Dynamical Study of the Vowel Sounds
Part II

By IRVING B. CRANDALL

SYNOPSIS: Comparative studies based on oscillographic records of the principal characteristics of vowel, semi-vowel, and consonant sounds, have contributed much to an understanding of the mechanism of speech. Analyses of the frequency spectra of vowels show almost invariably two principal resonance peaks which fact is suggestive of a double resonator to produce them.

The present paper is concerned with the mechanism of the double resonator system and a mathematical treatment thereof. Based on the volume, shape and coupling of the resonating chambers, some models of cardboard, tube and plasticene were made, and with which some experimental tests in the production of vowels were carried out. The best success was had with the sound *ä* (father) while fair results were obtained with the sound *ō*, *ă* and *ĕ*.

INTRODUCTION

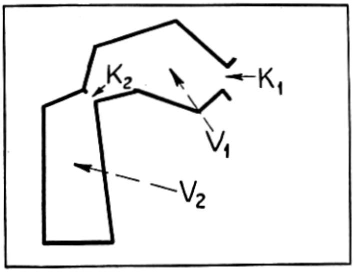
IN two earlier papers [1] a diagram has been given of the frequency spectra of the vowel sounds, based on analyses of a large number of accurate oscillographic records. In addition, there was given, in the second of these papers, a comparative study of the principal characteristics of vowel, semi-vowel and consonant sounds, and an account of certain studies made by other investigators whose methods and results have contributed to our understanding of the mechanism of speech.

Among the more original of recent contributions are those of Sir Richard Paget,[2] who has successfully employed multiple resonators to simulate almost all the vowel and consonant sounds. In getting together the material for the second paper from the Bell Laboratories, Paget's results for the *vowel sounds* were compared with ours only in a general way, and not in so detailed a manner as was followed in the discussion of consonant and semi-vowel sounds in that paper. It may be permissible to return to a consideration of the vowel sounds in the present paper, following Sir Richard Paget's idea of the double resonator as the instrument for vowel production. Indeed Sir Richard has pointed out to us that, since our own data on the spectra of the vowel sounds show almost invariably two principal resonance peaks, there must be a double resonator to produce them, thus harmonizing our results, at least for the male voices, with his own.

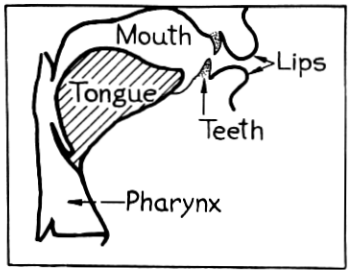
[1] I. B. Crandall and C. F. Sacia, "Dynamical Study of the Vowel Sounds," also "The Sounds of Speech," BELL SYSTEM TECHNICAL JOURNAL, III, 1924, p. 232; bid., IV, 1925, p. 586.

[2] *Proc. Roy. Soc.*, A102, 1923, p. 752; ibid., A106, 1924, p. 150.

Fig. 1*a* is a diagram of a double resonator. The volumes of the chambers are respectively V_1 and V_2; the conductivities [3] of the orifices are K_1 and K_2. In this structure the outer orifice corresponds to the mouth (see Fig. 1*b*), the outer cavity to the buccal cavity, the



1*a*



1*b*

Fig. 1*a* and 1*b*—Diagram of the mouth-pharynx system

inner orifice to the constriction between the soft palate and the back of the tongue, and the inner cavity to the pharynx. The source of sound in the back of the inner chamber is of course the glottis, or rather the periodic puffs of air to which the glottis gives rise, and we may remark that at resonance the apparatus is *driven at a node* (or pressure maximum) which is a condition for maximum efficiency. In Paget's models, a small opening was made at the back for the source of sound, which was a loosely stretched strip of rubber, mounted in a slit, and blown by an air stream. To be successful, in connection with the resonator model, in producing a vowel sound artificially, such a source must of course generate a sound whose fundamental is somewhere near that of the human larynx, and which has in addition a very extended range of harmonics; that is, for a bass voice, we should need a fundamental frequency of about 100, and additional sound energy scattered through the frequency range up to 4,000 or 5,000 cycles.[4]

[3] The average mass of air which surges to and fro in the orifice of a resonator is $\rho S^2/K$, in which ρ is the density of air, S the area of the orifice, and K the conductivity. K is a linear quantity, proportional to the width of the orifice, and is a measure of the ease of flow of fluid through it. It may be defined as the ratio of the (velocity) potential difference, between the two ends of the orifice, and the flux or current $(S\dot{\xi})$ flowing through the orifice.

[4] Sacia suggests that a source of sound giving a saw-toothed wave (rip saw tooth: one slanting and one vertical side) should be ideal for driving vowel resonators. (An experiment with such a device will be described later.) This wave shape corresponds to a fundamental and full retinue of harmonic tones, and should be of service in many ways in acoustic experiments.

Physical Features of the Mouth-Pharynx System

It is a curious fact that most of our data on the shape of the mouth cavities, position of the tongue, etc., for producing the different vowel sounds have been obtained by students of phonetics. There are of course excellent drawings of the mouth structure, in a few typical positions, given in the literature of anatomy; but for the finer differences, from one vowel sound to the next, we must rely on other sources. I know of no determination, for example, of the actual volumes of the mouth and pharynx, in any position for a typical individual, nor have I succeeded, by consulting anatomical experts, in obtaining the desired data.

