Appeared in Studia Phonetica Posnaniensia, Adam Mickiewicz University Press, vol. 1: 1987. pp. 149-171. Author's current address: Siemens Corporate Research, 755 College Road East, Princeton NJ 08540.

From Acoustic Tube to Acoustic Cues

Gary Kuhn
IDA-CRD
1 Thanet Road
Princeton, NJ 08540

Abstract

We illustrate the relation between sound propagation in the acoustic tube and formant-based acoustic cues for the phonetic dimension of place of articulation.

1 Introduction

We model an acoustic tube as a pressure source exciting an acoustic filter. We model the pressure source at frequency f as being given by $\cos(2\pi ft)$. We model the filter which transmits the acoustic disturbance with a vector of cross-sectional area measurements $a(x_i)$ at coordinates $x_i, i = 1, 2, ..., n$, where $x_1 = 0$ cm, $x_n = l$ cm, l cm is the length of the tube, and $dx = x_{i+1} - x_i$ is the uniform spacing of the measurements.

Ignoring losses, we can define the response of the filter to the source in terms of two functions of position and time. These functions are the excess pressure at frequency f, $p_f(x, t)$, and the particle velocity at frequency $f, u_f(x, t)$:

$$p_f(x,t) = p_f(x)\cos(2\pi ft),$$
$$u_f(x,t) = u_f(x)\sin(2\pi ft).$$

Each of these functions is a product of a function of x and a function of t. The functions of x are a pressure function, $p_f(x)$, and a particle velocity function, $u_f(x)$. The functions of t are the harmonic excitation, $\cos(2\pi ft)$, and the same function with a phase difference of $\pi/2$, or $\sin(2\pi ft)$.[1]

To represent the source and the filter response for voiced speech we model the source as having excitation at all frequencies which are integral multiples, m, of some lowest or fundamental frequency, f_0. Then the net response of the filter is a net pressure, $p(x, t)$, and a net particle velocity, $u(x, t)$, each defined as a sum over frequencies mf_0, $m = 1, 2, ..., M$, of the corresponding frequency-dependent quantities:

$$p(x,t) = \sum_f p_f(x,t),$$
$$u(x,t) = \sum_f u_f(x,t).$$

Finally, we can define the function called volume velocity, $\chi_f(x,t)$, which equals area times particle velocity:

$$\chi_f(x,t) = \chi_f(x)\sin(2\pi ft) = a(x)u_f(x)\sin(2\pi ft).$$

If the tube has uniform cross-sectional area, then volume velocity is related to particle velocity by a single multiplicative constant. The vocal tract has approximately uniform cross-sectional area for the vowel /ə/.

Starting with this physical model, here are some problems that we can solve:

1. Find the zeroes over frequency of p_f(lips), i.e., the formant frequencies.
2. Relate constriction of the uniform tube to changes in formant frequencies.
3. Relate changes in formant frequencies to acoustic cues for place of articulation.
4. Relate constriction of other tubes to changes in formant frequencies.
5. Relate changes in formant frequency to rules derived from listening tests.

In the next sections of this paper these problems are addressed one at a time.

2 Find the Formant Frequencies

To solve problem 1 we can apply Webster's horn equation [1,2,3,4]. Webster's equation relates the velocity potential, Φ, to the area function $A(x)$:

$$\frac{\partial^2}{\partial x^2}\Phi + \frac{\partial}{\partial x}\ln A\frac{\partial}{\partial x}\Phi - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\Phi = 0$$

where c is the velocity of sound ≈ 333m/s. The quantity Φ is related to pressure and particle velocity by two further equations [2]:

$$p = \rho\frac{\partial}{\partial t}\Phi,$$

$$u = -\frac{\partial}{\partial x}\Phi.$$

The quantity p_f(lips) does not depend on time. So we can simplify Webster's equation to get rid of the dependency on time. If we let $\Phi = \varphi_f(x)\sin(2\pi ft)$, then Webster's equation becomes

$$\varphi'' + (\ln A)'\varphi' + k^2\varphi = 0 \tag{1}$$

where $' = d/dx$ and $k^2 = (2\pi f/c)^2$.

The equation relating φ and u becomes

$$\varphi' = -u.$$

And, the equation relating Φ and p becomes

$$p = \rho 2\pi f\varphi_f(x)\cos(2\pi ft).$$

It is clear from this last equation that $\varphi_f(x)$ is proportional to $p_f(x)$, the pressure function of x. It follows that the zeroes over frequency of $\varphi_f(\text{lips})$ are identically the zeroes over frequency of $p_f(\text{lips})$. So, by solving for the zeroes of $\varphi_f(\text{lips})$ we will find all and only the formant frequencies.

To derive an expression for u', we can differentiate $\varphi' = -u$ and substitute in (1), obtaining

$$u' = k^2\varphi - (\ln a)'u. \tag{2}$$

This done, we make a choice about how to represent $(\ln a)'$ between measurements. Following Chiba and Kajiyama [2] we have chosen to use the average derivative of log area, $\bar{a}'(i)$, on each interval i, $i = 1, 2, ..., n-1$, where

$$\bar{a}'(i) = \ln[a(x_{i+1})/a(x_i)]/dx, a(x_i) > 0,$$

and to represent pressure, particle velocity and volume velocity as linear in each interval, so that all four functions are approximated as piecewise linear.

Next, we define boundary conditions for φ and u at $x = 0$, that is, at the back end of the tube. The boundary condition for the pressure quantity φ is $\varphi(0) = 1$. Together with the definition of the source as $\cos(2\pi ft)$, this condition implies that at (x,t) = (0,0) the air at the back of the tube is maximally compressed. The boundary condition for the particle velocity quantity u is $u(0) = 0$. This condition is consistent with a tube that is closed by an unyielding wall at the back, which approximates the nearly closed back of the vocal tract during voiced speech.

Now we can compute the pressure and particle velocity quantities, φ and u, transmitted from the left edge, $x = x_0$, to the right edge, $x = x_{dx}$ of interval i, $i = 1, 2, ..., n-1$. Note that the output pressure quantity of the preceding interval, $\varphi_{dx}(i-1)$, will be the input pressure quantity of the following interval, $\varphi_0(i), i > 1$. Similarly, the output particle velocity quantity of the preceding interval, $u_{dx}(i-1)$, will be the input particle velocity quantity of the following interval, $u_0(i)$.

