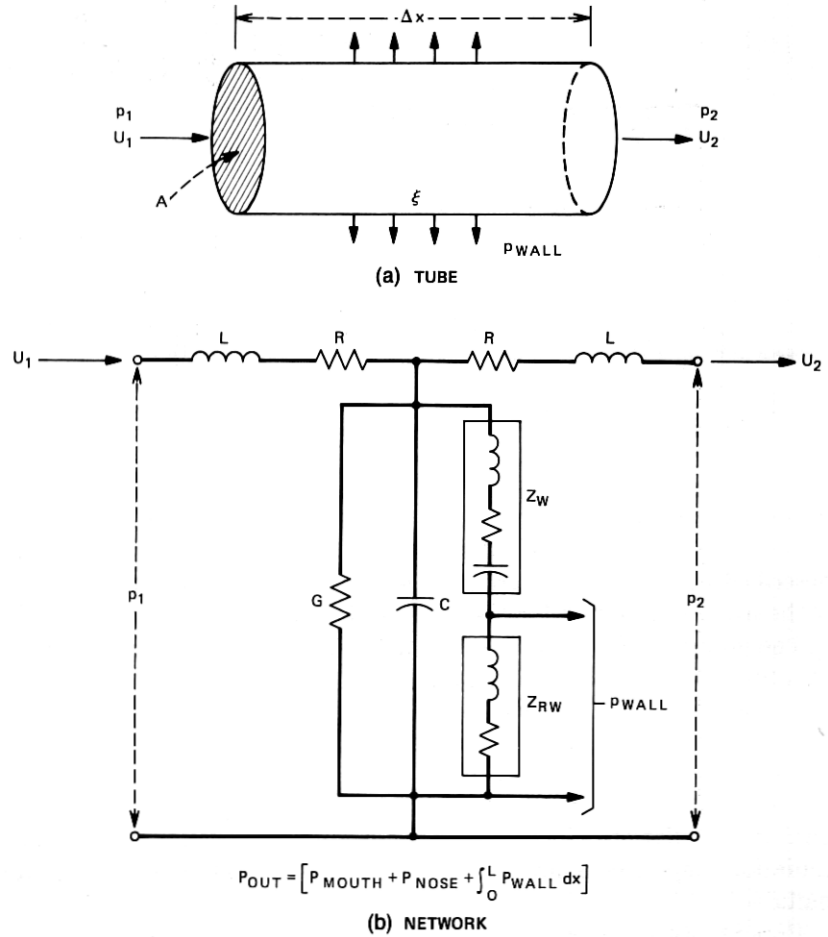

Fig. 2—Circuit representation of plane acoustic wave propagation in an elemental length of tube with yielding sidewalls.

velocity for plane-wave propagation are given in the circuit in Fig. 2b, in which L is the per-unit-length inertance of the air mass of the tube element, R the viscous loss at the sidewall, G the heat-conduction loss, C the acoustic compliance of the contained air volume, Z_w the acoustic equivalent mechanical impedance of the yielding wall, and Z_{rw} the radiation impedance of the wall, assumed to be that for a pulsating right circular cylinder.[6]

The total sound output from the model is, following the long-wave assumptions, the linear superposition of the mouth and nostril radiation plus the spatially summated wall radiation.

In addition, every T-section of the transmission-line network includes a means for introducing turbulent noise excitation. This capability is provided by a series random pressure source P_N with its internal resistance R_N, as shown in Fig. 3. This technique has been given in detail previously.[4] The intensity (or rather mean-square variance) of the random pressure source is controlled by the Reynolds number of the flow at each network section, while the internal resistance is similarly modulated according to the Bernoulli loss in a constriction.[5] In both instances, the specified value of cross-sectional area A and the calculated resulting volume velocity flowing through the serial source completely describe the control functions. That is, no additional input data are required.

More specifically, to simulate the conditions of turbulent-source generation in any section, the amplitude of the noise pressure is made directly proportional to the squared Reynolds number in excess of a

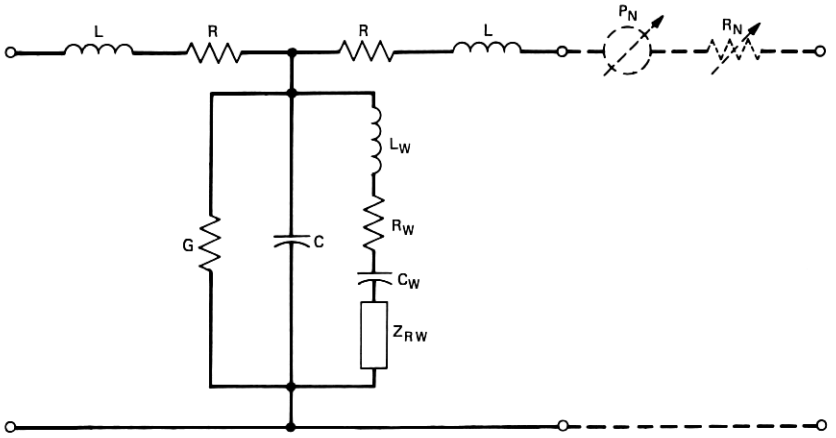

Fig. 3—Circuit representation of the turbulent noise source for each network section. The intensity of the random pressure source, P_N, and its self-resistance, R_N, are controlled by the volume velocity and cross-sectional area at each section.

critical (threshold) Reynolds number for turbulent flow.[4] The squared Reynolds number is proportional to U^2/A, whereas the internal resistance of the turbulent source is, to first order, a flow-dependent loss proportional to $|U|/A^2$. Therefore, as the prescribed section area becomes small in the presence of a large flow-velocity, turbulence conditions are indicated and the intensity of the noise source and the value of its internal resistance are increased. In the simulation, values of every dependent variable are calculated on a sample-by-sample basis to construct the time functions for the output sound pressure and all other pressure and velocity quantities.[4]

As a consequence of continually noting the magnitude of the calculated volume flow in each section and having the tract cross-sectional areas continually prescribed as input data, the synthesizer *automatically* introduces random noise excitation in any section when the Reynolds number is sufficiently high to indicate turbulent flow. The synthesizer, therefore, requires no additional data to produce voiceless sounds, but uses exactly the same control parameters to generate both voiced and voiceless sounds (or combinations of voiced and voiceless sounds). As Fig. 1 has shown, these control parameters are P_s, Q, A_{g0}, N, and $A(x)$.

As a practical matter in the computer implementation, we use a P_N source produced from gaussian noise (or, rather, gaussian numbers) bandpass-filtered from 500 to 4,000 Hz. Further, to insure stability, the volume flow which modulates the serial noise source is low-pass filtered to 500 Hz. In other words, the noise source is modulated by low-frequency components of U, including the dc flow.

