SOME INFLUENCES OF THE GLOTTAL WAVE UPON VOWEL QUALITY

JAMES L. FLANAGAN

Of the varied articulations of speech sounds, vowel production is the simplest to describe acoustically. This is owing mainly to the fact that the source of excitation and the transmission properties of the vocal tract can be analyzed relatively independently. The acoustic output can consequently be determined from the component properties of the source and the transmission system. The physical reasons which permit such analysis are primarily that the source of excitation is spatially fixed, namely, at the vocal cords, and the source and transmission system do not greatly interact. At most frequencies, the acoustic impedance of the glottal source is appreciably higher than the driving point impedance of the tract, so that changes in vocal configuration do not greatly affect the operation of the voice source.

The shape of the vocal tract determines which vowel, of the ensemble of vowels belonging to a given language, is to be produced. The intensity, shape and periodicity of the glottal waveform largely determine the quality and personal characteristics associated with the vowel. The purpose of this paper is to consider, and to demonstrate, some of the ways in which the glottal wave can influence vowel quality. Toward this end, we will first recall the basic transmission properties of the vocal tract. Then we will consider certain characteristics of the glottal wave and how they might influence the ultimate output of the vocal system.

To recall the frame of reference, Figure 1 shows a sagittal-plane X-ray of an adult male vocal tract. The tract is in the approximate position for producing a high front vowel. Excitation occurs only at the glottis, by means of vocal cord vibration, and sound radiation takes place principally at the mouth. The velum, which normally seals off the nasal tract during the production of non-nasal sounds, is shown here slightly open, because phonation had not begun at this instant.

The length of the vocal tract is comparable to the wavelengths of frequencies in most of the audible spectrum. The tract cannot therefore be analyzed strictly as a lumped constant circuit but be must treated as a distributed system, or a non-uniform transmission line. Even so, one usually identifies two or more major cavities and constrictions in vowel production. To a rough approximation the tract is similar to the series of cylindrical pipes shown in Figure 2.

Because the glottal opening between the vocal cords is small compared with the cross-sectional area of the vocal tract, the acoustic impedance of the voice source is

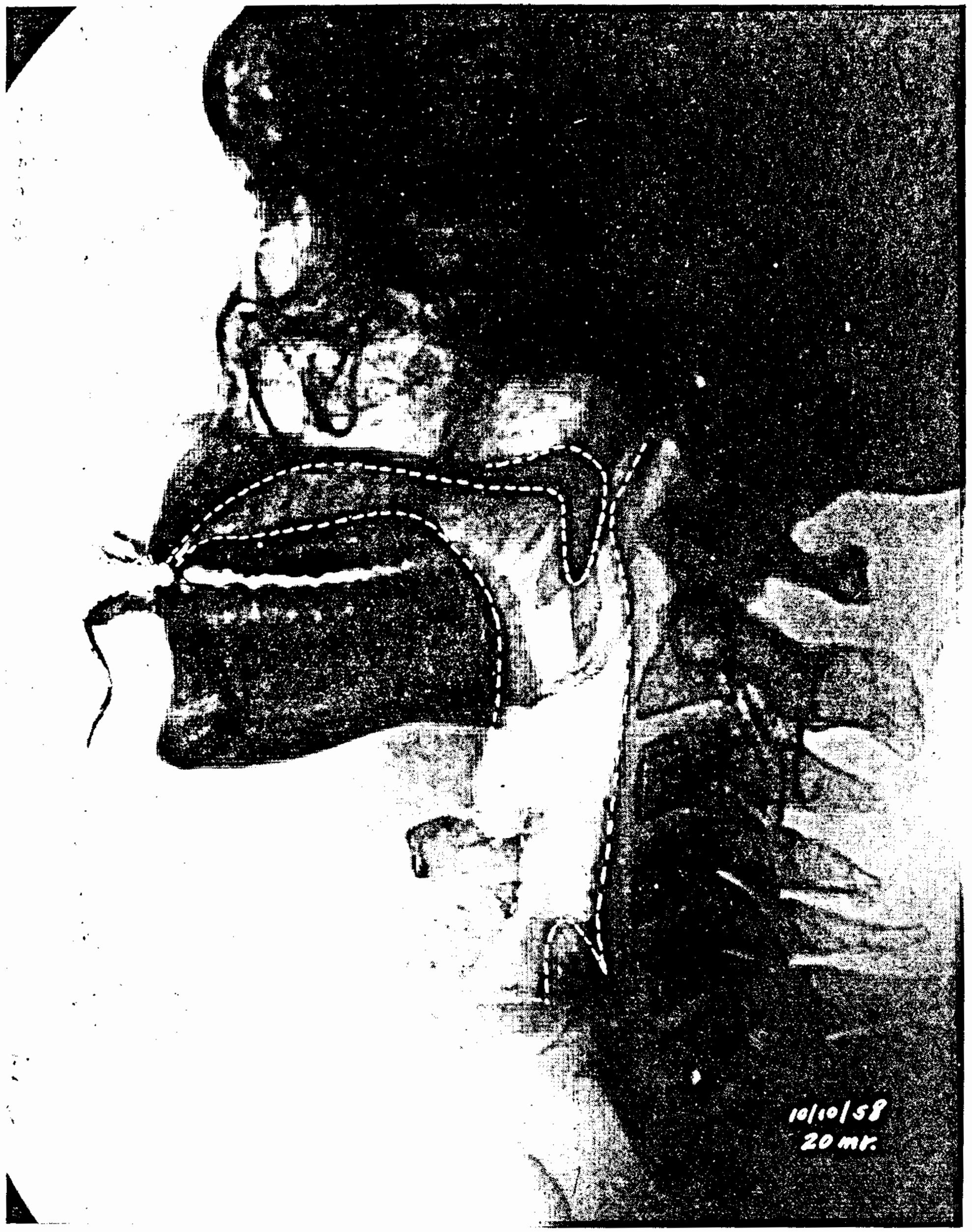

Fig. 1. Sagittal-plane X-ray of an adult male vocal tract.