In Fig. 2, there are shown certain conventional drawings, in median section, of the human mouth-pharynx region. These are taken from Rippmann's "Elements of Phonetics" (London, Dent, 1914) and were taken in turn by Rippmann from an article by Dr. R. J. Lloyd. In

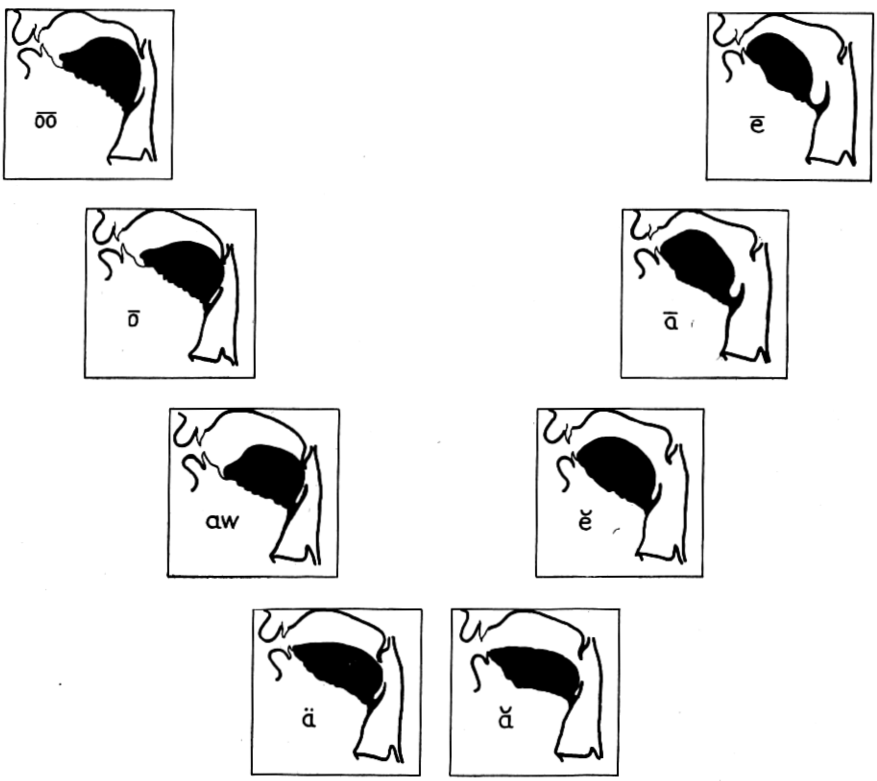


Fig. 2—Diagrams of vocal cavities for various vowel sounds

drawing conclusions from such diagrams as these, we must take care, of course, to use only the broadest features revealed by the series.[5]

It is evident that for the sounds on the left leg of the usual triangle (Fig. 3) (with the exception of short *u*), the inner orifice (that between the back of the tongue and the soft palate) is much constricted, and we

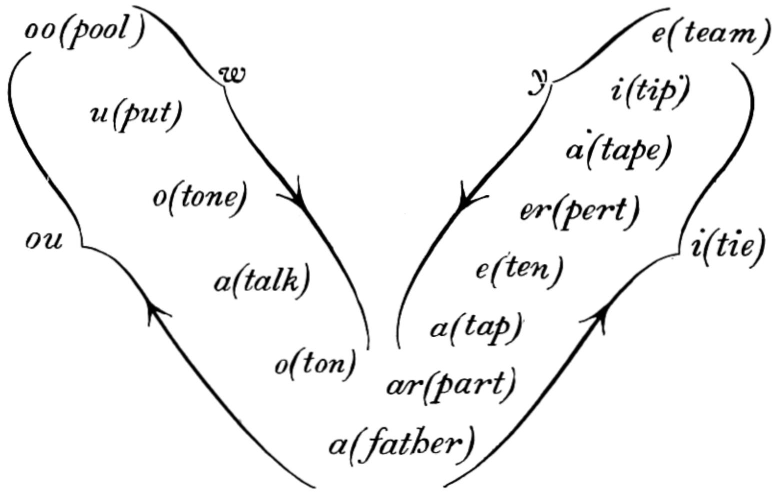


Fig. 3—Conventional vowel triangle

have here a loosely-coupled system to deal with. Also, due to the rearward position of most of the tongue structure, the mouth cavity appears larger than the pharynx. Here we must realize the horizontal width of these cavities, as well as their vertical extent. For the sounds on the right of the triangle, the tongue goes forward in such a way that the front cavity becomes the smaller of the two, and the connecting (inner) orifice becomes larger; the system then becomes closely coupled.

For some of the sounds it is not a difficult matter to get fair values for the conductivity of the mouth opening; this is approximately a circle, or an ellipse of moderate eccentricity in these cases. (The con-

[5] I understand that Prof. G. Oscar Russell, Director of the Phonetics Laboratory, Ohio State University, has made a remarkable series of clear X-ray photographs of tongue and mouth positions, for the various vowel sounds, some of which he has kindly shown me. He has worked out a special technique for making these pictures, and is now engaged in a thorough study of them, which will ultimately be published in monograph form. Unfortunately it is not possible to reproduce the pictures here; but it may be stated that the series follows (but in a more systematic way) the general course exhibited by the Rippmann diagrams shown in Fig. 2 of the present paper. The comparison between the results sketched in the present paper for the mouth cavities and results later to be published by Professor Russell should make a most interesting study.

ductivity of the circle is its diameter; we may take the conductivity of the ellipse as roughly equal to that of the circle of equal area.) In some cases, howevei (as for example, long ē), where teeth and lips are nearly closed together, the conductivity is certainly less than it appears on merely viewing the opening between the lips; hence a smaller value must be used. The conductivity of the inner orifice is even more uncertain, but in getting at this we are aided to some extent by a theoretical principle which will be given later. The diagrams at least offer some guidance in placing the various conductivities in thei order of relative magnitude.