The transmission line model of Fig. 1 is described by a set of linear and nonlinear differential equations in which all coefficients also vary with time. This set of differential equations is approximated by difference equations, as previously described,[2] and programmed in a laboratory computer for on-line control. Twenty network sections are used to approximate the vocal tract. This formulation has permitted initial experiments with physiologically-based control of the synthesis model.

III. ASSESSMENT OF WALL IMPEDANCE AND EFFECTS ON FORMANT BANDWIDTH

All elements of the transmission line network have been well established in previous work, with the exception of the wall-vibration shunt branch of the circuit in Fig. 2.

Assessment of wall effects in earlier calculations[5] utilized the only available mechanical impedance measurements of human tissue,

namely, chest, stomach, and thigh tissue. These data led to correct order-of-magnitude values for wall-vibration damping of formant resonances, but the values were clearly on the high side.

To better assess the wall impedance, we have done two things. First, we have used data on human formant bandwidths to estimate contributions to losses in our model. And second, we have made direct measurement of the mechanical impedance of the vocal-tract wall.[7]

Formant bandwidths have been measured for the human vocal tract by van den Berg,[8] Bogert,[9] Fujimura and Lindquist,[10] and Dunn.[11] Our programmed model allows us to adjust values of the wall-impedance parameters to effect three consistencies. It permits us to (*i*) adjust the wall-loss component to match glottis-closed formant bandwidths, (*ii*) adjust the inductive reactance of the wall to produce the observed mouth-closed, lowest value of first formant frequency of about 200 Hz, and (*iii*) choose a wall compliance to produce wall resonance substantially below 100 Hz. Small-signal-driven vibration of the cord oscillator in the model permits calculations of model response at any prescribed frequency. Furthermore, formant bandwidths measured on real speech allow additional cross-checks of parameters used in the model formulation, especially for the loss components of the cord-oscillator source. Application of this knowledge in our model yields the formant bandwidth behavior shown in Fig. 4.

In particular, Fig. 4 illustrates how the wall viscous loss parameter can be chosen to match glottis-closed formant bandwidth. This technique has recently been analyzed in extensive quantitative form by Sondhi.[12] The wall loss is selected to match measured formant bandwidths at formant frequencies around 300 to 500 Hz. In this frequency range, the contributions to formant bandwidth are mainly wall loss and glottal source loss. Viscosity, heat conduction at the walls, and mouth radiation resistance represent relatively small values (see Ref. 5, for example, for these calculation techniques). Note, too, that in Fig. 4 the vertically-sloping line of calculated bandwidth indicates the effect of wall impedance on the tuning of formant frequency. In the absence of additional data, we assume a uniform distribution of the per-unit-area wall impedance along the tract. The value we use for the mechanical per-unit-area impedance is $(1600 + j1.5\omega)$g/s/cm^2, where ω is the radian frequency. This value is confirmed well by our direct measurements of wall impedance.[7]

Formant bandwidths measured in real speech[11] permit a cross-check of the glottal oscillator parameters chosen in previous work.[1] Figure 4 shows that the contribution to formant damping of the glottal source falls into the correct range of real speech measurements. This is a

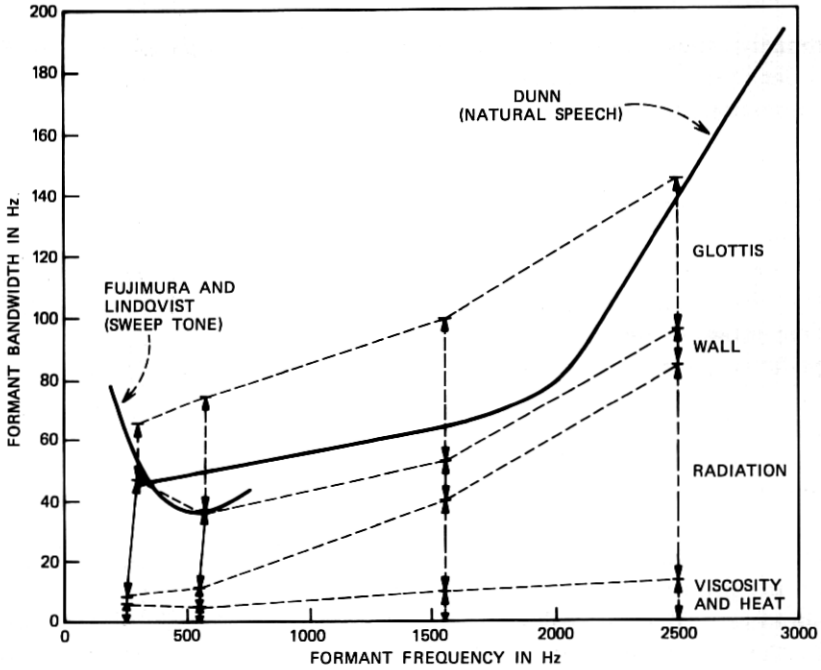

Fig. 4—Variation of formant bandwidth (damping) with formant frequency. The diagram shows the relative contributions of loss to the total formant bandwidth produced by the cord/tract model.

gratifying confirmation of cord parameters chosen strictly on other bases, namely, according to physiological properties and oscillatory behavior.[1]

The loss contributions of the glottal source in Fig. 4 are calculated for a nominal, midrange value of glottal rest area, namely $A_{go} = 0.05$ cm^2. The glottal contribution to formant bandwidth is, of course, a function of A_{go}. Figure 5 shows glottal loss contributions for other values of A_{go}. Note, especially, how the articulatory configuration of the tract influences the contribution of the glottal source to formant damping.