Two cases arise. Either $\bar{a}'$ is zero on the interval, or it is nonzero. If $\bar{a}' = 0$, then (2) simplifies to

$$u' = k^2\varphi.$$

From $\varphi' = -u$ we also have $u' = -\varphi''$, so

$$\varphi'' = -k^2\varphi.$$

The reader can verify by differentiating twice that

$$\varphi = \alpha\cos(kx) + \beta\sin(kx), \tag{3}$$

and

$$\varphi' = -\alpha k\sin(kx) + \beta k\cos(kx) = -u. \tag{4}$$

Solving for α and β at x_0 using Cramer's rule:

$$\alpha = \frac{\begin{vmatrix} \varphi_0 & \sin(kx_0) \\ \varphi_0' & k\cos(kx_0) \end{vmatrix}}{\begin{vmatrix} \cos(kx_0) & \sin(kx_0) \\ -k\sin(kx_0) & k\cos(kx_0) \end{vmatrix}} = \frac{\varphi_0 k\cos(kx_0) - \varphi_0'\sin(kx_0)}{k} \tag{5}$$

$$\beta = \frac{\begin{vmatrix} \cos(kx_0) & \varphi_0 \\ -k\sin(kx_0) & \varphi_0' \end{vmatrix}}{\begin{vmatrix} \cos(kx_0) & \sin(kx_0) \\ -k\sin(kx_0) & k\cos(kx_0) \end{vmatrix}} = \frac{\varphi_0'\cos(kx_0) + \varphi_0 k\sin(kx_0)}{k} \tag{6}$$

Substituting (5) and (6) in (3) and reordering terms,

$$\begin{aligned} \varphi(x) &= \varphi_0[\cos(kx)\cos(kx_0) + \sin(kx)\sin(kx_0)] \\ &+ \tfrac{\varphi_0'}{k}[\sin(kx)\cos(kx_0) - \cos(kx)\sin(kx_0)]. \end{aligned}$$

Applying the formulas for $\cos(a+b)$, and $\sin(a+b)$ we obtain

$$\varphi(x) = \varphi_0 \cos k(x - x_0) + \frac{\varphi_0'}{k}\sin k(x - x_0).$$

We find the pressure transmitted to the end of interval i, by setting $x = x_0 + dx$:

$$\varphi_{dx}(i) = \varphi_0\cos(kdx) + \frac{\varphi_0'}{k}\sin(kdx). \tag{7}$$

Either by substituting (5) and (6) in (4), or by differentiating (7),

$$\varphi_{dx}'(i) = -\varphi_0 k\sin(kdx) + \varphi_0'\cos(kdx) = -u_{dx}(i), \tag{8}$$

so the particle velocity at the end of the interval is

$$u_{dx}(i) = \varphi_0 k\sin(kdx) - \varphi_0'\cos(kdx). \tag{9}$$

The second case arises when $\bar{a}' \neq 0$. We can estimate the average value of pressure on interval i, $\varphi_e(i)$, as a linear function of the initial conditions for pressure and particle velocity on the interval. Let $\Delta x = dx/2$. Then, since pressure is piecewise linear, its average value on interval i occurs at $x = x_0 + \Delta x$, i.e. in the middle of the interval. The slope of pressure is obtained from $\varphi' = -u$ and the initial condition for particle velocity:

$$\begin{aligned} \varphi_e(i) &= \varphi_0(i) + \varphi'(i)\Delta x \\ &= \varphi_0(i) - u_0(i)\Delta x. \end{aligned}$$

Next, we can estimate the average value of particle velocity on interval i, $u_e(i)$, by integrating (2). To integrate (2) we are forced to approximate pressure as constant on the interval, so we use the average value just computed, $\varphi_e(i)$ [2]. Writing

$$\begin{aligned} c_1 = & \quad k^2\varphi, \\ c_2 = & \quad \bar{a}', \end{aligned}$$

so that (2) becomes

$$\frac{d}{dx}u = c_1 - c_2 u,$$

we integrate, exponentiate and solve for u:

$$\begin{aligned} \frac{1}{-c_2}\ln(c_1 - c_2 u) &= x + c_3 \\ c_1 - c_2 u = e^{-c_2 x + c_3} &= A e^{-c_2 x}, \end{aligned}$$

where $A = e^{c3}$, and so

$$u = \frac{c_1 - A e^{-c_2 x}}{c_2}. \tag{10}$$

The factor A,

$$A = (c_1 - c_2 u)e^{c_2 x},$$

can be estimated for interval i using $\varphi_e(i)$ and the boundary condition u_0 at $x = x_0$:

$$A = (c_1 - c_2 u_0)e^{c_2 x_o}. \tag{11}$$

Substituting (11) in (10):

$$\begin{aligned} u &= \tfrac{c_1 - (c_1 - c_2 u_o)e^{c_2 x_0} e^{-c_2 x}}{c_2} \\ &= \tfrac{c_1}{c_2}(1 - e^{-c_2(x - x_o)}) + u_0 e^{-c_2(x - x_o)}. \end{aligned} \tag{12}$$

Since particle velocity is piecewise linear, its estimated average value on interval i, $u_e(i)$, occurs at $x = x_0 + \Delta x$. In estimating $u_e(i)$, therefore, the indices of (12) simplify and we obtain

$$u_e(i) = \frac{c_1}{c_2}(1 - e^{-c_2 \Delta x}) + u_0 e^{-c_2 \Delta x}, \tag{13}$$

where $c_1 = k^2\varphi_e(i)$ and $c_2 = \bar{a}'(i)$.

Given these equations, we can now write a second expression for the average value of pressure on interval i. Our first expression, from the initial conditions on the interval, was

$$\varphi_e(i) = \varphi_0(i) - u_0(i)\Delta x,$$

but given the particle velocity at the middle of the interval, $u_e(i)$, we can compute

$$\varphi_c(i) = \varphi_0(i) - u_e(i)\Delta x.$$

Also, we can now write a second expression for the average value of particle velocity on interval i. Our first expression, from $\varphi_e(i)$ and integrating u', was

$$u_e(i) = \frac{c_1}{c_2}(1 - e^{-c_2\Delta x}) + u_0 e^{-c_2\Delta x},$$

but given $\varphi_e(i)$ and $u_e(i)$, we can now compute $u_c(i)$, where

$$u'(i) = k^2\varphi_e(i) - \bar{a}'(i)u_e(i),$$

and

$$u_c(i) = u_0(i) + u'(i)\Delta x.$$

When values are computed for these expressions, we hope that $\varphi_c(i)$ is close to $\varphi_e(i)$ and $u_c(i)$ is close to $u_e(i)$. Then we may decide to forget any differences. But if these pairs of quantities are not satisfactorily close, the traditional way around the problem has been the following:

1. note that $u_e(i)$, $\varphi_c(i)$ and $u_c(i)$ are all functions of $\varphi_e(i)$, and
2. perturb $\varphi_e(i)$ in an iterative numeric or graphic procedure until acceptable numbers are obtained [2].