The most serious lack of data relates to the volumes V_1 and V_2. I have made attempts to fill the mouth with water, and then measure this volumetrically, but of course this gives no hint of the volume of the pharynx. From these experiments, and other considerations, it seems that for an adult male the total volume $V_1 + V_2$ should be something over 100 cm.[3], and *nearly constant* for all the vowel sounds. That is to say, the change in V_1 and V_2 consists largely in a shift of volume from V_1 to V_2 (or vice versa) by the movement of the tongue; a proposition not so unreasonable anatomically, because competent advice states that a muscle, in taking its various shapes, preserves the same volume. Finally one would expect a somewhat larger total volume with the mouth wide open, for certain sounds, but this is partially compensated by the flattening of the cheeks in that position.

For the purposes of this study we shall consider $V_1 + V_2 \doteq 120$ cm.^{3} as one of the given data. But it may be stated in passing that these volumes should be much more accurately determined, preferably by anatomical experts.

It would be interesting to compare the results we shall obtain, for the dimensions of the resonator systems, with the actual data of Paget's resonators. But, on account of the four variables involved (K_1, K_2, V_1, V_2), there is no solution of a given case that is unique—that is to say, there are several combinations of different elements possible which will produce a given pair of natural frequencies. Hence such comparisons would often tell us little. Besides, in most cases it is impossible, from the figures given by Paget, to determine the sizes of his resonators, though their shapes are well shown in his drawings. Paget sometimes frankly imitated the structure of the mouth-pharynx system—not necessarily to scale—but sometimes, as in producing double (uncoupled) resonance by resonators in parallel, his models bore no relation to the structure of the natural system.

SPECTRA OF THE VOWEL SOUNDS

We shall take as fundamental data the average spectra of the vowel sounds (for male voices) as given by the writer's previous work with C. F. Sacia, and as given in Sir Richard Paget's chart. We thus treat Sir Richard's data as if they had been obtained analytically, and not synthetically, for the sake of taking the mean values of the two most complete series of data available, to get a better basis for calculation.

The two principal resonant frequencies for each sound are given in Table I. The lower characteristic frequency is denoted by $\omega_1/2\pi$; the other by $\omega_2/2\pi$. These characteristic frequencies are also shown in the chart of Fig. 4.

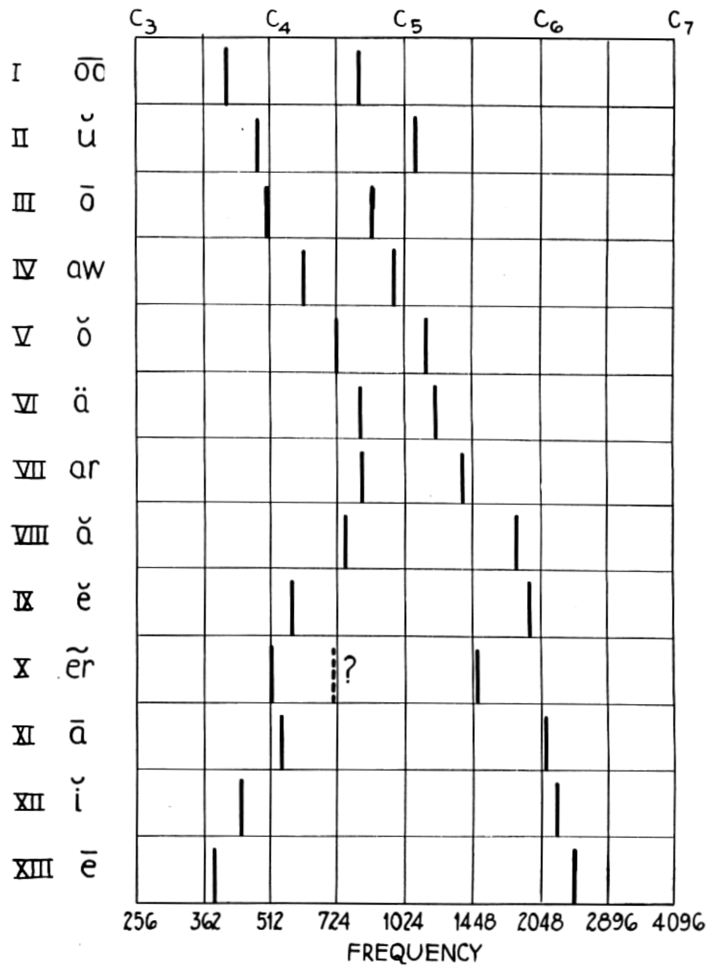


Fig. 4—Principal resonant frequencies, vowel sounds

TABLE I

NATURAL OR CHARACTERISTIC FREQUENCIES OF THE VOWEL SOUNDS

(*Male Voices*)