IV. DYNAMIC BEHAVIOR OF THE CORD/TRACT MODEL

How does this physiologically-based model of the vocal cords and vocal tract behave under dynamic control? Time-varying control inputs in the present study are P_s, Q, A_{go}, and $A(x)$. An obvious major problem is the determination of realistic values of these parameters. As a first cut, fairly realistic data can be derived from direct measurements of lung pressure during speech,[13] laryngeal muscle

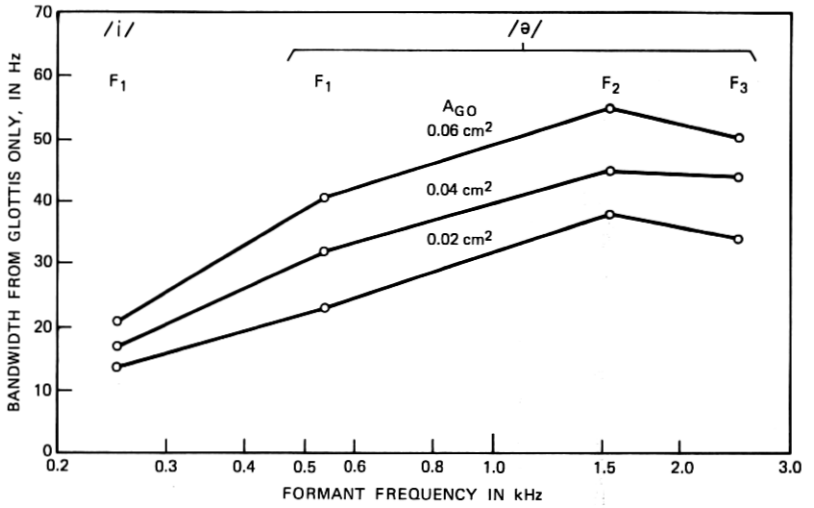

Fig. 5—Variation of the glottal loss contribution to formant bandwidth as a function of formant frequency. The parameter is the neutral glottal area, A_{go}.

electromyography,[13] glottal transillumination,[14] glottal pulses,[15] and cine X-rays of the vocal tract.[16] An important element at present is that all these data are not simultaneously available for a given subject. Experiments now under way aim to provide some simultaneous measurements.[17]

The physiological literature does provide adequate bases for dynamic tests on some simple utterances, using idealized input controls. We have, therefore, made first tests on vowel-consonant-vowel syllables (v-c-v) in which stress may be on either initial or final vowel, and where the intervocalic consonant is a voiced or unvoiced labial stop. These combinations also provide a convenient vehicle for exposing other physiologically realistic properties of the cord/tract model.

Figure 6 shows the synthesis of the syllable /'abə/. Input controls are indicated in the top three traces. Because of the initial stress, P_s falls during the labial closure to a lower value. Because the intervocalic stop is voiced, A_{go} is maintained in a position favorable to cord oscillation throughout. Cord tension, not shown, is also maintained constant. Any pitch changes are effected solely by P_s variation and by the interaction of tract load on the cord oscillator. Articulatory shape $A(x)$ changes from /a → b → ə/. Because of space limitation in the illustration, only the mouth area, A_m, is displayed.

Response behavior of the model to these input controls is shown in the bottom five traces: the sound spectrogram of the total output sound; A_g; U_g; the pressure waveform of the total output sound P;

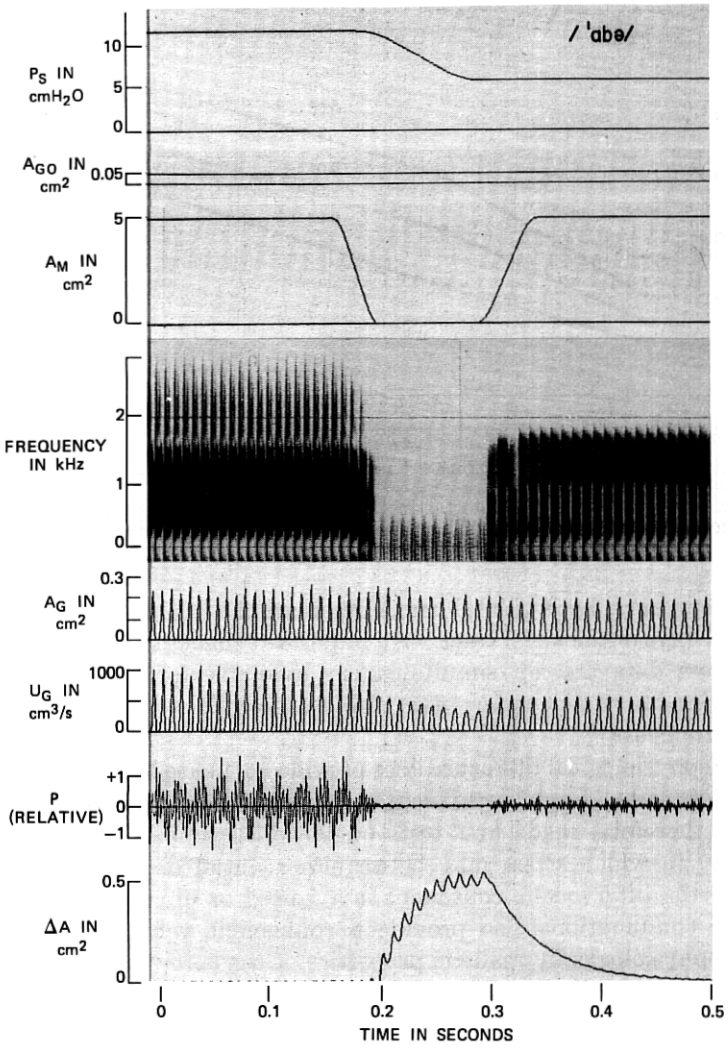

Fig. 6—Control functions and sound output from the cord/tract model synthesizing the syllable /ˈabə/. The effects of sound radiation from the yielding sidewalls is evidenced in synthesized sound, and the vibration of the mouth cavity wall is illustrated by the ΔA trace.

and the incremental change in area ΔA of the oral cavity in response to the contained sound pressure.

Several things are notable. In the sound spectrogram, notice the intense initial vowel /a/ with relatively elevated pitch (about 120 Hz) and with natural formant transition into the stop. Voicing continues