Alternatively, following a suggestion from I. Kessler [5], we can give an algebraic expression for the value of $\varphi_e(i)$ that minimizes the following error criterion:

$$E^2(i) = [\varphi_c(i) - \varphi_e(i)]^2 + [u_c(i) - u_e(i)]^2. \tag{14}$$

Let

$$z = \varphi_e(i), \tag{15}$$

and

$$u_e(i) = zV + a,$$

where $V = k^2(1 - e^{c_2\Delta x})/c_2$, and $a = u_0(i)e^{-c_2\Delta x}$. Then let

$$\begin{aligned} u'(i) &= k^2 z - c_2(zV + a), \\ u_c(i) &= u_0(i) + u'(i)\Delta x, \end{aligned} \tag{16}$$

$$\begin{aligned} \varphi'(i) &= -u_e(i), \\ \varphi_c(i) &= \varphi_0(i) - (zV + a)\Delta x \\ &= b - zV\Delta x, \end{aligned} \tag{17}$$

where $b = \varphi_0(i) - a\Delta x$, so that we can define the pressure error, $E_\varphi(i)$, as

$$\begin{aligned} E_\varphi(i) &= \varphi_c(i) - \varphi_e(i) \\ &= b - zY, \end{aligned}$$

where $Y = V\Delta x + 1$, and the particle velocity error, $E_u(i)$, as

$$\begin{aligned} E_u(i) &= u_c(i) - u_e(i) \\ &= zW + d, \end{aligned}$$

where $W = k^2\Delta x - V(c_2\Delta x + 1)$, and $d = u_0(i) - a(c_2\Delta x + 1)$. Then (14) becomes

$$E^2(i) = (b - zY)^2 + (zW + d)^2,$$

and the error as we change $\varphi_e(i)$ has a minimum at

$$\frac{d}{dz}E = W(zW + d) - Y(b - zY) = 0,$$

or

$$\varphi_e(i) = z = \frac{bY - dW}{Y^2 + W^2}. \tag{18}$$

Calculating $\varphi_e(i)$ from (18), $u_e(i)$ from (15), $u'(i)$ from (16), and $\varphi'(i)$ from (17) we obtain error minimizing values for the pressure transmitted to the end of the interval, $u_{dx}(i)$, from

$$\begin{aligned} \varphi_{dx}(i) &= \varphi_e(i) + \varphi'(i)\Delta x \\ u_{dx}(i) &= u_e(i) + u'(i)\Delta x. \end{aligned}$$

Repeating our calculations over $n-1$ intervals, we obtain estimates of the pressure and particle velocity transmitted to the open end of the tube. Repeating our calculations over the frequencies of interest, we find those frequencies for which φ_f(lips) is zero, or the formant frequencies of the tube with the given area function, thus solving problem 1.

Note that we have used the term "formant frequency" to mean exactly a "natural resonance frequency" of the tube [6]. We should keep the notion of a formant frequency distinct from that of a "formant frequency estimate," e.g., a measurement made with the use of spectrograms.

We have compared formant frequencies calculated with these procedures to formant frequency estimates measured from spectrograms, for six vowels spoken while x-rays were made of the vocal tract area function [3]. The average magnitude of the error was 10% for F_1, 13% for F_2 and 8% for F_3.

3 Relate constriction of the uniform tube to changes in formant frequencies

Experience with Webster's equation has led at least three researchers to make or repeat the following statement:

> When part of a pipe is constricted, its resonant frequency becomes low or high according as the constricted part is near the maximum point of the volume (velocity) or of the excess pressure. [2,3,7] (19)

We emphasize this statement because it predicts qualitatively how individual formant frequencies should change as a tube is constricted at different x coordinates. To illustrate these predictions quantitatively for the uniform tube, we must decide how to model constriction of the tube, and we have to compute the x coordinates of the extrema of volume velocity and pressure at each formant frequency for the tube before it is constricted. Then we can compare the predictions of statement (19) with computed changes in formant frequencies.

To model constriction of the uniform tube, we let $A(x)$ be a vector of 17 cross-sectional area measurements, each equal to 4 cm^2, with a spacing of 1 cm. Then we constricted this 16 cm uniform tube on centimeter-long intervals starting with interval 1, at the closed back of the tube and moving centimeter by centimeter to interval 16, at the open front of the tube.

As indicated at the top of Figure 1, constriction at each interval was modeled in six steps. First, the uniform tube was left unconstricted on interval i: $a(x_i)$ and $a(x_{i+1})$ were left at 4 cm^2. Second, the uniform tube was constricted half-way to a target constriction of .5 cm^2 on interval i, that is to 2.25 cm^2. Third, the uniform tube was constricted to .5 cm^2 on interval i. Then, redundantly, the same amounts of constriction were modeled in reverse order for steps four, five, and six. The first three formant frequencies were computed for the tube at each of the six steps, by the procedure given in Section 2.

At the bottom of Figure 1, the formant frequencies for constriction of interval 12 (4-5 cm from the lips), are plotted. The abscissa is labeled with the step number; the ordinate is labeled in kHz. The lines connect the values obtained for each formant frequency over the six steps. This formant pattern suggests both the formant frequencies and the formant frequency changes that should occur if a 16 cm vocal tract were used to make a vowel-constriction-vowel articulation that starts and ends with the vowel /ə/, and has the constriction located on the interval 4-5 cm from the lips.

To locate the extrema of volume velocity and pressure before the tube is constricted, we turn to Figure 2. In the top panel of Figure 2, we display $\chi_f(x)$ and $-\chi_f(x)$, the volume velocity function of x and its negative, at the first three formant frequencies. The curves are shifted vertically by an amount proportional to the formant frequency. Over time, volume velocity at each frequency varies between the higher and the lower of the two curves. The zeroes of volume velocity at each frequency occur at those x coordinates where the two curves intersect. The extrema occur where the two curves are maximally separated. Note that there is an additional zero and extremum for each higher formant. Note also that in the fourth panel of Figure 2 we display the approximate position of the fixed articulators along a 16 cm vocal tract [2].

According to the top panel of Figure 2, there are extrema of volume velocity at 3.2 cm, 9.6 cm, and 16 cm (the lips) at the third formant frequency. Also, there are extrema of volume velocity at 5.3 cm and 16 cm for F_2, and at 16 cm for F_1. In the same panel, there are zeroes of volume velocity for F_3 at 0 cm, 6.4 cm and 12.8 cm. Also, there are zeroes of volume velocity for F_2 at 0 cm and 10.7 cm and for F_1 at 0 cm.

In the uniform tube, volume velocity is related to particle velocity by a single multiplicative constant, and we know that $\varphi' = -u$. It follows that the coordinates of the zeroes of volume velocity are the coordinates of the extrema of pressure.

According to statement (19), constriction at an extremum of volume velocity for a particular formant should lower that formant frequency. Therefore, in the second panel of Figure 2, labeled "Effect on resonance," we have placed a minus sign in the row for each formant, on each centimeter-wide interval which contains an extremum of volume velocity in panel one. Similarly, we have placed a plus sign in the row for each formant on each centimeter-wide interval which contains an extremum of pressure in panel one.

In other words, the minuses and pluses indicate statement (19)'s predictions about what should happen to each formant frequency if the uniform tube is constricted on any of the marked intervals. The prediction for changes in formant frequencies on the unmarked intervals might reasonably be that they vary monotonically with the "nearness" to the neighboring marked intervals.

The third panel of Figure 2, labeled "Characteristic formants," displays one graph of formant frequencies on each interval $i, i = 1, 2, ..., 16$. Across intervals, the abscissa still indicates centimeters for panel 3. But within each interval the abscissa indicates the constriction step. The formants are at the frequencies for constriction step 1 at the left edge of the interval and for constriction step 6 at the right edge of the interval. For example, the formant graph from Figure 1, for constriction of interval 12, can be recognized on interval 12. Interpreting the constriction steps as occurring over time, the ith graph enables us to see what the formant frequencies should look like as a constriction on interval i is made and then subsequently released.