Sound	$\omega_1/2\pi$				$\omega_2/2\pi$			
	Crandall and Sacia	Paget (centered about)	Mean	Equiv. pitch	Crandall and Sacia	Paget (centered about)	Mean	Equiv. pitch
I. *oo* (p*oo*l) . . .	431	383	407	$G_3\#$	861	724	793	$G_4\#$
II. *u* (p*u*t) . . .	575	362	473	B_3-	1,149	966	1,058	C_5+
III. *o* (t*o*ne) . .	575	430	502	C_4-	912	790	851	A_4
IV. *a* (t*a*lk) . . .	645	558	602	$D_4\#$	1,024*	886	955	B_4
V. *o* (t*o*n) . . .	724	703†	713	$F_4\#-$	1,218	1,116†	1,167	D_5
VI. *a* (f*a*ther) .	861	790	825	$G_4\#+$	1,149	1,254	1,202	$D_5\#$
VII. *ar* (p*ar*t) . .	861	767	814	$G_4\#$	1,290	1,491	1,390	F_5+
VIII. *a* (t*a*p) . . .	813	703	758	G_4-	1,825	1,824	1,825	$A_5\#$
IX. *e* (t*e*n)	609	527	568	D_4-	1,825	1,932	1,879	B_5-
X. *er* (p*er*t) . . .	{542, 700‡}	470	{506, 700}	C_4	1,448	1,534	1,491	G_5-
XI. *a* (t*a*pe) . .	609	470	540	$C_4\#$	2,048	2,169	2,108	C_6+
XII. *i* (t*i*p)	512	362	437	A_3	2,170	2,298	2,234	$C_3\#+$
XIII. *e* (t*e*am) . .	431	332	381	G_3	2,435	2,434	2,435	$D_6\#+$

* Poorly resolved, in our charts.

† In Paget's notation, for the sound *o* as in *not*.

‡ Considering *er* to have triple resonance.

The main resonances of most of these sounds are so pronounced that it is not at all difficult to take the correct data from the original charts, gnoring the less-essential minor peaks. In only one case (*a* as in *talk*) does our original chart fail to resolve the two principal peaks, but they are partially resolved even in this case, so that there is no great uncertainty in the figure given. In the case of the sound *er*, a third frequency is shown in the diagram. Reasons will be given later for considering this sound to be produced by a system of three degrees of freedom.

MECHANISM OF THE DOUBLE RESONATOR SYSTEM

There are two ways of studying the action of the double resonator at its resonant frequencies. If we drive the back of the inner chamber with a source of prescribed motion, then the greatest motion in the orifices will be obtained when the driving point impedance of the system *as viewed from the back of the inner chamber* is infinite. Or, equally, if we drive the system (by sound waves, say) from the front orifice, then the greatest motion will be obtained for those frequencies for which the driving point impedance of the system *as viewed from without* is zero. By either method we should be able to deduce the natural frequencies of the system; the second method is chosen here because it involves less labor.

In Rayleigh (II, p. 191, eq. 12) it is shown that the natural frequencies of a double resonator of the type described are the roots ω_1, ω_2, of

$$\omega^4 - \omega^2(n_1^2 + n_2^2 + n_{12}^2) + n_1^2 n_2^2 = 0, \qquad (1)$$

in which

$n_1 = c\sqrt{\dfrac{K_1}{V_1}}$, the natural frequency of the outer resonator, with inner orifice closed;

$n_2 = c\sqrt{\dfrac{K_2}{V_2}}$, the natural frequency of the inner resonator alone;

$n_{12} = c\sqrt{\dfrac{K_2}{V_1}}$,

and c is the velocity of sound. Equation (1) is easily obtained by writing the equations of motion of the system, for zero applied forces and zero damping, and placing the determinant of the coefficients of the amplitudes or velocities equal to zero.[6] (This is equivalent to placing the driving point impedance, as viewed from the front orifice, equal to zero.) If $n_{12} \doteq 0$ (the case of a very constricted inner orifice), the roots of (1) are simply n_1, n_2.

We neglect damping in the system in order to get an easily-managed solution for the natural frequencies. Damping arises in two ways: (1) from sound absorption by the soft (tissue) lining of the cavities, and (2) by radiation from the mouth. Both are very variable, that due to radiation particularly so because of the considerable change in size of the mouth opening from one vowel sound to another. A great deal can be learned of the mechanism of the system by studying only the natural frequencies, and although it is not entirely impracticable to solve the problem with an allowance for radiation damping, we shall ignore this here.

The general procedure in this study will be to take as known from the vowel spectra the actual natural frequencies ω_1, ω_2 of the system, and to find the most reasonable values for the four quantities K_1, K_2, V_1, V_2, in order that these natural frequencies may result. We thus reconstruct the hypothetical resonator, or throat-mouth system which produces the vowel sounds. If we take

$$n_{12} = c\sqrt{\frac{K_2}{V_1}} = n_1\sqrt{\mu}, \qquad \left(\mu = \frac{K_2}{K_1}\right), \qquad (2)$$

[6] A typical solution of a double resonator problem is given in the author's "Theory of Vibrating Systems and Sound," Van Nostrand (1926), pp. 59–64. The double resonator as a sound amplifier is discussed by E. T. Paris, *Science Progress*, XX, No. 77 (1925), p. 68.

we may rewrite (1) as

$$\omega^4 - \omega^2[n_1^2(1 + \mu) + n_2] + n_1^2 n_2^2 = 0. \tag{1a}$$

If it were not for μ, we could determine from (1a) the ratios K_1/V_1 and K_2/V_2 from the known data ω_1, ω_2. As will appear later, we can make reasonable assumptions with regard to μ; but it is obvious that even then two further assumptions are required to fix K_1, K_2, V_1, V_2 in absolute value. These we supply by assuming a fixed total volume $V_1 + V_2$ for the system, and a certain conductivity K_1 for the mouth opening, which is the most easily observed element of the system.