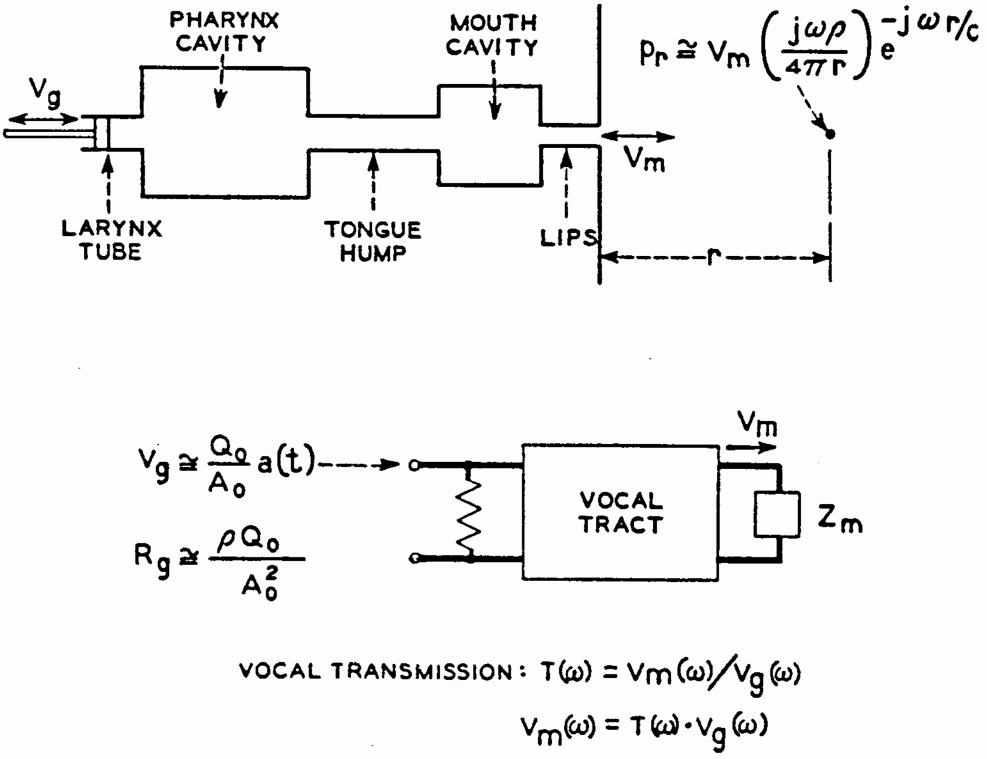
PHARYNX CAVITY
MOUTH CAVITY
V_g
LARYNX TUBE
TONGUE HUMP
LIPS
V_m
r
$P_r \cong V_m \left(\frac{j\omega\rho}{4\pi r}\right) e^{-j\omega r/c}$
$V_g \cong \frac{Q_0}{A_0} a(t)$
$R_g \cong \frac{\rho Q_0}{A_0^2}$
VOCAL TRACT
V_m
Z_m
VOCAL TRANSMISSION: $T(\omega) = V_m(\omega)/V_g(\omega)$
$V_m(\omega) = T(\omega) \cdot V_g(\omega)$

Fig. 2. Schematic representation of the vocal tract.

generally high compared with the input impedance of the tract. The glottal volume velocity is therefore dependent primarily upon glottal area and subglottic pressure, and is little affected by tract configuration. For this reason, the glottis can be approximated as a source of constant volume velocity. In the upper left part of the diagram the glottal source is represented as a piston producing the volume velocity V_g into a pipe the size of the larynx tube. The pharynx cavity is represented by a pipe of large diameter, and the tongue-hump constriction, the mouth cavity and the lip tube are indicated by pipes of appropriates size. The volume velocity produced at the mouth is denoted as V_m. In the frequency domain, it is related to the sound pressure at a fixed point, r distance in front of the mouth, by approximately the spherical source relation shown at the right of the slide.

The vocal system can be represented by a block diagram in the fashion shown in the lower part of the figure. A more detailed consideration of the glottal source[1] shows that it can be approximated as a volume current whose time waveform is the quantity $[Q_0/A_0]$ a(t), where Q_0 is the mean volume flow, A_0 the mean glottal area,

[1] J. L. Flanagan, "Some properties of the glottal sound source," *Jour. Speech and Hearing Res.*, 1, 99–116 (1958).