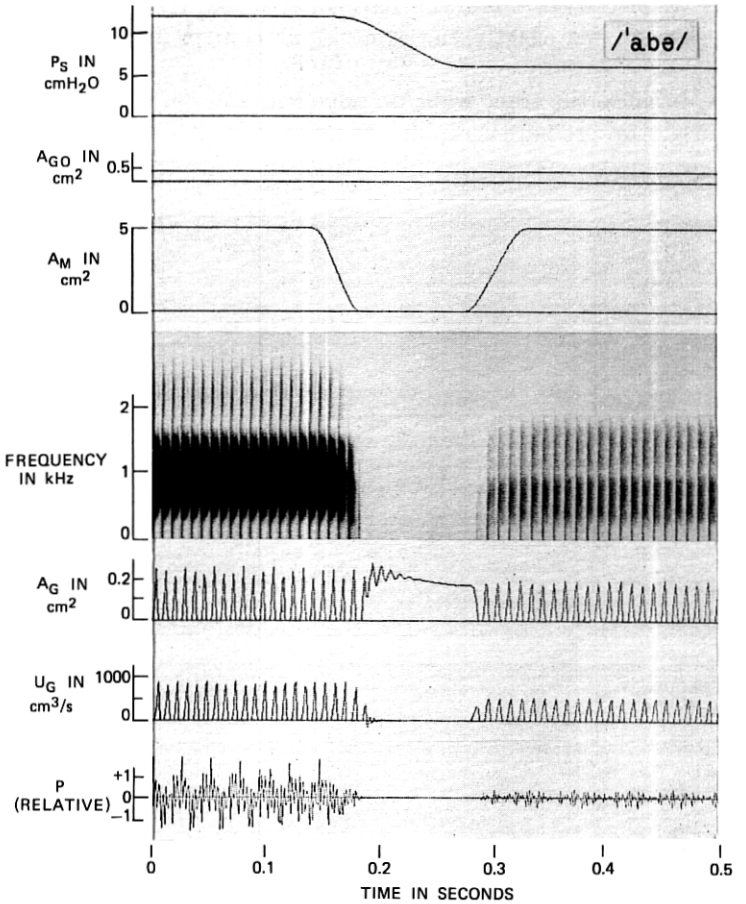

Fig. 7—Behavior of the cord/tract model when the sidewalls are made rigid. The input control functions are identical to those of Fig. 6. The synthesized syllable is /'abə/. Note especially that the cord-oscillator ceases vibration during the mouth closure.

throughout the labial closure, at slightly reduced pitch (about 95 Hz), and with the sound output coming solely from the wall radiation. The sound level during the lip closure is on the order of 20 dB lower than the mouth-radiated vowels. Natural transition into the final vowel follows, with voicing at reduced pitch (104 Hz) and intensity.

The waveforms of the A_g and U_g oscillations confirm the spectrogram display, as does the waveform of output pressure. The wall-radiated sound is dramatized by examining the incremental area change in the yielding-wall oral cavity. The area perturbation is seen to follow pitch-synchronously the glottal pulses of U_g.

It is instructive to contrast this soft-wall behavior with that which obtains when the tract is made hard-walled; i.e., by letting $Z_w \to \infty$. This behavior, for exactly the same input control data, is shown in Fig. 7.

Now, because the tract walls do not yield and permit enlargement, the transglottal pressure is rapidly diminished during the labial closure,

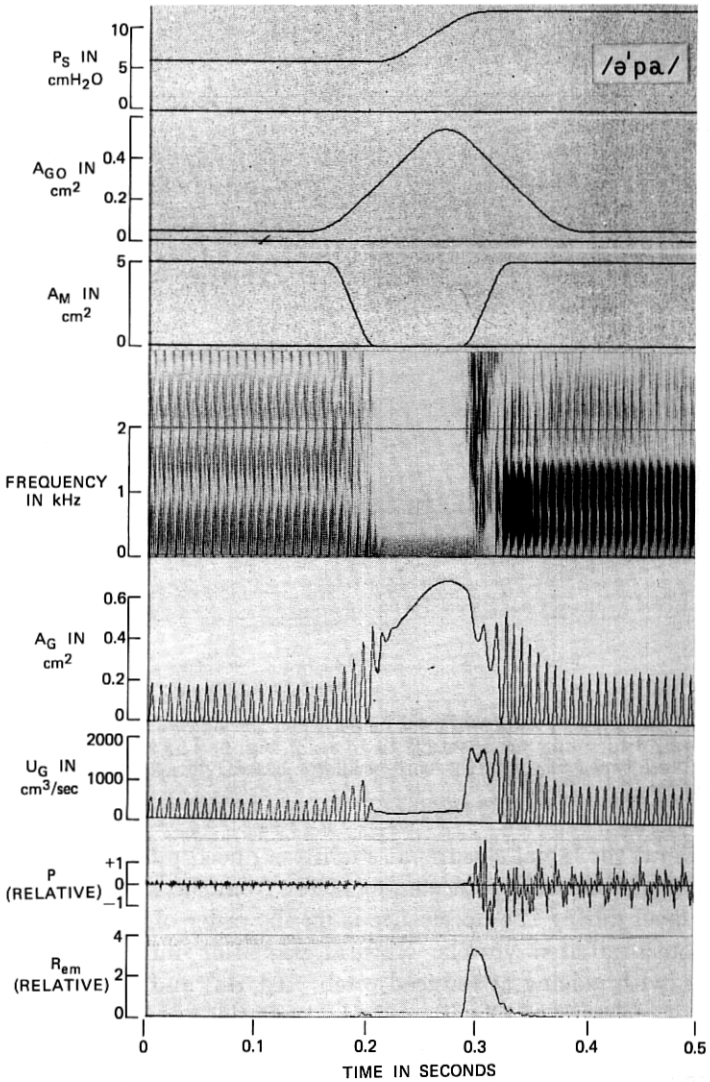

Fig. 8—Control functions and synthesized sound for the syllable /ə'pa/. Sound generation for the voiceless consonant is produced from the distributed random noise sources that are modulated by the Reynolds number for every network section.

and cord oscillation rapidly ceases during the /b/ consonant. Also, no sound is radiated from the tract walls, and only silence prevails during the lip closure. Offset and onset of cord oscillation, with lip closure and release, appears abnormal when compared to transillumination data taken on human vocalization. This latter factor may be more important perceptually than the actual absence of sound during lip closure.

Dynamic behavior for a voiceless intervocalic labial stop is displayed in Fig. 8. The syllable is /ə'pa/, with stressed second vowel. Again, control function input is indicated by the top three traces. Only mouth area A_m is again displayed, and cord tension Q is held constant. Note now, however, the A_{go} control effects voiced-voiceless switching by moving from a value that sustains cord oscillation to one that does not.

The spectrogram of the sound output shows the low-intensity, low-pitch initial vowel with natural formant transitions into the stop. Cord oscillation ceases during the closure because the cords are overtly pulled apart. (The lateral and posterior crico-arytenoid muscles accomplish this in the human larynx.) The cords come back together as the lip closure is released, and oscillation starts with an abrupt bounce that is quite characteristically seen in glottal transillumination data on humans.[18] Natural formant transition is made into the final, high-intensity, relatively-higher-pitched vowel.