Proceeding in the manner outlined, it will be possible to take the series of the vowel sounds and fit to each sound a doubly resonant system such that the whole series forms a more or less coherent group.

The following is an outline of the type of calculations required. If we write, from (1a),

$$\begin{aligned} n_1^2(1 + \mu) + n_2^2 &= \omega_1^2 + \omega_2^2, \\ n_1^2 n_2^2 &= \omega_1^2 \omega_2^2, \end{aligned} \tag{3}$$

and eliminate n_2^2, we have

$$\left.\begin{matrix} n_1^2 \\ n_1'^2 \end{matrix}\right\} = \frac{\omega_1^2 + \omega_2^2 \pm \sqrt{(\omega_1^2 + \omega_2^2)^2 - 4(1 + \mu)\omega_1^2\omega_2^2}}{2(1 + \mu)}; \tag{4}$$

also, if we eliminate n_1^2, we have

$$\left.\begin{matrix} n_2^2 \\ n_2'^2 \end{matrix}\right\} = \frac{\omega_1^2 + \omega_2^2 \mp \sqrt{(\omega_1^2 + \omega_2^2)^2 - 4(1 + \mu)\omega_1^2\omega_2^2}}{2}. \tag{4a}$$

In these equations it will be noted that (n_1^2, n_2^2) $(n_1'^2, n_2'^2)$ each represent possible combinations of simple resonators which will give, on coupling, the observed frequencies ω_1, ω_2. In other words, for given (comparable) conductivities K_1, K_2, of the two orifices, the outer resonator may be small, and the inner resonator large ($V_1 < V_2$), corresponding to the (separate) natural frequencies

$$\begin{aligned} n_1^2 &\equiv c^2\frac{K_1}{V_1} = \frac{\omega_1^2 + \omega_2^2 + \sqrt{(\omega_1^2 + \omega_2^2)^2 - 4(1 + \mu)\omega_1^2\omega_2^2}}{2(1 + \mu)}, \\ n_2^2 &\equiv c^2\frac{K_2}{V_2} = \frac{\omega_1^2 + \omega_2^2 - \sqrt{(\omega_1^2 + \omega_2^2)^2 - 4(1 + \mu)\omega_1^2\omega_2^2}}{2}, \\ n_1^2 &> n_2^2; \end{aligned} \tag{5}$$

or, if $V_1 > V_2$, we must apply the other pair of equations

$$n_1'^2 \equiv c^2\frac{K_1}{V_1} < n_2'^2 \equiv c^2\frac{K_2}{V_2} \tag{5a}$$

using the lower signs in (4) and (4*a*). Thus in reconstructing the resonator cavities from the vowel data, we must take care to use, for each particular vowel, that pair of solutions (n_1, n_2, or n_1', n_2') which places the front and rear cavities in correct order for relative size. From the discussion given above of the data on position of the tongue, sections of the cavities, etc., the application of this principle is a relatively easy matter.

The matter of fixing the coupling factor is not so straightforward. For the loosely coupled systems (*oo* to *ar*, the vowels on the left leg of the triangle, Fig. 3), it appears that the maximum allowable coupling factors μ (that is, the values of μ for which the radicals in (4) and (4*a*) vanish) are so small that it seems reasonable to adopt them forthwith.[7] In these cases we have the single solution

$$\left.\begin{aligned} n_1^2 &= \frac{\omega_1^2 + \omega_2^2}{2(1+\mu)}, \\ n_2^2 &= \frac{\omega_1^2 + \omega_2^2}{2}, \qquad n_1^2 < n_2^2; \\ \mu &= \frac{(\omega_1^2 - \omega_2^2)^2}{4\omega_1^2\omega_2^2} = \frac{K_2}{K_1}. \end{aligned}\right\} \tag{6}$$

In this situation (since the ratio V_2/V_1 is fixed if n_1^2/n_2^2 and K_2/K_1 are fixed) all the quantities V_1, V_2, K_1, K_2 are determinate as soon as we fix either K_1 or $V_1 + V_2$. The practice followed will be to set a value for K_1 and check this by noting the value of $V_1 + V_2$ to which it leads; thus by trial and error the most reasonable values for the resonator constants for the loosely coupled systems can be found. Incidentally, we shall note in all these cases that the solution requires V_1 to be larger than V_2.