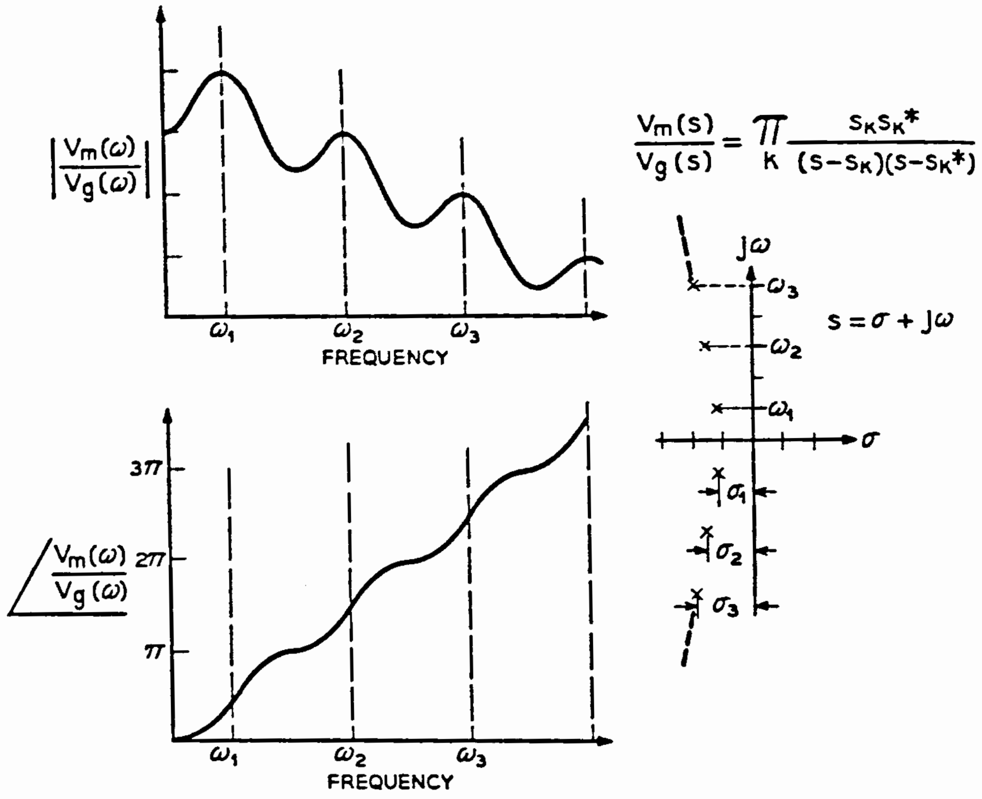

Fig. 3. Frequency characteristics of vocal tract transmission.

and a(t) the a.c. component of the glottal area wave. Further its inherent resistance is about $\rho Q_o/A_o^2$ where ρ is the air density. The output mouth volume velocity, V_m, is, in effect, driven through the radiation impedance at the mouth, Z_m. The power dissipated in the real part of the radiation impedance is the sound power radiated.

The vocal transmission can be defined in terms of the ratio of the (Fourier) frequency transforms for the mouth and glottal volume currents. This is indicated by the function T(ω) at the bottom of the figure. This function is of course determined by the vocal tract shape and it is a frequency domain description of the articulatory configuration. The product of T(ω) and the glottal function $V_g(\omega)$ specifies the mouth output. We want to look at the frequency properties of the individual functions T(ω) and $V_g(\omega)$ in a little more detail.

The nature of the vocal transmission, T(ω), is well known to speech researchers.[2] Its frequency dependence is indicated qualitatively in Figure 3. The amplitude-vs-

[2] G. Fant, *Acoustic Theory of Speech Production* (Mouton and Co., 's-Gravenhage, 1960).