The U_g flow continues without cord oscillation through the lip closure, as the tract wall yields and enlarges the volume forward of

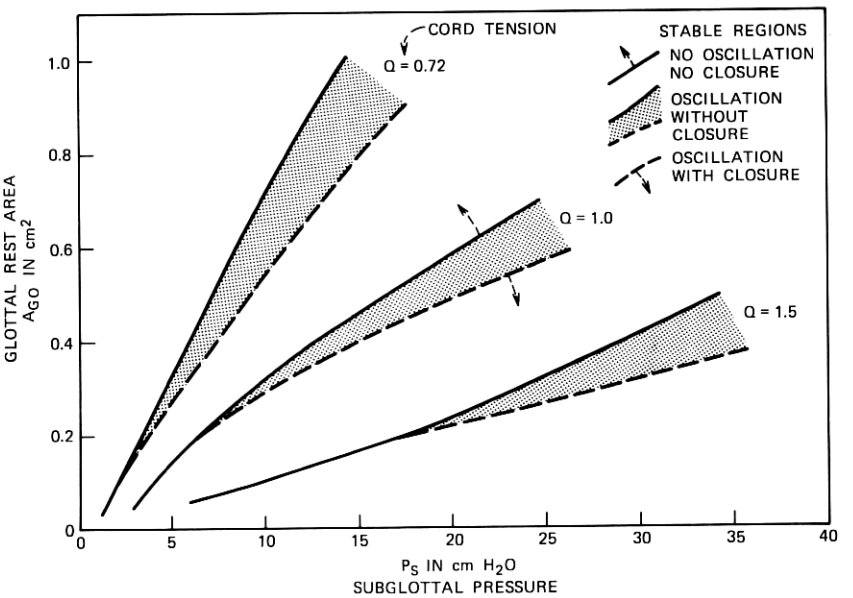

Fig. 9—Regions of stable oscillation for the vocal-cord model.

the cords. As the lips release, the U_g flow reflects a relatively large dc component before oscillation commences. This flow is the source of aspiration in the consonant release.

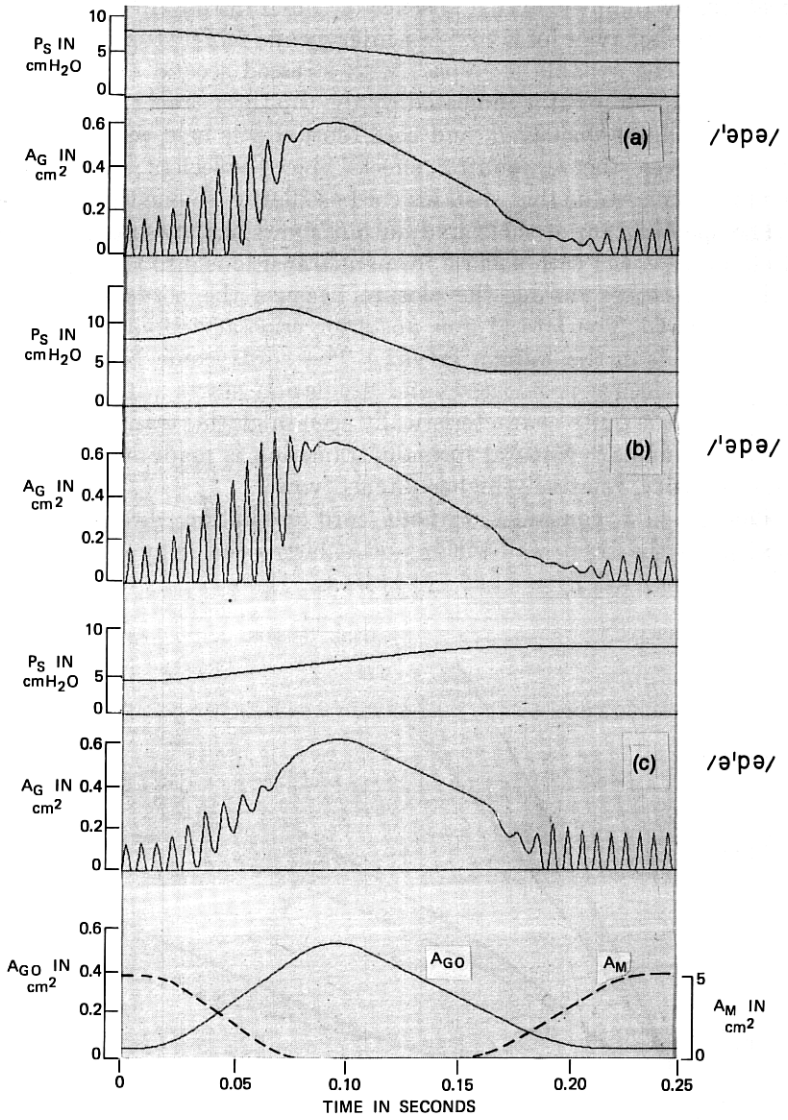

Fig. 10—Behavior of the vocal-cord oscillation as a function of subglottal pressure. The dynamics of tract motion and the control of the vocal cord neutral area are the same for each diagram. Subglottal pressure for conditions (a) and (b) correspond to initial stressed vowel, whereas condition (c) represents final stressed vowel. Note especially the delayed voicing onset in (b).

The automatic turbulence generation is indicated by the lower trace in Fig. 8, which is the squared Reynolds number for the volume flow at the lips. As discussed previously, turbulence (noise) intensity is monotonely related to this function, in excess of a threshold value.[5] The high spike in R_{em} at about $t = 0.3$ s indicates a turbulent burst of noise with approximately this amplitude envelope. The sound output pressure waveform and the spectrogram show the result of this turbulence generation. The result is consistent with aspirated releases seen in the /p/ consonant. Furthermore, auditory assessment of the synthesized sound indicates a natural-sounding syllable.

This synthesis also highlights the importance of the A_{go} control for switching between voiced and unvoiced sounds. A more detailed indication of this behavior is shown in Fig. 9. Three distinct regions of stable cord behavior are indicated. For given cord parameters, stable behavior is determined by the interplay of P_s and A_{go}.

An additional examination of dynamic behavior dramatizes the so-called delayed voicing onset. The syllable /əpə/ is generated with the A_{go} and A_m controls shown at the bottom of Fig. 10. The cord tension, Q, is maintained constant. Lung pressure, P_s, however, is varied to correspond to initially stressed vowels (conditions a and b) and a finally-stressed vowel (condition c). Notice especially in condition b, the initially rising, then abruptly falling P_s conspires with the first opening, and then closing A_{go} control to produce substantial delay in the resumption of voicing. This is found characteristically in human speech.[19]

V. AUTOMATIC GENERATION OF CORD/TRACT CONTROL

Ultimately, we wish to use the cord/tract model as an end-organ for speech synthesis. What are the prospects for obtaining the necessary controls automatically by rule?