Let us call a change in formant frequency over time a "formant transition." The formant transitions in panel three of Figure 2 are, in general, consistent with statement (19). Where a lowered formant frequency was predicted, the change in that formant frequency is negative going into the constriction, and then positive coming out. Where a raised formant frequency was predicted, the change in that formant frequency is positive going into the constriction, and then negative coming out. Note, however, that the marked interval is sometimes adjacent to the interval with the greatest change in formant frequency in the predicted direction: for example, F_3 moves lower on interval 3 than on interval 4. Note also that on some intervals intermediate between the marked intervals, the direction of formant transition shows a slight reversal from constriction step 2 to constriction step 3: for example, F_3 on interval 2 rises at step 2 and falls at step 3.

4 Changes in formant frequency are acoustic cues to place of articulation

For phonetics, the significance of (19) and Figure 2 is that they predict a physical basis for acoustic cues to "place of articulation," as we now explain.

Panel 5 of Figure 2 is a phonetician's array of phonetic symbols for consonants of American English[2]. Each row of this array represents a "manner of articulation" for the consonants. Thus the /m/ in the first row differs from the /b/ in the second row by being a nasal instead of a stop consonant. Each column of the array represents a "place of articulation" of the consonants. Thus the same /m/, in column 16, differs from the /n/ in column 13 by being a bilabial nasal instead of an alveolar nasal.

At some (row, column) positions of the array, two phonetic symbols appear. In these positions, the two consonants differ by their "voicing": the consonant to the left is "voiced," and the consonant to the right is "voiceless." Thus the (stop, bilabial) /b/ differs from the (stop, bilabial) /p/ by being voiced. Place of articulation is, then, a phonetic dimension along which the consonants of American English, and, we believe, all languages, can be ordered.

Panel 5 has been aligned with the rightmost six columns of the physical arrays in panels 1-4. In this position, the phonetic dimension of place of articulation correlates well with the physical dimension which gives the location of the fixed articulators. Thus, a bilabial consonant constricts the vocal tract at the lips, a labial-dental consonant constricts the vocal tract a little further back, at the lips and teeth, and so on.

One implication of this alignment is that the phonetic sequence /ə- any bilabial consonant -ə/ should have formant frequencies that look like those directly above, in graph 16 of panel 3. Another implication is that the sequence /ə- any labiodental consonant -ə/ should have formant frequencies like those in graph 15, and so on. In short, Figure 2, predicts formant patterns for the various places of articulation from the distributions of volume velocity and pressure for the 16 cm uniform tube.

Listening tests have shown that acoustic stimuli with formant patterns like those of panel 3 can function for human listeners as sufficient acoustic cues to the place of articulation categories of panel 5. For example, speakers of American English listened to stimuli whose time-varying formants look like those at steps 4 through 6 in graphs 11 through 16. The formant amplitudes rose rapidly at the beginning of the stimuli. The listeners perceived a consonant-vowel stimulus. The listeners were asked to label the consonant either /g/, /d/ or /b/. The predominant response was /g/ for patterns like graph 11, /d/ for patterns like graph 13, and /b/ for patterns like graph 16 [8]. Analogous results have been obtained in labeling experiments with nasals, [9] fricatives, [10] and sonorants [11].

Other observations and experimental results reveal, however, that Figure 2 is an oversimplification of the relation between formant patterns and place of articulation. For one thing, the inventory of place of articulation categories to be mapped onto constriction coordinates varies over languages. French, for example, has a voiced pharyngeal sonorant /R/ that could be aligned at interval 6 [12]. For another, Figure 2 implies that consonants of identical place of articulation have identical formant patterns. It turns out that consonants of identical place of articulation have only similar formant patterns, for reasons that include the following:

4.1 Vocal tract length and vocal tract proportions vary over talkers.

As a result, different formant frequencies and different formant frequency ratios are observed over talkers even for sounds perceived phonetically as the same [13,14,15,16].

4.2 The relation of physical place of constriction to phonetic place of articulation can be many to one.

For example, constriction coordinates that cover at least intervals 10 through 12 have to be used to model the velar consonants [3].

4.3 Air flow characteristics vary over manner of articulation.

4.3.1. At the glottis, increased flow due to abducted vocal cords can occur during a fricative as well as at the release of an affricate or stop. As a result, F_1 can be weak or absent during a fricative as well as at the release of an affricate or stop [9,12].

4.3.2. At the constriction, high-speed turbulent flow can occur during a fricative as well as at the release of an affricate or stop. As a result, high-frequency noise can occur during a fricative as well as at the release of an affricate or stop [9,12]. Zero flow at the constriction occurs for the occlusives (flaps, affricates, stops and nasals), with the result during the flaps, stops and affricates that all formants briefly disappear [12,17,18].

4.3.3. At the velum (the soft palate), flow is diverted through the nose for nasals, which causes "nasal" formants to appear, at different frequencies [12,19].

4.4 Airflow characteristics vary over voicing.

4.4.1. At the glottis, the vocal cords are more completely abducted for voiceless fricative, affricates and stops. The results include increased flow, a total absence of any fundamental frequency of vocal cord vibration (the "voice bar" in the spectrum), and the presence of glottal noise or "aspiration" [12].

4.4.2. At the constriction, the increased flow creates conditions for stronger and perhaps also longer turbulence during a fricative, as well as at the release of an affricate or stop [12].

4.5 Constriction characteristics vary over manner of articulation.

4.5.1. The amount of constriction tends to be less for sonorants than for fricatives, and less for fricatives than for occlusives. Where there is less constriction, formant frequencies deviate less from those for the uniform tube, and less from those for other places of articulation [3,20].

4.5.2. The rate of constriction can be slower for sonorants than for fricatives, and slower for fricatives than for occlusives. Thus the time derivative of formant frequency can be smaller (as above) and the time interval of transition can be longer, for less constricted sounds [21].

4.5.3. The duration of maximum constriction can be longer for sonorants and fricatives than for flaps, affricates and stops, so the formant frequencies can exhibit their maximum deviation from the frequencies of the uniform tube during a longer period of time for the former than for the latter [12].

4.6 Constriction characteristics vary over voicing.

4.6.1. The duration of maximum constriction varies over voicing, with voiceless consonants tending to be longer [22].

4.6.2. The timing of constriction varies over voicing, with longer vowels tending to occur before voiced consonants [23].

For reasons such as these, the formant patterns for consonants with the same place of articulation will not be identical, but only similar. To sum up, in a fixed phonetic environment, here /ə- consonant -ə/, the direction of formant transitions appears not to vary across consonants at a single place of articulation, but the same cannot be said for the excitation, extent and timing of transitions.

5 Relate changes in formant frequencies across area functions

The most important shortcoming of Figure 2 is that it does not give formant patterns for constrictions applied to a starting non-uniform area function. Statement (19), however, predicts that these formant patterns, too, will depend on the distributions of volume velocity and pressure. So, in Figure 3, we show area functions for six vowels, /ieɯəoa/ [2], volume velocity at the first three formant frequencies for each area function, and the formant patterns for constriction of each area function. The area functions are either 16 or 17 cm long. Constriction of each centimeter long interval i was modeled in six steps as before: $a(x_i)$ and $a(x_{i+1})$ were first left unconstricted, then constricted half-way to .5 cm^2, then constricted to .5 cm^2, and then back again. No starting $a(x)$ was less than .5 cm^2.