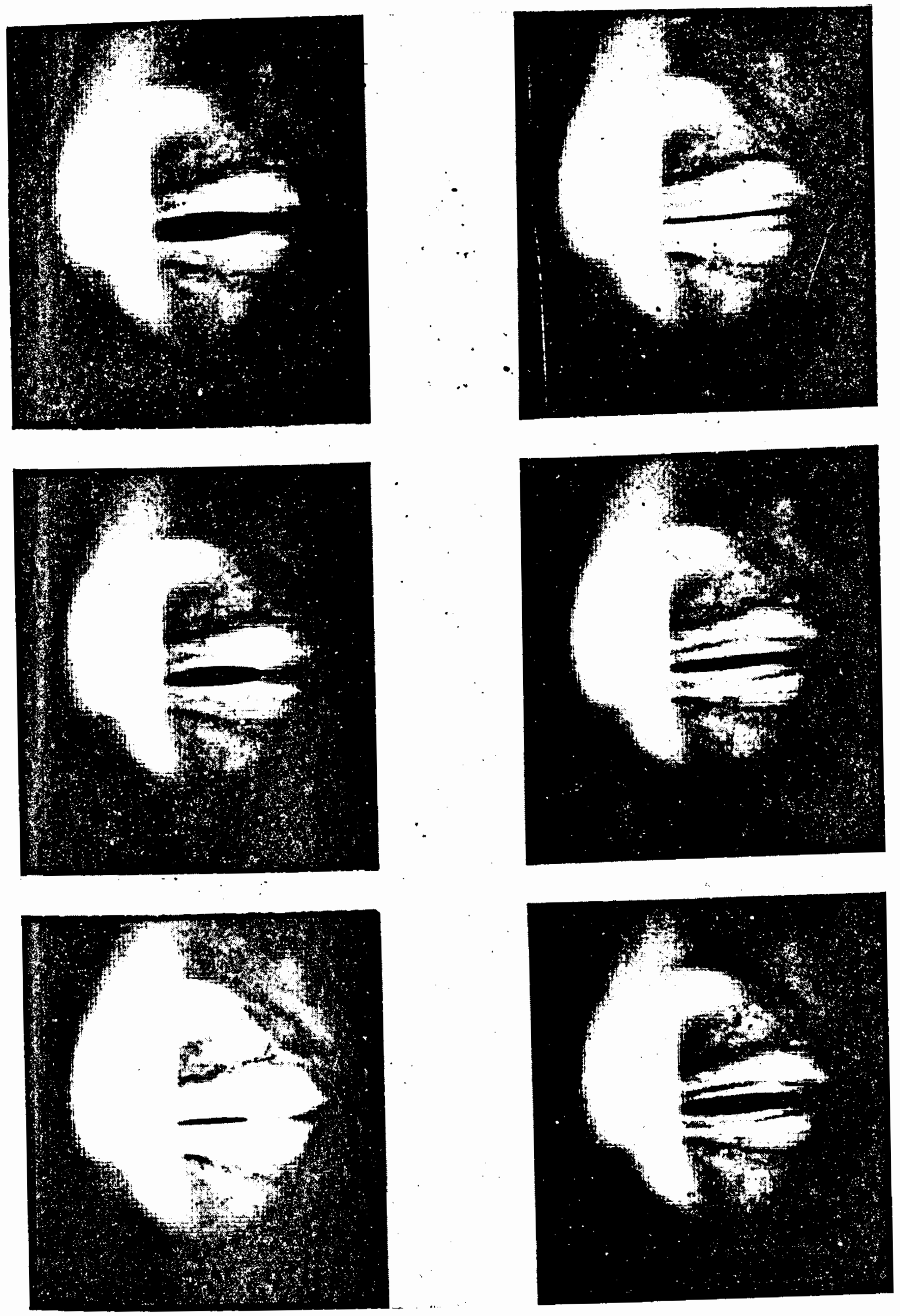

Fig. 4. Selected phases in one period of vibration of the vocal cords.

frequency response of the vocal tract is shown at the upper left. It is characterized by peaks or maxima spaced along the frequency axis. These peaks, or formants, are the natural resonances, or normal modes of vibration, of the vocal tract. Their frequency spacing is dependent upon the tract configuration, and their bandwidth or damping generally increases with formant number. For our present purposes, one of the important things to notice is that the amplitude response never dips very precipitously, and never drops to zero. The valleys between formant peaks are relatively shallow.

The phase response of the vocal transmission is illustrated at the lower left. The phase angle advances in lag by π radians each time a formant peak is passed on the frequency axis. The slope of the phase curve is representative of the propagation time from the glottis to the lips, and is approximately the tract length divided by the sound velocity.

For those partial to Laplace transform techniques, the complex frequency representation of the vocal transmission places the normal modes in evidence as shown at the right of the figure. The vocal transmission can be characterized by a distribution of simple, complex conjugate poles. These are the s_k's in the product function, and are shown by the X's in the s-plane diagram. The frequency responses at the left are obtained from $V_m(s)/V_g(s)$ by letting $(s = j\omega)$ and taking the magnitude for the amplitude response, and the angle for the phase response. As is commonly known, the first three normal modes, or formants, usually lie in the frequency range below 3000 cps for an adult male. The damping or resonant bandwidth for a given formant, is approximately constant and is determined principally by viscous, heat-conduction, glottal and radiation losses in the tract.

As indicated in Figure 2, the amplitude spectrum of the vocal output is the product (or db sum) of the vocal transmission spectrum and the spectrum of the glottal source. If the glottal source produced sharp repetitive pulses of volume flow, its spectrum would be a line spectrum whose envelope would be relatively flat or uniform, and whose lines would be spaced at the pitch frequency. In such a case the spectral envelope of the vowel output would be essentially that of the vocal transmission, such as illustrated in Figure 3. The glottal source does not, of course, produce sharp pulses of volume velocity. Its spectral envelope is generally irregular rather than flat. Because of this, it influences the shape of the ultimate vowel spectrum. We would like to consider a little more particularly what these effects might be.

We mentioned earlier that the glottal volume flow is similar in waveform to the area of the vocal cord opening. The nature of the glottal opening is very familiar to phoneticians, but for the sake of the completeness it is illustrated in Figure 4. These photographs are successive phases in one cycle of vibration of an adult male's vocal cords. The pitch is about 125 cps and the elapsed time for the cycle is about 8 msec. For a constant subglottic pressure, the acoustic volume velocity is approximately proportional to the area of the glottal opening viewed here. Actually, a more accurate estimate of the volume velocity can be had from a knowledge of the sub-

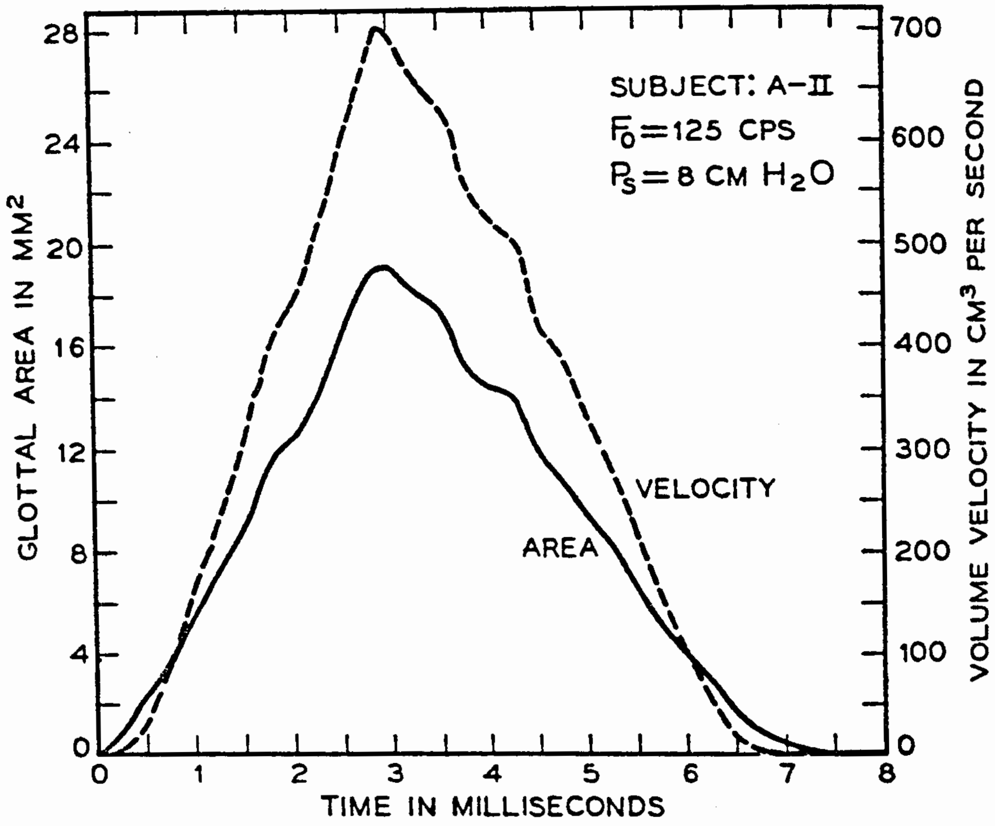

Fig. 5. *Solid curve*: glottal area wave derived from high-speed motion pictures of the vocal cords. *Dashed curve*: calculated volume velocity wave.