In recent work on synthesis-by-rule, Coker and Umeda[20] generated synthetic speech from printed text using programmed algorithms for articulatory dynamics and for speech prosody. Their speaking machine includes a pronouncing dictionary, a syntax and prosody analyzer,

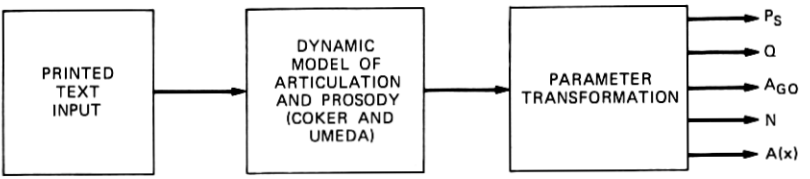

Fig. 11—Automatic generation of control functions for the cord/tract synthesizer from printed text input.

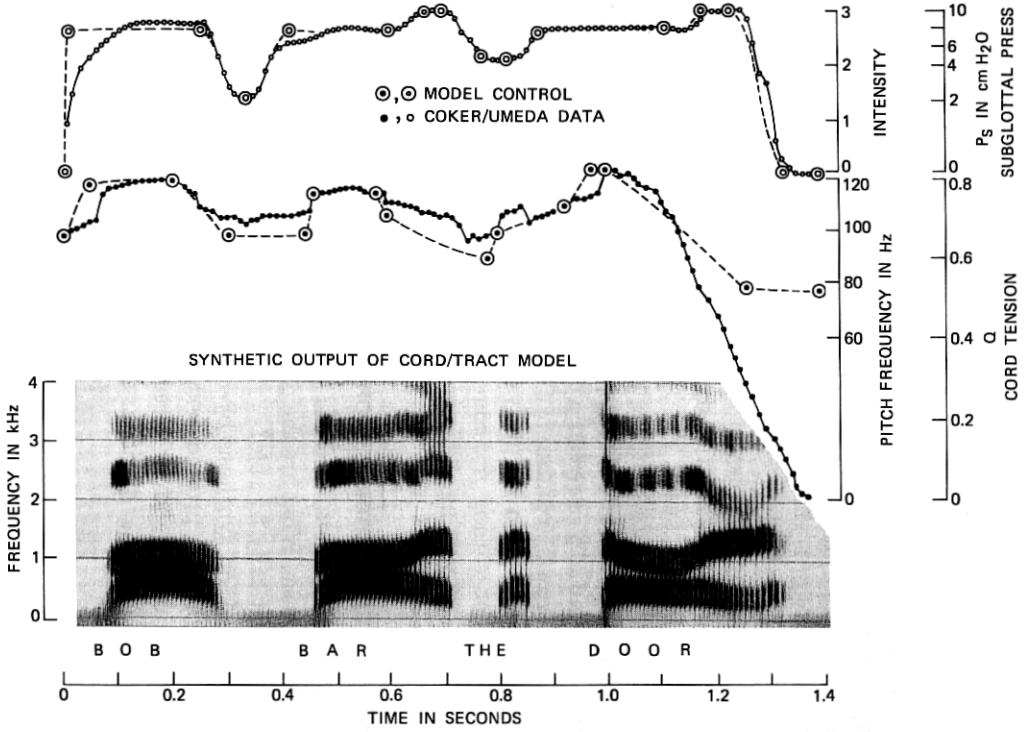

Fig. 12—Example of automatic synthesis of a voiced sentence from printed text using the cord/tract synthesizer.

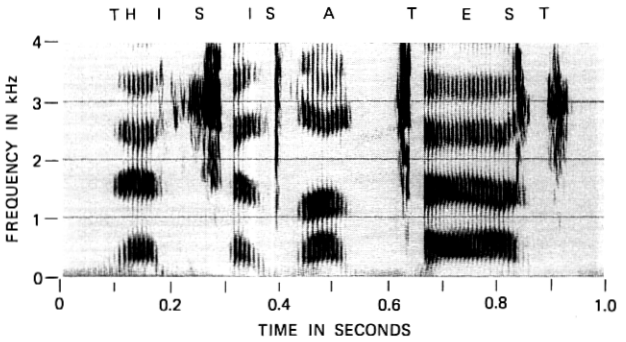

Fig. 13—Example of automatic synthesis from printed text for a sentence containing voiceless consonants.

and a dynamic model of vocal-tract shape. The text synthesis program calculates several functions that can be transformed into the parameters needed for the control of our cord/tract synthesizer. The sequence of conversions is illustrated in Fig. 11. As determined from the Coker-Umeda machine, overall sound intensity can be related to P_s, voice pitch to Q and P_s, voiced-unvoiced switching to A_{go}, and tract shape to N and $A(x)$.

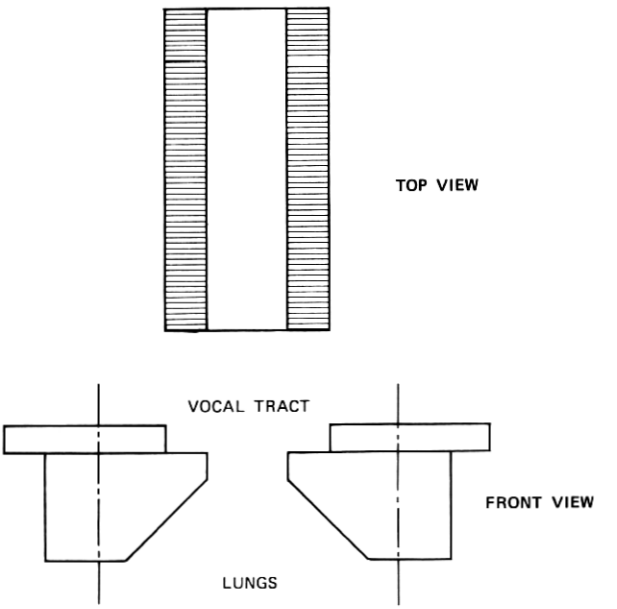

Fig. 14—Format of the computer movie illustrating dynamic behavior of the vocal-cord model.