The vowel short *ă* marks the transition between the loosely-coupled systems already considered and the closely-coupled systems for the sounds from short *ĕ* to long *ē* on the right leg of the triangle. Short *ĕ* is also the first vowel sound of the series to have a high frequency resonance of frequency greater than 1,500 cycles. We might be in a

[7] These values of the coupling factors are not inconsistent with the diagrams of the mouth cavities shown in Fig. 2. Aside from complicating the calculations, the effect of taking still smaller values for μ (keeping K_1 constant) is merely to lower V_2 in proportion as K_2 is decreased. For example, taking $\mu = \mu$ max. for the sound *aw*, we arrive at the solution $V_1 = 119$ cu. cm., $V_2 = 22$ cu. cm., if $K_1 = 2.1$ cm. as given in Fig. 5. Now if we take $\mu = \frac{1}{2}\mu$ max., we get $V_1 = 121$, $V_2 = 10$ cu. cm. Thus no great change has been made in the total volume $V_1 + V_2$, except that we get a value for V_2 which seems unreasonably small. The most satisfactory course, in the case of the loosely coupled systems, is to use the maximum allowable coupling factors.

dilemma here, as to which pair of solutions (5 or 5*a*) to apply, since solutions are possible in which the two cavities V_1 and V_2 are of comparable size in this case. It is nearly certain, however, that the front cavity, V_1, is greater than V_2 in this case, but it is not certain that the highest possible value of μ ($\mu = 1$) is the one to use. A compromise was made, setting $\mu = .80$, and using equations (5*a*) for the solution. We shall see later that a resonator built according to these specifications performs sufficiently well to justify these assumptions. With this sound we have finished with equations (6) and (5*a*) and for the last time we have $V_1 > V_2$.

For the last 5 sounds (short *ĕ* to long *ē*) the maximum possible coupling factors range from 1.75 to 9.4; it has been found advisable to shade these and use factors ranging from 1.25 to 5.0. A choice now has to be made between solutions (5) and (5*a*); and since the tongue comes so far forward in these cases, we adopt at once the first solution, according to (5), which leads to the relation $V_1 < V_2$ in all these cases.

Discussion of the Results

The calculated results are shown in the chart, Fig. 5. Because of the speculative character of some of the assumptions made it is reasonable to call attention only to certain outstanding features of the chart. Among the first seven (loosely-coupled) systems the sound *u* (as in *put*), if placed *second*, would seem definitely out of order, because of the magnitude of the coupling factor, or (what is the same thing) the greater separation of the characteristic frequencies. There is no escape from the larger inner orifice for this system, and the effect which it produces. This sound simply does not conform to the habits of its (assumed) neighbors; otherwise the first seven sounds form a coherent group. In classifying short *u* Paget takes the dilemma by the horns, and places it *first*, that is, preceding all the other sounds of this group. This arrangement is adopted in Fig. 5.

There will be noticed in the chart a tendency to expand the total volume, $V_1 + V_2$, for the rounder and more open sounds. This is in a deliberate attempt to allow for the effect of opening the mouth a little wider in these cases.

The last 5 sounds (from short *ĕ* to long *ē*) form a fairly coherent group, except for the non-conforming member *er*. Paget places *er* preceding short *ă* in the series; it seems to the writer a hybrid of the short *ĕ* (or long *ā*) and the *r* sound, but its low frequency resonance (ca. 500) requires a large volume for either V_1 or V_2, and this can only be back of the tongue (V_2) because of the contraction of V_1 when the tip of the tongue is raised for the *r* sound. If we let $K_1 = 1.5$ cm., and

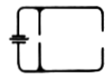
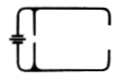
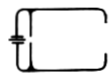
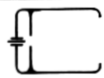
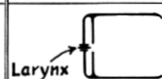
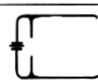
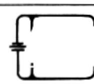
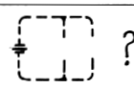
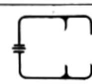
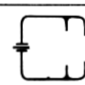
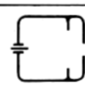
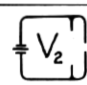
Sound	μ max.	μ used K_2/K_1	K_2	K_1	V_2+V_1	System (Schematic)	V_2	V_1
II ŭ (put)	.80	.80	.96	1.20	137		42	95
I o͞o (pool)	.50	.50	.45	.90	134		34	100
III ō (tone)	.31	.31	.45	1.50	146		28	118
IV a̤ (talk)	.23	.23	.48	2.10	141		22	119
V ŏ (ton)	.26	.26	.73	2.8	134		23	111
VI ä (father)	.15	.15	.52	3.5	126		15	111
VII ar (part)	32	.32	1.12	3.5	127		23	104
VIII ă (tap)	1.00	.80	2.0	2.5	123		23	100
X er (pert)	1.75	1.00	1.5	1.5	118		73	45
IX ĕ (ten)	2.27	1.25	2.25	1.8	117		77	40
XI ā (tape)	3.34	1.8	2.34	1.30	101		73	28
XII i (tip)	6.10	3.0	2.4	.80	102		80	22
XIII ē (team)	9.4	5.0	3.0	.60	104		83	21

Fig. 5—Schematic diagrams of doubly-resonant systems for vowel sounds

assume maximum coupling, i.e., $\mu = 1.75$, we get $V_1 = 98$ cu. cm. and $V_2 = 62$ cu. cm., which seems absurd; if we assumed for *er* a system of only *two* degrees of freedom, the most reasonable course would be to give μ a smaller value (say, unity) and solve on the basis that $V_2 > V_1$ which would give (if $K_1 = 1.5$) $V_1 = 45$ cu. cm., $V_2 = 73$ cu. cm., and $K_2 = 1.5$ cm. These data are entered (very tentatively) in Fig. 5; here again we revise the previous order, and place *er* between short *ă* and short *ĕ*.