glottic pressure and glottal area (see ref. 1), using results which have been determined by van den Berg[3] and his associates, and by Wegel.[4]

By way of illustration, a typical area wave obtained from high speed photography, and its calculated velocity wave are shown in Figure 5. These data are for an adult male uttering the vowel /a/, at a pitch of 125 cps and at a relatively loud level. The leading and trailing edges of the velocity wave are slightly steeper than the area function because of the nonlinear resistance of the glottal orifice. This generally has the effect of making the high frequency components of the velocity spectrum slightly more intense than those of the corresponding area wave. The glottal wave is, of course, susceptible to very large variations in shape and period, but this one is typical enough to illustrate a point about its influence on the vowel spectrum.

We can study the spectrum of the glottal wave from a Fourier analysis of a single

[3] Jw. van den Berg, J. T. Zantema and P. Doornenbal, Jr., "On the air resistance and the Bernoulli effect of the human larynx," *J. Acoust. Soc. Am.*, 29, 626–631 (1957).

[4] R. L. Wegel, "Theory of vibration of the larynx", *Bell Systm. Tech. J.*, 9, 207–227 (1930).

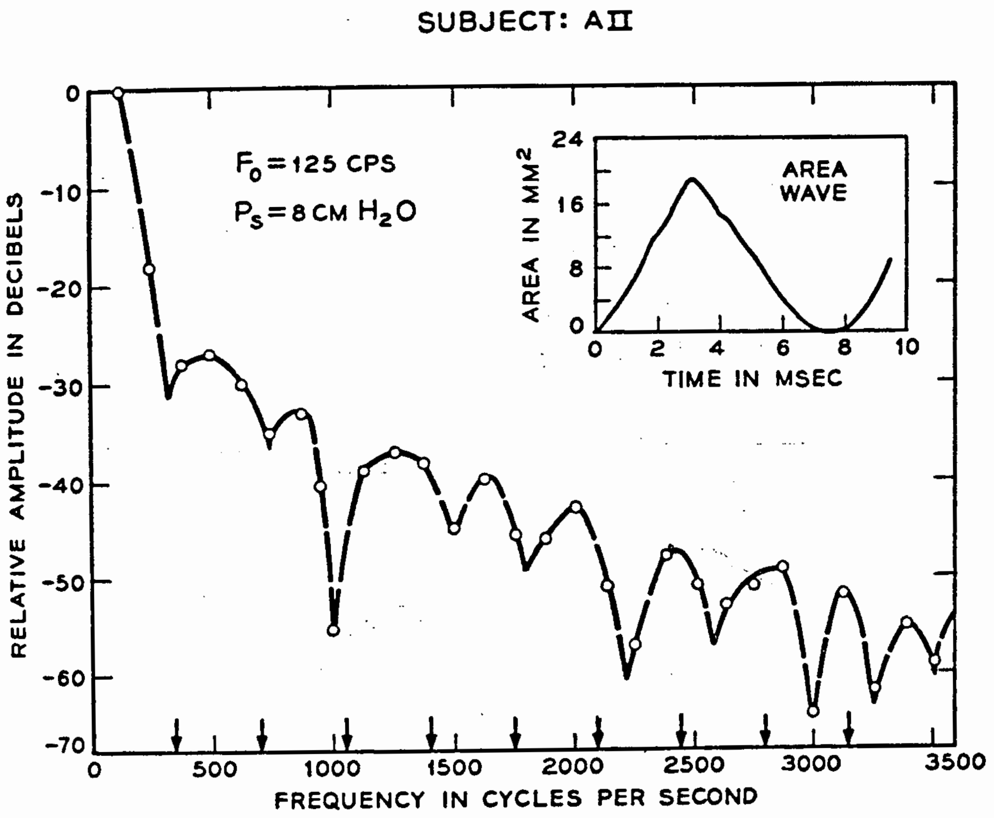

Fig. 6. Amplitude spectrum computed for the area wave of Fig. 5. Each point indicates the amplitude of a harmonic line.

period. This is easily calculated on a digital computer. Suppose we consider the area wave. Its computed spectrum is shown in Figure 6. Each plotted point is the amplitude of a harmonic line. One sees that the envelope of this line spectrum displays many irregularities, with a number of dips or holes in the amplitude response. These dips are occasioned by the so-called zeros of the glottal wave. The zeros occur at the specific values of frequency which make the frequency-transform of the glottal wave equal to zero. Except in particular cases, the zeros usually lie at complex frequencies, and the dips in the Fourier spectrum, as we have here, do not go completely to zero for real frequencies.