Three observations can be made directly from Figure 3. First, there are large proportional changes across area functions. For example, the area at the lips for the vowel /o/ or /ɯ/ is approximately one quarter of that for the vowel /a/. Also the area at the middle of the mouth for the vowel /i/ is an order of magnitude smaller than that for the vowels /a/ or /o/. Finally, the area of the pharynx for /a/ or /o/ is approximately an eighth of that for the vowel /i/.

Second, those extrema of volume velocity and pressure that are not at the lips or the glottis do move somewhat. We only show volume velocity but since $\varphi' = -u$, the extrema of pressure vary similarly to the zeroes of volume velocity. For example, volume velocity for F_3 has its first zero at 4 cm from the lips for /i/, 3 cm for /ɯ/, 2.3 cm for the rounded /o/, 3 cm for /a/, 3.2 cm for /ə/ and 3.5 cm for /e/.

Third, the extreme downward or upward movements of the formants still occur near the coordinates of the extrema of volume velocity and pressure. For example, volume velocity at F_3 for /i/ has its first zero at 4 cm from the lips, and the formant patterns on the fourth and fifth intervals show maximum upward movement of F_3. Volume velocity at F_3 for /ɯ/ has its first zero at slightly more than 3 cm from the lips, and the formant pattern on the fourth interval shows maximum upward movement. And volume velocity at F_3 for /o/ has its first zero at 2.3 cm from the lips, and the formant pattern on the third interval shows maximum upward movement.

The formant patterns of Figure 3 are aligned at the lip interval, enlarged and superimposed in Figure 4. Note that F_2 is shifted 500 Hz and F_3 1000 Hz, so that F_1, F_2 and F_3 do not overlap. Some changes in formant transitions with vowel formant frequency are noticeable. For example, the coordinate of the maximum upward movement of F_1 moves forward from 14 cm as the vowel F_1 decreases. Also the amount of downward transition for F_1 with a lip constriction, or for F_2 with a pharyngeal constriction, increases with the vowel formant frequency.

If we take the difference between each formant frequency and its step 1 or vowel formant frequency, then we can compare just the formant transitions. This has been done in Figure 5. Note how similar these formant transitions are to the formant frequency pattern for constriction of the uniform tube. Where a plus or minus transition was predicted for the uniform tube in Figure 2, the same direction of transition is observed in Figure 4. So, F_1 always rises on the interval starting 15 cm from the lips, and falls on the interval at the lips. F_2 always rises on the interval 15 cm from the lips, falls at 10 cm, rises at 5 cm and falls at 0 cm from the lips. F_3 rises at 15 cm, falls at 12 cm, rises at 9 cm, falls at 6 cm, rises at 3 cm and falls at 0 cm from the lips. These observations suggest that the distributions of the extrema of volume velocity and pressure are after all very similar across vowel area functions [4].

Large and variable transitions occur 5-6 cm from the lips (F_3 down and F_2 up), 3-4 cm from the lips (F_3 up), and 0-1 cm from the lips (F_2 down). These are the approximate coordinates of the widely used velar, apical and bilabial articulations. Small and relatively invariant transitions occur 2-3 cm from the lips, at the approximate coordinate of the rarer lingua-dentals. Apparently, languages are more likely to signal place of articulation with transitions that are more extreme, even if they are not more invariant.

6 Relate changes in formant frequency to rules derived from perception tests

Psychologists have experimented with rules for synthesizing stimuli to produce the desired percepts of place of articulation. In the best-known rules, the notion of a target formant frequency or "formant locus" was introduced [9,12,24]. As an example of how a formant locus can be used, look at the falling and then rising transitions of F_1 on the lip interval of Figure 4. Starting from the six different vowel formant frequencies, six different amounts of downward formant movement take place as the constriction is made on this interval of the tube. A formant locus makes possible one rule for synthesizing all six patterns. The rule is this: start at the vowel formant frequency, go half-way toward a locus frequency of 200 cycles, and then return. This single rule is very adequate for describing all the F_1 transitions on this interval.

However, no similar rule can describe the transitions modeled on the lip interval for either F_2 or F_3: there is no single target frequency toward which all the transitions may be said to point. Some transitions are shallow and at high frequencies while others are deep and at low frequencies. Futhermore, no single target frequency appears for F_2 or F_3 anywhere in the intervals corresponding to the places of articulation for English.

Still, formant loci have had success as a basis for rule synthesis. To understand this, we can compare Figure 6 with Figures 4 and 5. Figure 6 plots the F_2 and F_3 loci proposed for each English place of articulation [10, 12]. Place of articulation is on the abcissa; the proposed F_2 and F_3 loci are listed, each ordered by increasing frequency, on the ordinate. An "x" in the grid indicates the F_2 or F_3 locus for place of articulation at that column.

The proposed F_2 loci move higher in frequency as place of articulation moves from bilabial back to velar. The proposed F_3 loci move higher and then lower as place moves over the same values. Similarly, the average frequency of F_2 at extreme constriction in Figure 5, and the average deviation of F_2 in Figure 6, move higher as place of articulation moves from bilabial back to velar. Also the average frequency of F_3 at extreme constriction in Figure 5, and the average deviation of F_3 in Figure 6, move higher and then lower as place moves from bilabial back to velar.

In other words, the proposed formant loci may give the best single frequency toward which to point the formants. Here "best" means that these loci produce the fewest confusions of intended and perceived place of articulation.

For the velar place of articulation, however, it was found quite early that the single locus had to be replaced by a locus that varies with vowel context [9]. Thus, the velar F_2 locus given in parentheses in Figure 6, is solely for the context of unrounded nonback vowels. Then with just a few loci, far fewer than the number of vowels, identifiable and more natural stimuli were produceable by rule [25].

Context-dependent loci are now being used for all places of articulation in automatic speech recognition [26]. They are called "pseudo-loci" and are defined as the difference between the formant frequency seen at extreme constriction and that at the loudest part of the vowel. Pseudo-loci appear to insure that the transitions to be synthesized or recognized are more like the deviations expected from Figure 5.

7 Summary

We have reviewed sound propagation in the acoustic tube for the purpose of computing formant frequencies. We derived changes in formant frequencies by constricting a model tube. We related the changes in formant frequencies to acoustic cues for the phonetic dimension of place of articulation. We compared changes in formant frequencies by constricting tubes with different cross sectional areas. And we related trends in the changes in formant frequencies to formant based cues motivated by listening tests. To sum up, we have used the physical model to illustrate the relation between sound propagation in the acoustic tube and formant based acoustic cues for the phonetic dimension of place of articulation.

8 Acknowledgements

The author is grateful to D. L. DeGeorge, A. S. House, I. J. Kessler, D. I. Lieberman, J. B. Lurie, and C. A. Olano for many helpful discussions of this work.

9 Footnotes

1. In the acoustic tube, forward-going and reflected waves at frequency f require one quarter of a period to advance from coordinates where identical pressures intersect (extreme pressure) to coordinates where identical particle velocities intersect (extreme particle velocity). For a review of sound propagation in an elastic medium, see H. Levitt and D. Monk. 1981. "Sound transmission in tubes: a simplified treatment and an analysis of errors of approximation." In A. S. House, Ed., Proceedings of a Symposium on Acoustic Phonetics and Speech Modeling. Princeton: Institute for Defense Analyses – Communications Research Division. 2: Paper 14.