It is not at all certain, however, from the spectra of the *er* sound (see chart, Fig. 13, in the paper "The Sounds of Speech") that it is produced by a system of only two degrees of freedom; the analyses of the female voices gave 3 definite peaks, and we note that when the tip of the tongue is raised, for this sound, there is a third cavity betewen the tongue and the lips which is doubtless significant. There will be noted, with a question mark, a third line (of frequency about 700, for the male voices) in the spectrum of *er* shown in Fig. 4. I have attempted, from the three lines shown in Fig. 4, and some simple assumptions regarding the volumes and conductivities, to obtain a rough solution, using 3 degrees of freedom for this sound; but none of these results are entered in the chart, because they appear to be unreasonable.[8]

No attempt has been made to subject the semi-vowel sounds (*l*, *ng*, *n*, *m*) to dynamical calculations. It is evident from their spectra (cf. "The Sounds of Speech") that they are produced by systems of three or four degrees of freedom, which is to be expected, if, in addition to mouth and pharynx, the tongue, naso-pharynx, or

[8] By trial and error it was hoped that some triply-resonant system could be found which would give the spectrum of *er*, as shown in Fig. 4. After solving more than a dozen of these systems, the best fit was one in which $V_1 = 31$, $V_2 = 63$, $V_3 = 31$ cu. cm.; $K_1 = K_2 = 1$ cm., $K_3 = \frac{1}{2}$ cm. The calculated frequencies for this system are 445, 890, and 1,520 cycles. The trouble with this solution is that the middle cavity (V_2, between the tongue and the roof of the mouth in this case) is the largest of the three, which does not seem reasonable. A model made to these specifications, and tried by the method described later, gave a sound something like *er* but not so satisfactorily that one could accept this as a solution. Consequently it is not entered in Fig. 5.

At first, in a number of these attempted solutions, the innermost chamber, V_3, was taken as the largest of the three. These all led to too great a separation of the two lower resonant frequencies to be acceptable.

The sound *er*, in addition to the three resonances about as shown in the chart, may contain a component of higher frequency; or it may be due to a progressive variation or modulation of the two principal frequencies shown in the chart. Some of Paget's results suggest this; and if this is so, it would be a most difficult vowel to imitate with a *fixed* resonator. It is possible that X-ray pictures may reveal some point hitherto overlooked in the mouth adjustment for this sound.

nasal cavities are brought into play. The calculations required would be too cumbersome for the present paper. It is rather a tribute to Paget's experimental skill that he was able to synthesize these more complicated sounds with resonators of more than two degrees of freedom and so arrive at their characteristics.

It is not thought that the calculations given herein suffer appreciably due to the omission of damping factors from the dynamical equations. It would be almost impossible to take correct values of damping constants from the speech spectra; there is a better chance of doing this from the records of the sounds themselves, but even so, they cannot be determined with anything like the precision of the natural frequencies.

To summarize the results, we have an idealized system of two degrees of freedom, loosely coupled for one group of the sounds, closely coupled for the remaining sounds, with fair indication of the transition between the two groups. We have the assumption of virtually constant total volume of the two cavities, and an indication of how this volume should be apportioned between them in most cases. We also have a rough determination in most cases of the conductivity of the inner orifice between the two cavities.

Some Experimental Tests

It would be of interest if we could now make models of all the systems considered, excite them in some suitable way, and establish their essential validity from the character of the sounds produced. This might seem unnecessary, on account of Sir Richard Paget's extended work; it seemed worth while, however, to attempt a few models, using cardboard tubes and plasticene for the structure.

The most success was had with the sound *a* (f*a*ther). A model was made to scale (Fig. 6), using the data of the chart—but of course

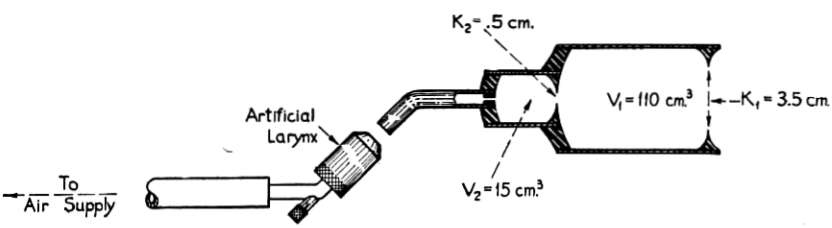


Fig. 6—Double resonator model for *ä*, and method of attaching artificial larynx

we should expect similar results from somewhat larger or smaller models, provided the ratios $K_1 : K_2 : V_1 : V_2$ were maintained; the chief point here is the variation in damping with the sizes of the orifices, and the requirement that any orifice should be smaller than the mean dimension of the adjacent volume, in order that the usual resonator theory may apply.

The model when gently blown with a slow current of air through the small hole in the back gave a good *whispered ä*; but some difficulty was experienced in exciting it correctly for a *voiced ä*. It was first connected, at the rear, to an artificial larynx,[9] keeping the connecting hole small in order to preserve the dynamical characteristics of the main system. When the artificial larynx was blown (though it did not function well with the output orifice so small), a recognizable *voiced ä* was produced by the apparatus; but this was not as good as the whispered sound first described. (We have here the point made at the beginning: that the driving system, to imitate the vocal cords successfully, must give a low pitched tone, very rich in partials.) The artificial larynx was then replaced by a telephone receiver excited by the (rip) saw-toothed A.C. wave of 100 fundamental frequency, arranged by Mr. Sacia. A rather poor sustained *ä* sound resulted, quite deficient in volume, because of weak driving through the small hole in the back. Altogether, the artificial larynx, with its intermittent or variable excitation, came the nearest to producing a voiced *ä*; and the sound was similar to that produced by a person actually using the artificial larynx inserted in the side of the mouth, in the usual manner, for this sound.