The main point to mention here, however, is that it is easily possible for a minimum of the glottal spectrum to coincide with a peak, or formant, of the vocal transmission, and for the formant to be thereby de-emphasized or even nullified in the output. The positions of the dips in the glottal spectrum are functions only of the glottal wave shape and depend upon the pitch only to the extent that the waveshape depends upon pitch. The question might be posed as to how significant are the effects of the

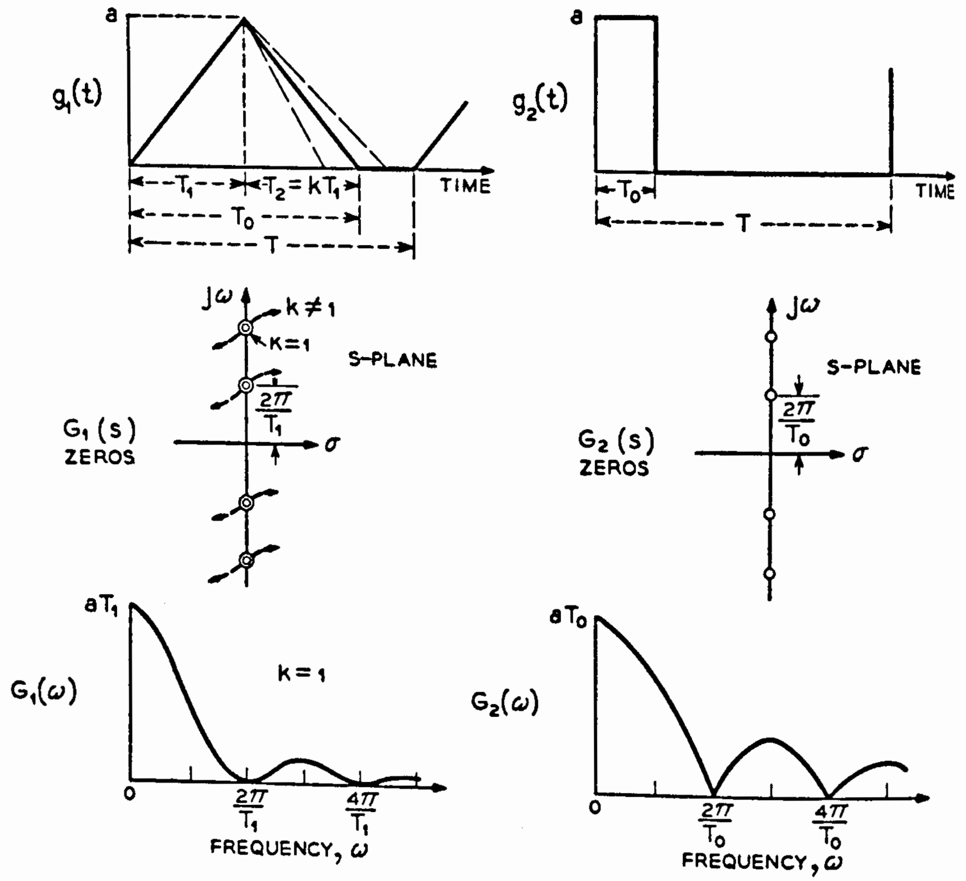

Fig. 7. Time and frequency relations for triangular and rectangular waveforms.

glottal zeros in speech. Certainly speech is a dynamic event, and one does not produce the same glottal waveform for all voiced sounds. There are, however, sustained voiced passages in speech during which the glottal zeros can affect vowel quality.

When one studies high speech motion pictures of vocal fold vibration, one finds that the area wave can range all the way from relatively sharp, rectangular forms to almost sinusoidal forms, depending upon vocal effort, intensity, pitch, and other factors. The triangularly shaped wave is a commonly seen one, both in its symmetrical and asymmetrical forms. The spectral zeros of the symmetrical triangular and rectangular waves occur at real frequencies. The waves have particularly simple spectral representations which it is useful to recall. The spectra for these idealized waveforms are shown in Figure 7.

The time waveforms for the functions are shown at the top of the diagram. The triangular wave has its opening phase during time T_1, and its closing phase during $T_2 = kT_1$. If it is symmetrical $T_1 = T_2$, or $k = 1$. The rectangular wave has open

time T_0 and period T. The amplitude spectra for these two waves are at the bottom of the figure. Both have their spectral zeros at real frequencies. In the case of the symmetrical triangle, the zeros are spaced at frequencies which are integral multiples of $1/T_1$, and the spectrum is a $(\sin^2 x/x^2)$ function. The zeros are double zeros, as is evidenced by the fact that not only does the spectral amplitude go to zero, but its slope also goes to zero. Suppose, for example, a vowel having its first formant at about 500 cps is excited by this wave. The first double zero of the excitation would fall at the formant frequency if $T_0 = 4$ msec. A glottal open time of this value is commonly encountered in speech. For the rectangular wave the zeros are single and are spaced at integral multiples of $1/T_0$. Its spectrum is a $(\sin x/x)$ function and the zeros are single ones.

The zero diagram on the complex frequency place is shown for both functions in the middle diagrams. For the symmetrical cases the zeros fall on the $j\omega$ axes. As the triangle is made asymmetrical, however, that is as *k* is made different from unity, the double zeros part and move off the axis and out into the plane. Their trajectories are dependent upon the wave shape in a nonsimple manner, and it is necessary to solve a transcendental equation to find their locations. This situation actually corresponds to many real-life conditions, and we are currently carrying out some computations for these types of situations on a digital computer. Unlike the vocal tract poles which, by stability criteria, are confined to the left half of the s-plane, the zeros may occupy positions in the right half of the s-plane. The latter is particularly the case for glottal waves whose closing phase is shorter than the opening phase.

When a glottal zero falls close to, or coincides with a pole, or formant of the vocal tract, it can partially or completely nullify the formant. Under such conditions it may completely eradicate a formant peak from the output spectrum. One convenient way to study such effects is with an electrical synthesizer for vowel sounds. An arrangement to do this is illustrated in Figure 8. In the block diagram at the bottom of the figure, the vocal tract transmission and its associated formants are simulated by a set of resonant electrical circuits. These are labelled vowel synthesizer. The formants, or pole positions, are set into the circuits. The synthesizer is then excited by a function generator whose output waveform has a shape and pitch analagous to a given glottal wave. Through this means, the poles of the vocal tract, and the zeros and pitch of the glottal wave can all be varied independently.