2. After Peterson, G. B., and J. E. Shoup. 1966. A physiological theory of phonetics. J. Speech Hearing Research. 9: 5-67.

10 REFERENCES

[1] Webster, A. G. 1919. Acoustical impedance, and the theory of horns and of the phonograph. Proceedings of the National Academy of Sciences. 5: 275-282.

[2] Chiba, T. and M. Kajiyama. 1941. The vowel its nature and structure. Reprinted in 1958 by the Phonetic Society of Japan.

[3] Fant, C. G. M. 1960. Acoustic Theory of Speech Production. 's Gravenhage, Mouton.

[4] Fant, C. G. M., and S. Pauli. 1974. Spatial characteristics of vocal tract resonance modes. Preprints of the Speech Communication Seminar. Speech Transmission Laboratory, Stockholm. 2: 121-132.

[5] Kessler, I. J. 1982. Personal communication.

[6] *Stevens, K. N., A. House. 1961. An acoustical theory of vowel production and some of its implications. J. Speech Hearing Res. 4: 303-320.

[7] Mermelstein, P. 1967. Determination of the vocal-tract shape from measured formant frequencies. J. Acoust. Soc. Amer. 41: 1283-1294.

[8] Harris, K. S., H. Hoffman, A. Liberman, P. Delattre, F. Cooper. 1958. Effect of third-formant transitions on the perception of the voiced stop consonants. J. Acoust. Soc. Amer. 30: 122-126.

[9] *Cooper, F. S., P. Delattre, A. Liberman, J. Borst, L. Gerstman. 1952. Some experiments on the perception of synthetic speech sounds. J. Acoust. Soc. Amer. 24: 597-606.

[10] Delattre, P. 1962. Formant transitions and loci as acoustic correlates of place of articulation in American fricatives. Studia Linguistica. 16: 104-121.

[11] Lisker, L. 1957. Minimal cues for separating /w,r,l,j/ in intervocalic production. Word. 13: 257-267.

[12] Delattre, P. 1968. From acoustic cues to distinctive features. Phonetica. 18: 198-230.

[13] Fant, C. G. M. 1975. Non-uniform vowel normalization. Speech Transmission Laboratory Quarterly Progress and Status Report. 2-3/75: 1-19.

[14] *Peterson, G. E., H. Barney. 1952. Control methods used in a study of the vowels. J. Acoust. Soc. Amer. 24: 175-184.

[15] *Ladefoged, P., D. Broadbent. 1957. Information conveyed by vowels. J. Acoust. Soc. Amer. 29: 98-104.

[16] Rand, T. C. 1971. Vocal tract size normalization in the perception of stop consonants. Haskins Laboratories Status Report on Speech Research. 25/26: 141-146.

[17] Hughes, G., M. Halle. 1956. Spectral properties of fricative consonants. J. Acoust. Soc. Amer. 28: 303-310.

[18] *Halle, M., G. Hughes, J. P. Radley. 1957. Acoustic properties of stop consonants. J. Acoust. Soc. Amer. 29: 107-116.

[19] Fujimura, O. 1962. Analysis of nasal consonants. J. Acoust. Soc. Amer. 34: 1865-1875.

[20] Lieberman, P. H., D. Klatt, W. Wilson. 1969. Vocal tract limitations on the vowel repertories of rhesus monkey and other non-human primates. Science. 164: 1185-1187.

[21] *Liberman, A. M., P. Delattre, L. Gerstman, F. Cooper. 1956. Tempo of frequency change as a cue for distinguishing classes of speech sounds. J. Exp. Psych. 52: 127-137.

[22] Lisker, L. 1957. Closure duration and the voiced-voiceless distinction in English. Language. 33: 42-49.

[23] *House, A. S., G. Fairbanks. 1952. The influence of consonant environment upon the secondary acoustical characteristics of vowels. J. Acoust. Soc. Amer. 25: 105-113.

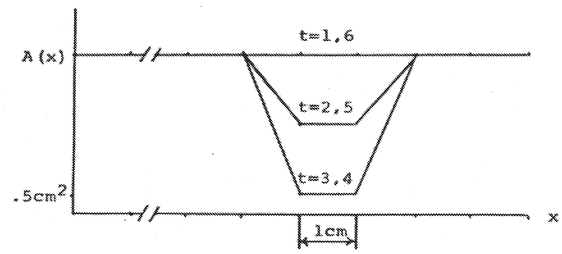
[24] *Delattre, P. C., A. M., Liberman, F. S. Cooper. 1955. Acoustic loci and transitional cues for consonants. J. Acoust. Soc. Amer. 27: 769-773.

[25] Holmes, J. N., I. G. Mattingly, J. N. Shearme. 1964. Speech synthesis by rule. Language and Speech. 7: 127-143.

[26] Demichaelis, P., R. DeMori, P. Laface, M. O'Kane. 1983 Computer recognition of plosive sounds using contextual information. ASSP 31: 359-377.

* Reprinted in Lehiste, I., ed. 1967. Readings in Acoustic Phonetics. (Cambridge, Mass. M.I.T. Press.)

```
c Appendix 1. FORTRAN subroutine for computing PHIf(X), Uf(X).
      SUBROUTINE niwe(n,dx,f,phi(1))    !called for each freq f
      implicit integer (a-z)
      parameter (size=16)
      real dlnadx(size)                 !d(log area)/dx
      real phi(size), avphi(size)       !phi, avg phi
      real u(size), avu(size)           !u, avg u
      real dphidx(size), dudx(size)     !d(phi)/dx, d(u)/dx
      real a,b,c,c2,d,delx,dx,f,k,V,W,Y

c initialize
      delx=dx/2.                        !dist. to midpoint of section
      u(1)=0.                           !init. condition for part vel
      c=33333.                          !velocity of sound, cm/sec
      k=2.*3.14159*f/c                  !radians*cycles/cm

c for each tube section
      do i=1,n-1                        !n sections, each of length dx
      c2=dlnadx(i)                      !derivative of log area

c if cylindrical tube section: exact values of phi, u at end of section
      IF (c2.eq.0) THEN
      phi(i+1)=phi(i)*  cos(k*dx)-u(i)*sin(k*dx)/k
      u(i+1)  =phi(i)*k*sin(k*dx)+u(i)*cos(k*dx)
      ELSE

c otherwise, conic section: estimate avg phi, avg u in section
      a=u(i)*exp(-c2*delx)
      b=phi(i)-a*delx
      d=u(i)-a*(c2*delx+1)
      V=k**2*(1-exp(-c2*delx))/c2
      W=k**2*delx-V*(c2*delx+1)
      Y=V*delx+1
      avphi(i)=(b*Y-d*W)/(Y*Y+W*W)      !minimizes phierr**2 + uerr**2
      avu(i)=avphi(i)*V+a

c estimate d(phi)/dx, d(u)/dx in section
      dphidx(i)=-avu(i)
      dudx(i)=k**2*avphi(i)-c2*avu(i)   !k^2*phi = c1

c approximate phi and u at end of section
      phi(i+1)=avphi(i)+dphidx(i)*delx
      u(i+1)=avu(i)+dudx(i)*delx
      ENDIF

c continue with next section
      enddo
      return
      end
```



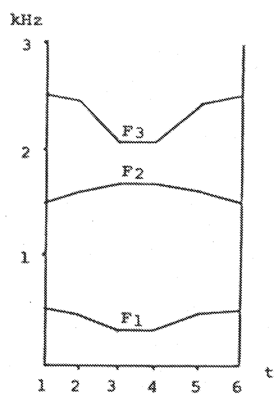


Figure 1. Modelling constriction of interval i.