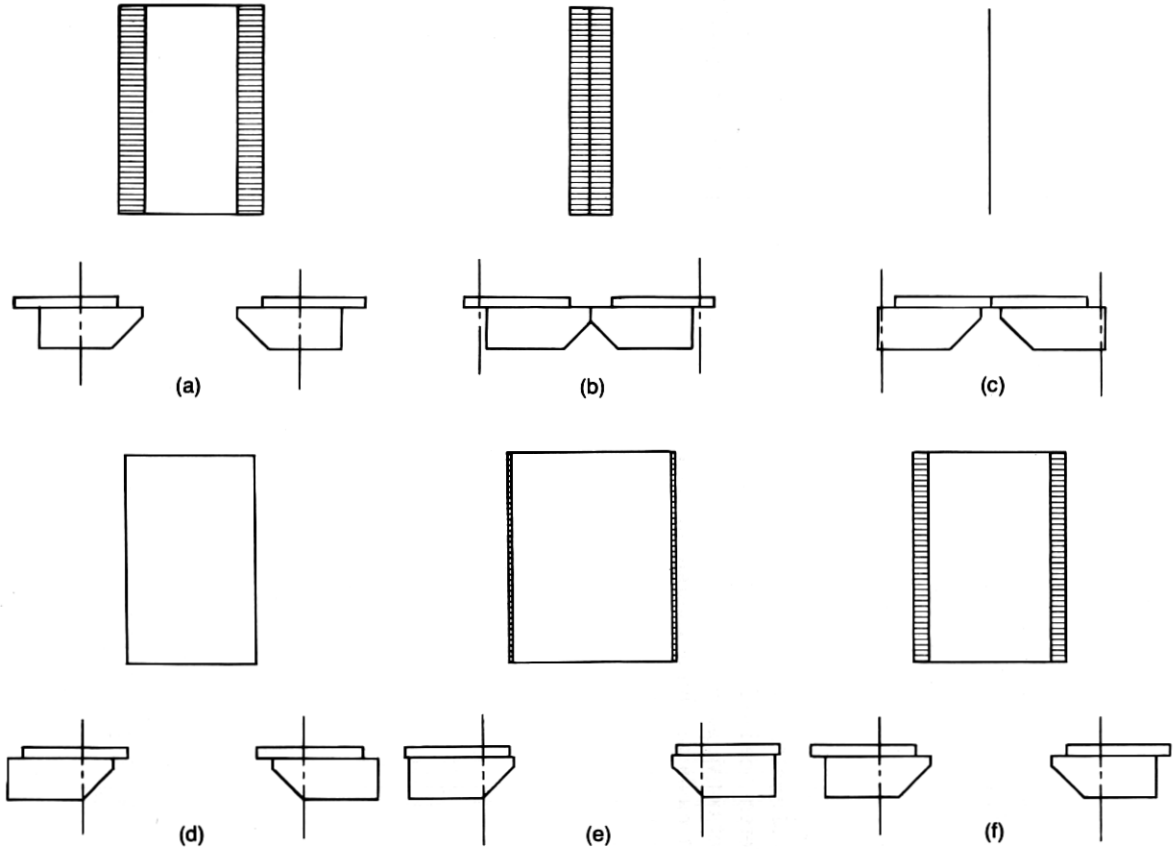

Fig. 15—Frame sequence from the computer movie illustrating vocal-cord vibration.

With the collaboration of Coker and Umeda, we have made an initial trial at synthesis of connected speech by making a transformation of the prosody and area-function output of their text-synthesis machine. An illustration of this first attempt to marry the two systems is shown in Figs. 12 and 13. Figure 12 includes plots to show how the machine-determined values of voice pitch frequency and intensity are transformed into the Q and P_s parameters required by our synthesizer. The spectrogram of Fig. 11 shows completely automatic synthesis of a voiced sentence. Figure 12 shows automatic synthesis of a sentence containing voiceless sounds.

VI. SLOW-MOTION COMPUTER PICTURES OF CORD AND TRACT BEHAVIOR

To aid in visually assessing the complex control and interaction of the model components, we programmed high-speed microfilm displays of the cord and tract motion. The 16-mm movie film, when shown at 24 frames/s, corresponds to a 100:1 slowdown of real time. One can, therefore, examine detailed cord motion and cord/tract interactions.

One display shows details of the two-mass vocal-cord model under dynamic control. The film format is given in Fig. 14 and shows simultaneously a top view of the glottal opening and a front (anterior-posterior) view of the two-mass cord model. Some prints of frame sequences are given in Fig. 15. The time between displayed frames is 20 ms.

A second display, given in Fig. 16, shows a schematized side view of the whole vocal system. The vocal tract is simplified to four cyl-

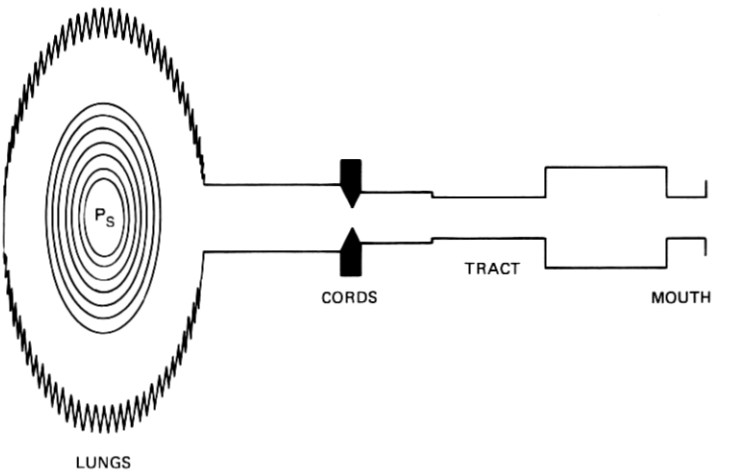

Fig. 16—Format of the computer movie showing dynamic articulatory relations between lung pressure, cord motion, and tract shape.

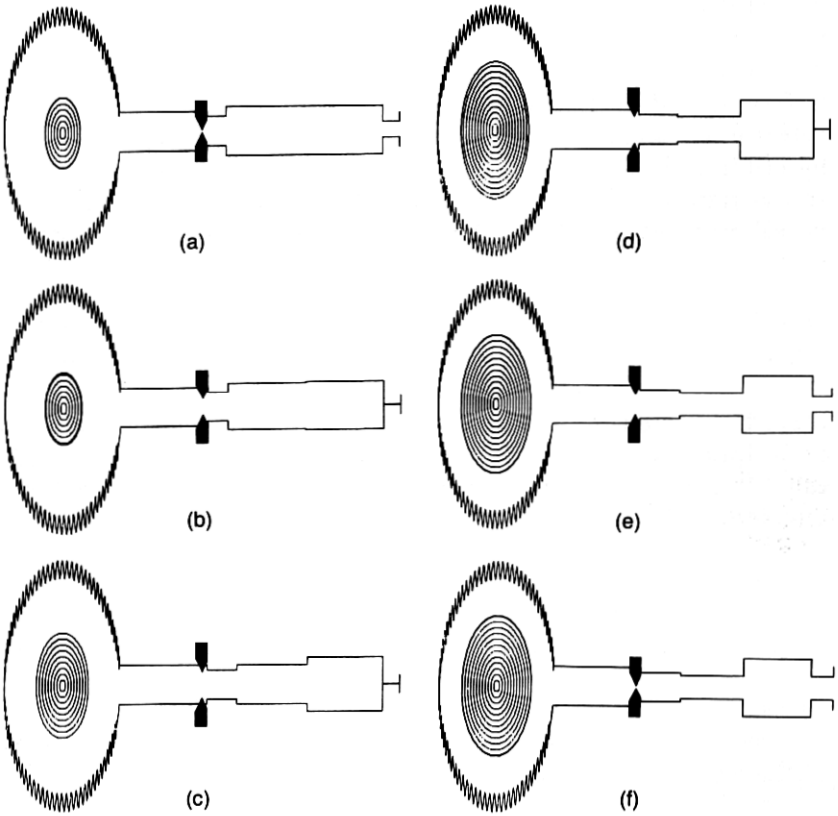

Fig. 17—Frame sequence from the computer movie illustrating dynamic behavior of the cord/tract synthesizer. The sequence is taken from the synthesis of /ə'pa/.