Two idealized conditions of glottal excitation, for the same vowel, are represented by the data at the top of the diagram. In both cases the vowel is $|\Lambda|$ and the glottal wave is approximately a symmetrical triangle. The first two formants are represented by the X's in the s-plane diagrams at the right. For this vowel the first two formants occur at $j\omega$ frequencies of about 600 cps and 1200 cps. The double zeros of the glottal wave are also indicated in the same diagrams, and occur at integral multiples of $1/T_1$. In case 'A' the open time of the glottal wave is taken as 4 msec which places its zeros at about 500, 1000, 1500 cps, etc. In the case labeled 'B' the open time of the wave is taken as 2.5 msec which spaces the zeros at 800, 1600, 2400 cps, etc. In case

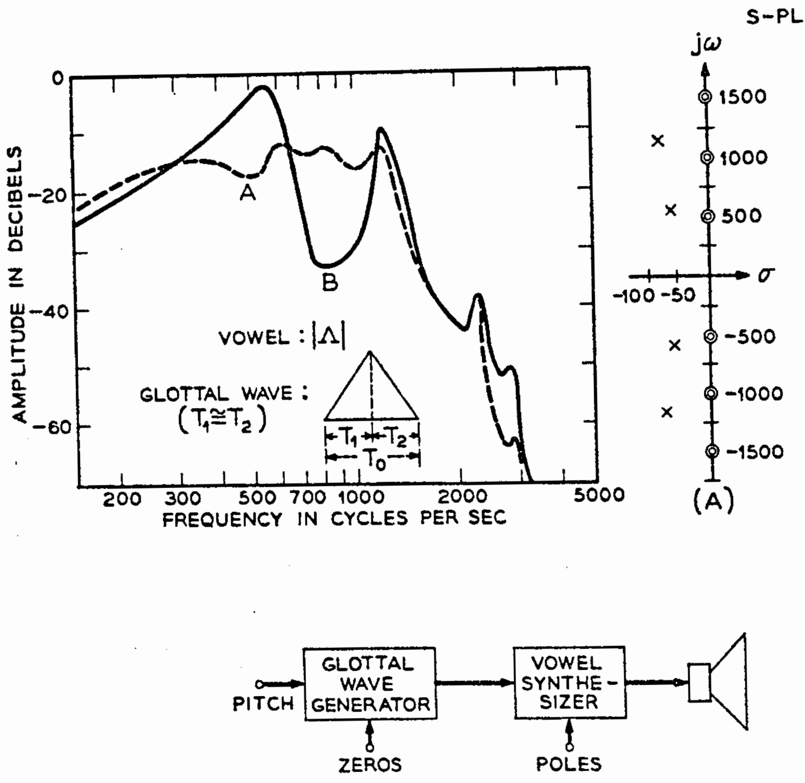

Fig. 8. Effect of glottal zeros upon the measured spectrum of a synthetic vowel sound.

'A' the glottal zeros fall near the first two formants. In case 'B' they more or less straddle the formants.

When these conditions are set up on the synthesizer, and the ouput spectrum actually measured on a wave analyzer, the result is shown at the upper left of the diagram. In this particular situation the effect upon the first formant is the greatest. In cases 'A' the glottal zeros completely level the first formant, whereas in the second case, the glottal zeros greatly emphasize the valley between the first and second formants. In neither case does the measured amplitude spectrum actually go to zero at the frequency of the zeros because the synthesized glottal wave is not precisely symmetrical. Its zeros actually lie off the jω axes of the s-planes by a little bit. We can demonstrate these two conditions from a magnetic tape recording. On the recording, the pitch is given a falling inflection from about 120 to 105 cps, with a small amount of 7 cps vibrato to improve naturalness. This inflection does not influence the zero location.[5] In considering the vowel quality, most persons have no difficulty in hearing a difference.

As another, and perhaps more dramatic situation, let us consider an excitation of the synthesizer such as shown in Figure 9. The top row of oscillograms show three different glottal waves. The second row shows the waveforms of the vowel synthesizer output for the vowel /ə/ when it is excited by the top waves. The lower diagrams are the pole-zero plots for each condition. These glottal waves are not particularly realistic, but they represent extreme cases which are easy to treat analytically. In particular, the second two conditions can be crudely likened unto the phenomenon of diplophonia which has been discussed by Smith[6] and others. This is a situation in which the vocal cords produce two glottal puffs per period, usually one strong one and a weaker one.