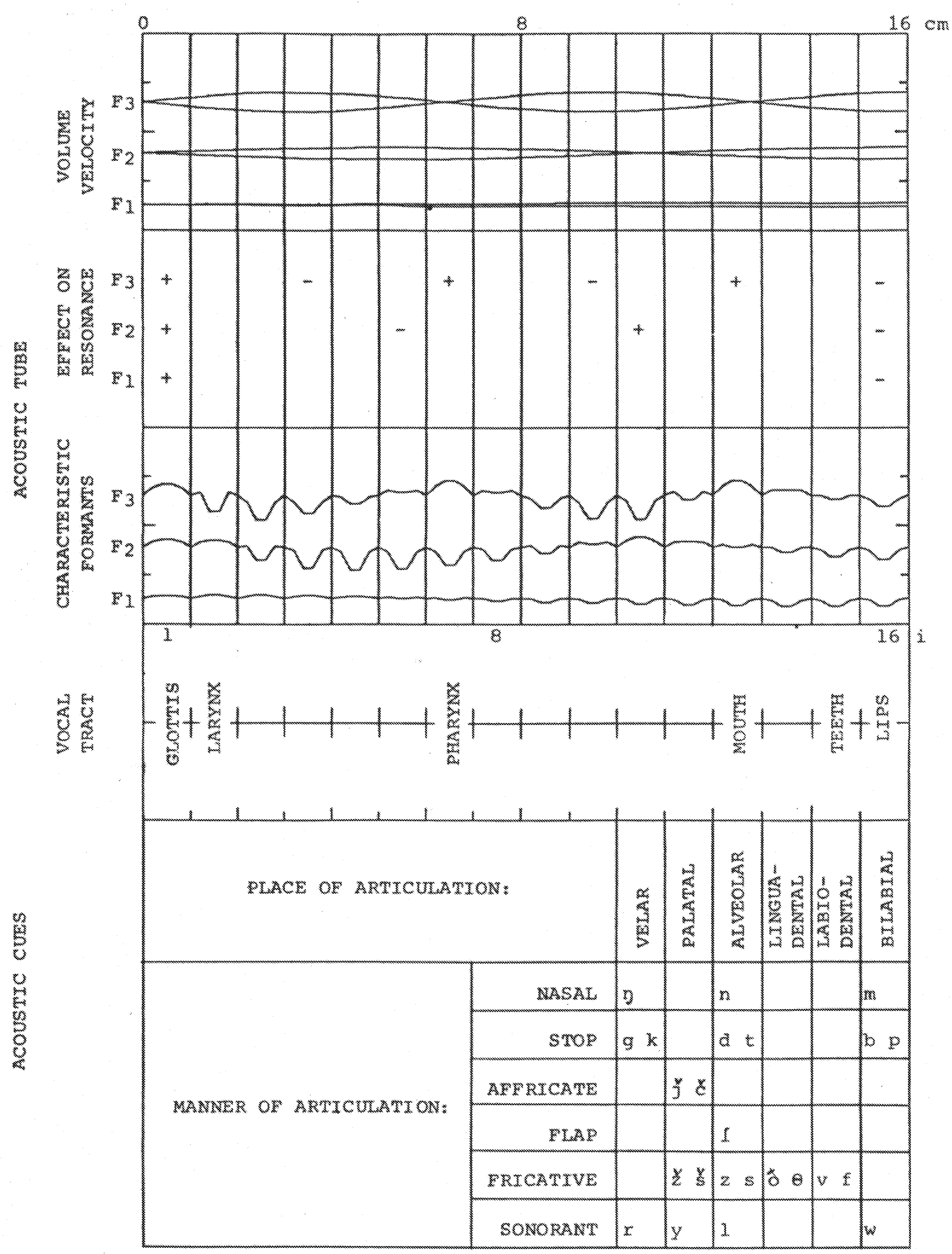


Figure 2. Acoustic tube and acoustic cues: /ə- constriction -ə/.

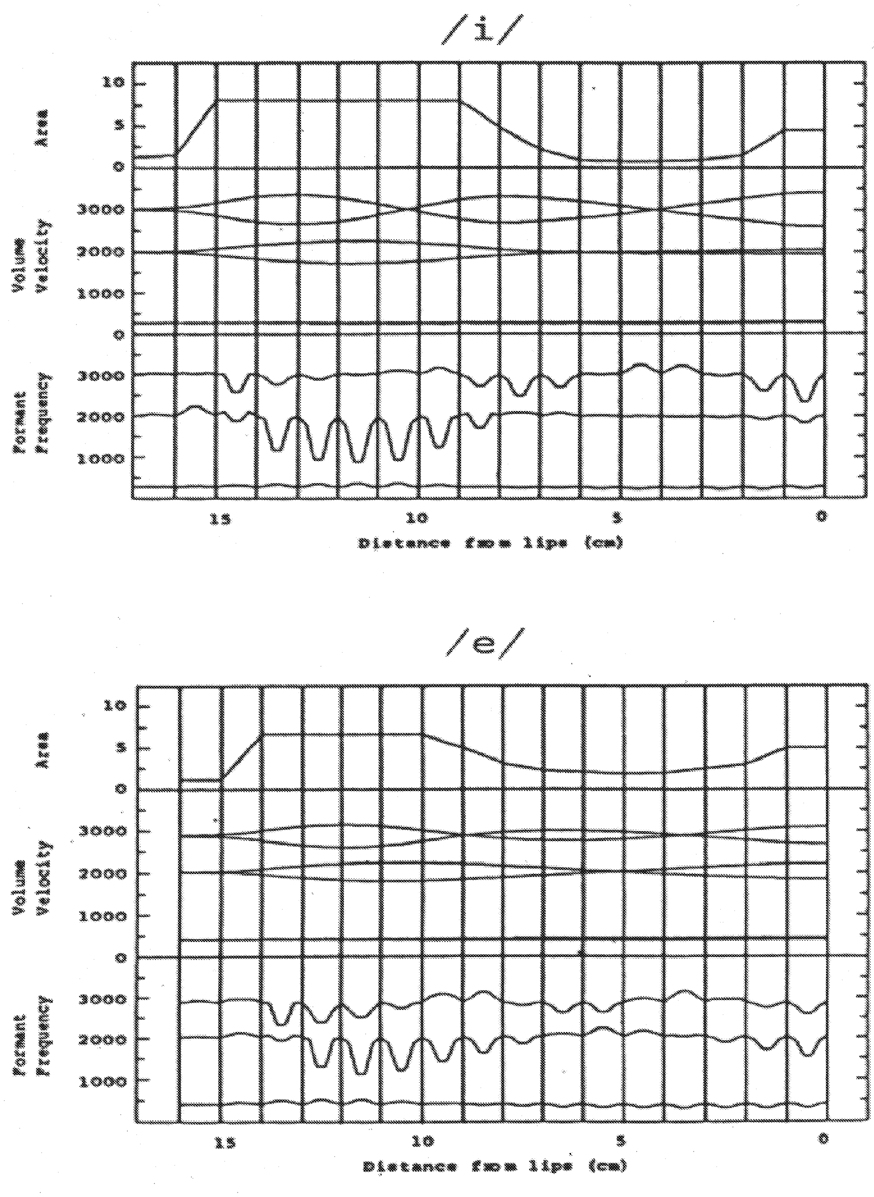


Figure 3a. Acoustic tube: /V - constriction - V/, V = /ie/.