indrical sections only, but with lengths and areas that change with time. The magnitude of subglottal pressure is represented by the elliptical contours in the lung volume which expand or contract with time. Figure 17 shows a sequence of motion frames, spaced by 20 ms, for generation of the syllable /ə'pa/.*

VII. CONCLUSION

Initial experiments with this formulation of cord and tract properties suggest that the physiologically-based control functions have distinct advantages in terms of "built-in" information. That is, much natural behavior—such as vagaries of voicing onset and offset, fine-structure pitch fluctuations occasioned by tract motion, and voicing behavior dur-

* The data of Fig. 17 correspond to those of Fig. 8.

ing occlusion—is produced automatically in the model. In other words, faithful modeling of significant physiological parameters leads to input control data that can be rather parsimonious. It is therefore not necessary to describe input commands with intricate, high-information-rate detail. The model is able to generate many of these intricacies of natural behavior from relatively simple input control.

If continued work proves the cord/tract formulation to indeed possess the desired physiological constraints and attributes, the synthesis approach would also seem promising as a relatively sophisticated end-organ synthesizer which could be driven by models of prosody and articulation, such as provided by the Coker-Umeda text-synthesis system. This is an ultimate long-range goal.

Further than this, however, the model promises some extensive potential for studying the dynamics of real speech. Feasibility is presently being examined for automatically adapting the model's synthetic output to match real speech waveforms (for example, in a least-squares sense). Gradient-climbing adaptive algorithms are being examined for this.[21] Obvious difficulties are model nonlinearities and multiple local-minima traps which may be encountered. Continued work will determine whether these analytical questions can be solved.

Finally, since the present cord/tract synthesis model incorporates the technique for automatic generation of turbulence devised earlier,[4] this feature permits detailed study of the remarkably delicate articulatory timing the human employs in transitions between voiced and voiceless sounds. The cord/tract model therefore fills a critical need for a framework within which to organize and assess articulatory measurements now being accomplished.*

REFERENCES

1. K. Ishizaka and J. L. Flanagan, "Synthesis of Voiced Sounds from a Two-Mass Model of the Vocal Cords," B.S.T.J., *50*, No. 6 (July-August 1972), pp. 1233–1268.
2. J. L. Flanagan and L. L. Landgraf, "Self-Oscillating Source for Vocal-Tract Synthesizers," IEEE Trans. Audio and Electroacoustics, *AU-16* (March 1968), pp. 57–64.
3. K. Ishizaka and M. Matsudaira, "What Makes the Vocal Cords Vibrate," 6th Int. Congr. Acoust., *II* (August 1968), pp. B, 9–12.
4. J. L. Flanagan and L. Cherry, "Excitation of Vocal-Tract Synthesizers," J. Acoust. Soc. Amer., *45*, No. 3 (March 1969), pp. 764–769.
5. J. L. Flanagan, *Speech Analysis, Synthesis and Perception*, Second Edition, New York: Springer Verlag, 1972.
6. S. N. Rshevkin, *A Course of Lectures on the Theory of Sound*, New York: Macmillan, 1963, pp. 400–405.
7. K. Ishizaka, J. C. French, and J. L. Flanagan, "Direct Determination of Vocal-Tract Wall Impedance," J. Acoust. Soc. Amer., *55* (April 1974), p. S79(A).

* In a cooperative study of dynamics of articulation with Dr. T. Shipp, Veterans Hospital, San Francisco, California.

8. J. W. Van der Berg, "Transmission of the Vocal Cavities," J. Acoust. Soc. Amer., *27*, No. 1 (January 1955), pp. 161–168.
9. B. P. Bogert, "On the Band Width of Vowel Formats," J. Acoust. Soc. Amer., *25*, No. 7 (July 1953), pp. 791–792.
10. O. Fujimura and J. Lindqvist, "Sweep-Tone Measurements of Vocal-Tract Characteristics," J. Acoust. Soc. Amer., *49*, No. 2 (February 1971), pp. 541–558.
11. H. K. Dunn, "Methods of Measuring Vowel Formant Bandwidths," J. Acoust. Soc. Amer., *33*, No. 12 (December 1961), pp. 1737–1746.
12. M. M. Sondhi, "A Model for Wave Propagation in a Lossy Vocal Tract," J. Acoust. Soc. Am., *55*, No. 5 (May 1974), pp. 1070–1075.
13. T. Shipp and R. E. McGlone, "Laryngeal Dynamics Associated with Vocal Frequency Change," J. Speech and Hearing Res., *14* (December 1971), pp. 761–768.
14. M. Sawashima, "Movements of the Larynx in Articulation of Japanese Consonants," Annual Bulletin, Research Institute of Logopedics and Phoniatrics, Univ. Tokyo, No. 3 (1968), pp. 11–20.
15. M. M. Sondhi, "Measurement of the Glottal Pulses," 86th Meeting, Acoust. Soc. Amer., *53* (October 1973), G-1.
16. J. S. Perkell, *Physiology of Speech Production: Results and Implications of a Quantitative Cineradiographic Study*, Boston: M.I.T. Press, 1969.
17. J. French, C. H. Coker, M. M. Sondhi, K. Ishizaka, and J. L. Flanagan, "Measurement of Articulatory Dynamics," Proc. Int. Congr. Acoustics, London (July 1974).
18. C. H. Coker, unpublished laboratory data on transillumination.
19. J. Lindqvist, "Laryngeal Articulation Studied on Swedish Subjects," STL-QPSR, 2–3 (1972), pp. 10–27.
20. C. H. Coker, N. Umeda, and C. P. Browman, "Automatic Synthesis from Ordinary English Text," IEEE Trans. Audio and Electroacoustics, *AU-21*, No. 3 (June 1973), pp. 293–298.
21. E. Hafer, "Speech Analysis by Articulatory Synthesis," M.S. Thesis, Dept. Elec. Eng., Northwestern Univ., June 1974.