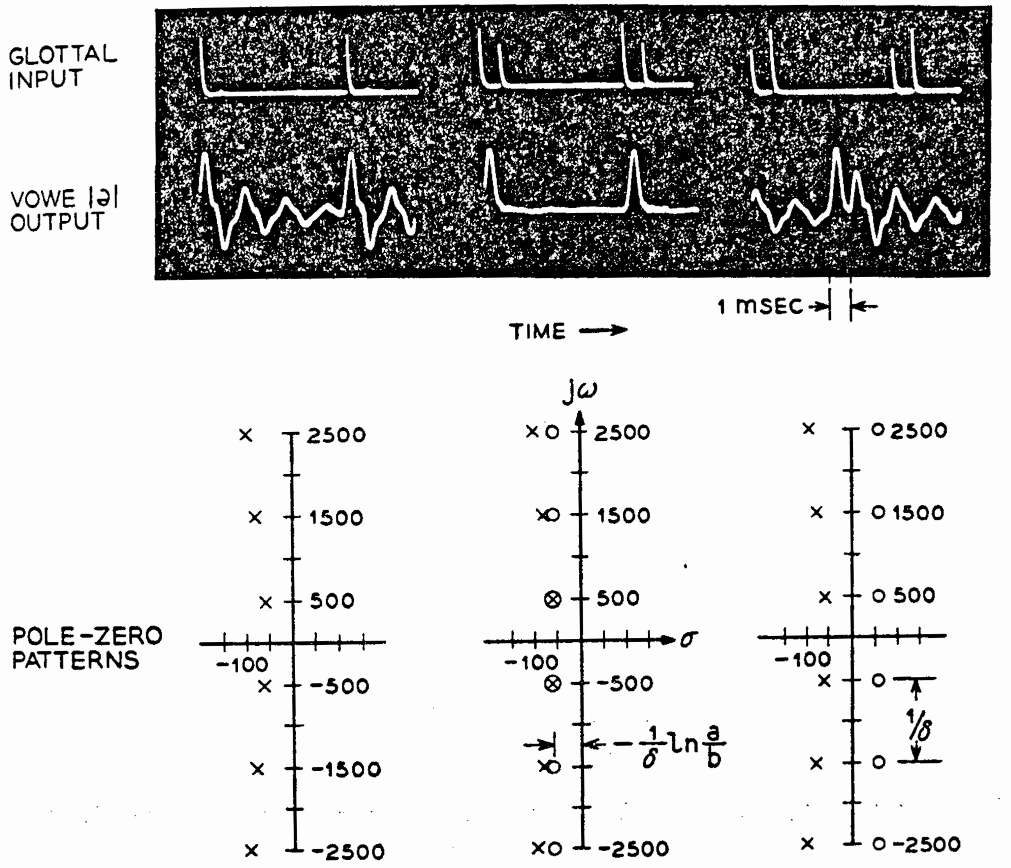

Fig. 9. Nullification of vowel formants by left-half plane (middle column) and right-half plane (right column) glottal zeros. These glottal excitations can be crudely likened unto the phenomenon of diplophonia.

In the first column, the glottal excitation of the synthesizer is very sharp, repeated pulses. This excitation has no spectral zeros in the audible frequency-range, so the pole-zero diagram for the vowel output exhibits only poles, corresponding to the

[5] The first part of a tape recording, demonstrating this condition, was played at this point.

[6] S. Smith, "Diphloponie und Luftschallexplosionen," *Archiv Ohren-usw. Heilk. u Z. Hals-usw. Heilk.*, 173, 504–508 (1958).

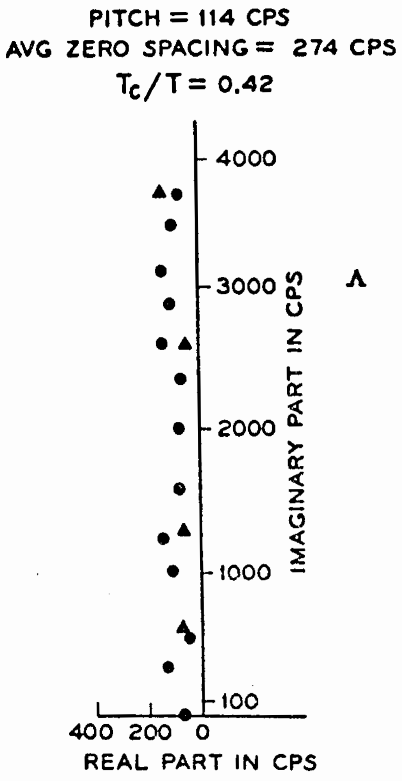

Fig. 10. Pole-zero diagram of a best fit to the spectrum of a natural vowel sound (after Mathews, Miller and David).

formants of the vowel /ə/. In the second case the zeros fall in the left half plane, and are made to nearly coincide with the vowel formants. In this very extreme case, the zeros of the source nearly cancel completely the poles of the tract, so that the overall frequency response of the source and system is a constant. The waveform of the synthesizer output is very nearly repeated impulses. This output has an amplitude spectrum which is very nearly flat and is completely devoid of formant peaks. In the third case, the zeros lie in the right half of the s-plane, just opposite the poles. The output vowel waveform exhibits no particularly peculiar features, but like the second case its amplitude spectrum is very nearly flat. Its phase spectrum of course differs considerably from that of the second case. All of these conditions are recorded on the magnetic tape.[7]

[7] The tape recording demonstrated some paired comparisons for the vowel /ə/ as well as for the vowels /ɜ, u, æ, ɑ/. The first sample was the given vowel generated with the single-pulse excitation shown in the left column, followed by the same vowel generated with the pole – cancelling excitation shown in the middle column. – The second sequence was designed to demonstrate the small perceptual difference between the excitations shown in the second and third columns for which only the phase spectra differ. The samples were: first the vowel generated by the excitation of column one; then column two; then column one again; then column three; then column two; and finally column three.

The perceptual differences between excitation 1–2 and 1–3 are quite pronounced and easily heard; the differences between excitations 2–3 are very small.

As one final point, how closely are these idealized, synthetic situations related to real speech? In some recent work at the Bell Laboratories, Mathews, Miller and David[8] used a computer to fit pole-zero patterns to the spectra of real speech. Figure 10 shows the best fit they obtained for one particular vowel utterance. The X's represent the poles, and the dots represent the zeros. Although the analysis method does not decide whether the zeros lie in the left or right half planes, it is clear that there are a substantial number of them present in the spectrum. We feel, therefore, that a synthesis technique such as described here may prove helpful in studying the subjective effects of glottal zeros, and that some of the idealizations made here may not be particularly unrealistic.

Bell Telephone Laboratories, Inc.
Murray Hill, N.J.

[8] M. V. Mathews, J. E. Miller and E. E. David, Jr., "Pitch synchronous ânalysis of voiced sounds," *J. Acoust. Soc. Am.*, 33, 179–186 (1961).