Very fair results were also obtained with a model, built according to specifications, for the sound long *ō*. Models were next attempted for short *ă* and short *ĕ*. First, a model was made with two volumes $V_1 = 80$ cu. cm., $V_2 = 45$ cu. cm., and having the *three* openings K_1, K_2, and the hole in the rear of V_2 (for a cork fitting connecting the larynx) each about 2.5 cm. in diameter. It was thought that, when blown from the rear of V_2, it would give a recognizable short *ă* sound; and that when reversed, i.e., when the cork fitting was inserted in K_1 so that V_1 and V_2 became interchanged, it would give short *ĕ*. The result was that the sounds produced were nearly alike, and quite unsatisfactory in both cases! However, when the conductivities were modified, so that $K_1 = 2.5$, $K_2 = 2.0$, for short *ă*, and $K_1 = 2.0$, $K_2 = 2.5$ for short *ĕ*, the volumes being interchanged as before, the results were much better. As described here, the model for short *ĕ* approximates in dimensions the data entered in Fig. 5, but the model for short *ă* ($V_1 = 80$, $V_2 = 45$ cu. cm.) does not quite have the theoretical division of total volume (namely, $V_1 = 100$, $V_2 = 23$ cu. cm.) entered in the chart. The partition was therefore moved back, until this condition was obtained, with the result that the short *ă* sound was given at least as well as before.

[9] Previously described by H. Fletcher and C. E. Lane.

Attempts were also made at models for long *ā* and short *ĭ*, using the theoretical data. These seemed to give *whispered* sounds which suggested the true ones, but were not very satisfactory when excited by the artificial larynx. It is evidently more difficult to imitate the mouth structure by such simple means, when the outer conductivity (K_1, the orifice between lips and teeth) is small, and the inner orifice K_2 is large. And in addition it is likely that the artificial larynx does not supply sufficient high frequency energy to excite these sounds properly. There is also, of course, the difficulty of applying the simple resonator theory, when the conductivity of an orifice is comparable to one of the dimensions of the adjacent volume.

Conclusion

In this paper we have attempted to visualize the mechanism of the vowel sounds, on the basis of previous work, certain simple calculations, and a few rough experiments. It appears that the vowel sounds are usually produced by a double resonator system whose behavior in itself is thoroughly understood; but this does not by any means close the subject. A most interesting field of study remains in the excitation of the resonator system, to say nothing of the various factors which produce damping in the system itself.

We know from laboratory experiments that a reed (or a simple "squawker" made of rubber strip) is by itself a very poor imitation of the vocal cord apparatus. The artificial larynx, for example, will not vibrate properly unless a tube some 15 inches long is interposed between the "larynx" and the pressure reservoir by which it is blown. Correspondingly, we should expect the wind-pipe leading from the lungs to the human larynx to have a very important rôle in fixing the lower frequencies produced by the vocal cord apparatus. The mechanical problem indicated for study in this connection is the excitation of a reed-pipe with the reed at the distant end of the pipe, an inversion of the arrangement of ordinary wind instruments.

Consider the question of damping. In the apparatus used by J. Q. Stewart [10] (tuned electrical circuits excited by an interrupter) the damping could be systematically adjusted; this is the only case I know of, in experimenting with speech sounds, in which this adjustment was possible. In ordinary mechanical apparatus damping is difficult to control. Yet, damping is a significant element in the character of the constituent vibrations of either sustained or transient vowel sounds. For example, I have already pointed out [11] the close

[10] *Nature*, Sept. 2, 1922. "An Electrical Analogue of the Vocal Organs."

[11] "The Sounds of Speech," end of § V. Refer also to Records and Fig. 14 of that paper.

similarity between the spectra of *l* and long *ē*. In the semi-vowel *l* the characteristic high frequency (if viewed as a transient) decays much more rapidly than the corresponding vibration in the *ē* sound; this fact we have from the records themselves, but not from the frequency spectra. It may be that such phenomena as these will require a more definite adherence to the "transient" point of view in dealing with the vowel sounds, a matter previously discussed at some length.

The transitory or unstable qualities in the actual speech sounds almost defy imitation by mechanical means. There is, for example, the variation in fundamental frequency during the course of a vowel or semi-vowel sound which was pointed out in the paper "The Sounds of Speech." There is also the lengthening of the fundamental period for semi-vowels and voiced consonants as compared with vowel sounds; also the shortening of the fundamental cycle at the beginning of a voiced consonant.

Finally there is the question of classification of the speech sounds. We have already noted difficulties for some of the vowel sounds. It is likely that the vowel triangle or the arrangement of the vowels in a linear series will require modification. A satisfactory classification for all the sounds, from the dynamical standpoint, is at present an unsolved problem; but in conclusion one suggestion may be permissible. We might limit the application of the term "vowel sound" to those sounds which *can* be satisfactorily produced by the simple double resonator system. The more complicated vowel-like sounds (*l*, *ng*, *n*, *m* and possibly *r*) and some of the consonants can undoubtedly be related to systems of three or more degrees of freedom. A study of these systems is beyond the aims of the present paper; but it is to be hoped that such a study can be carried out, for the sake of the aid that mechanical theory offers in helping to visualize the mechanism of speech.