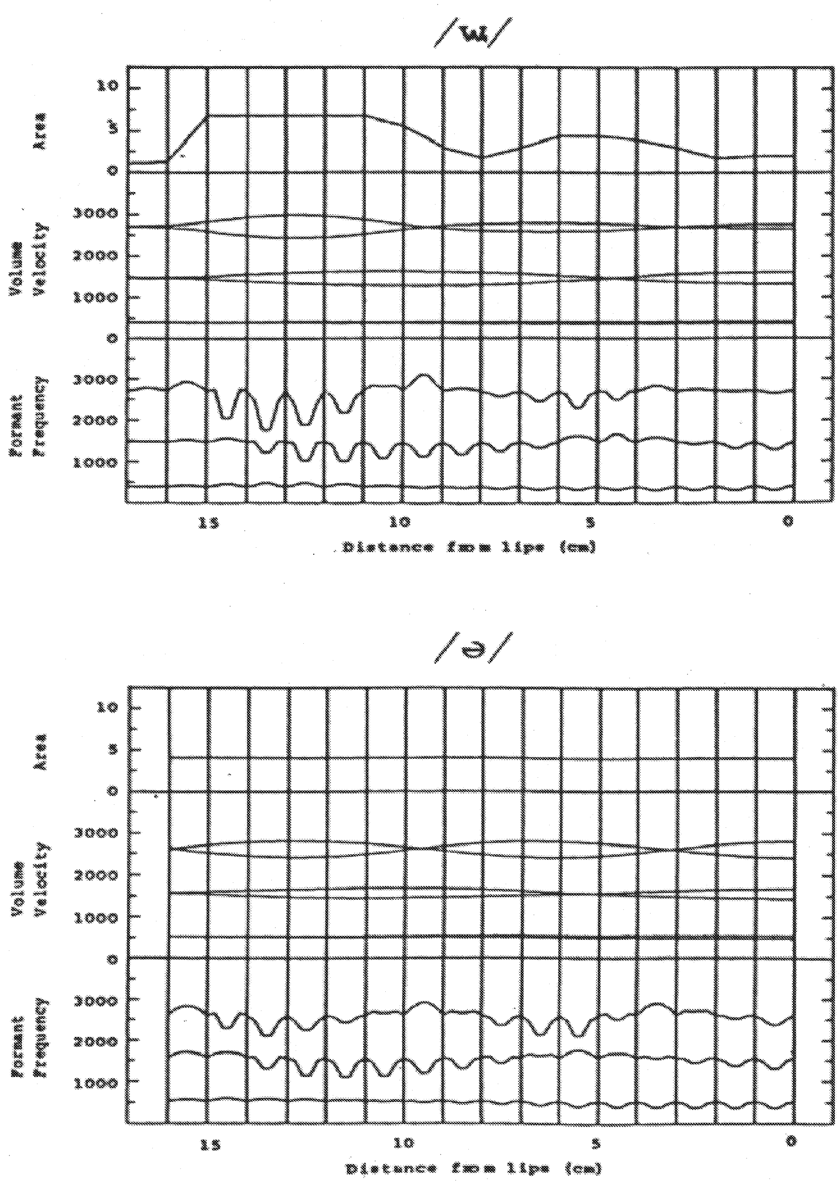
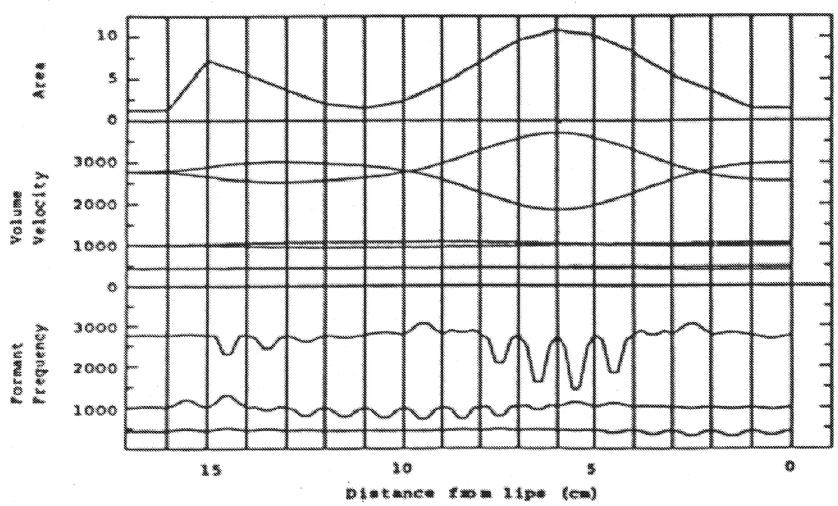


Figure 3b. Acoustic tube: /V - constriction - V/, V = /ɯə/.

/o/



/a/

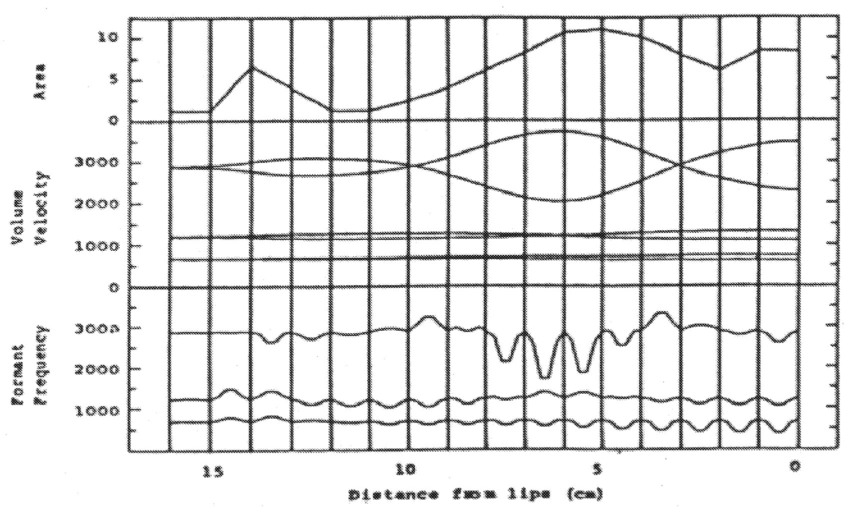


Figure 3c. Acoustic tube: /V - constriction - V/, V = /oa/.

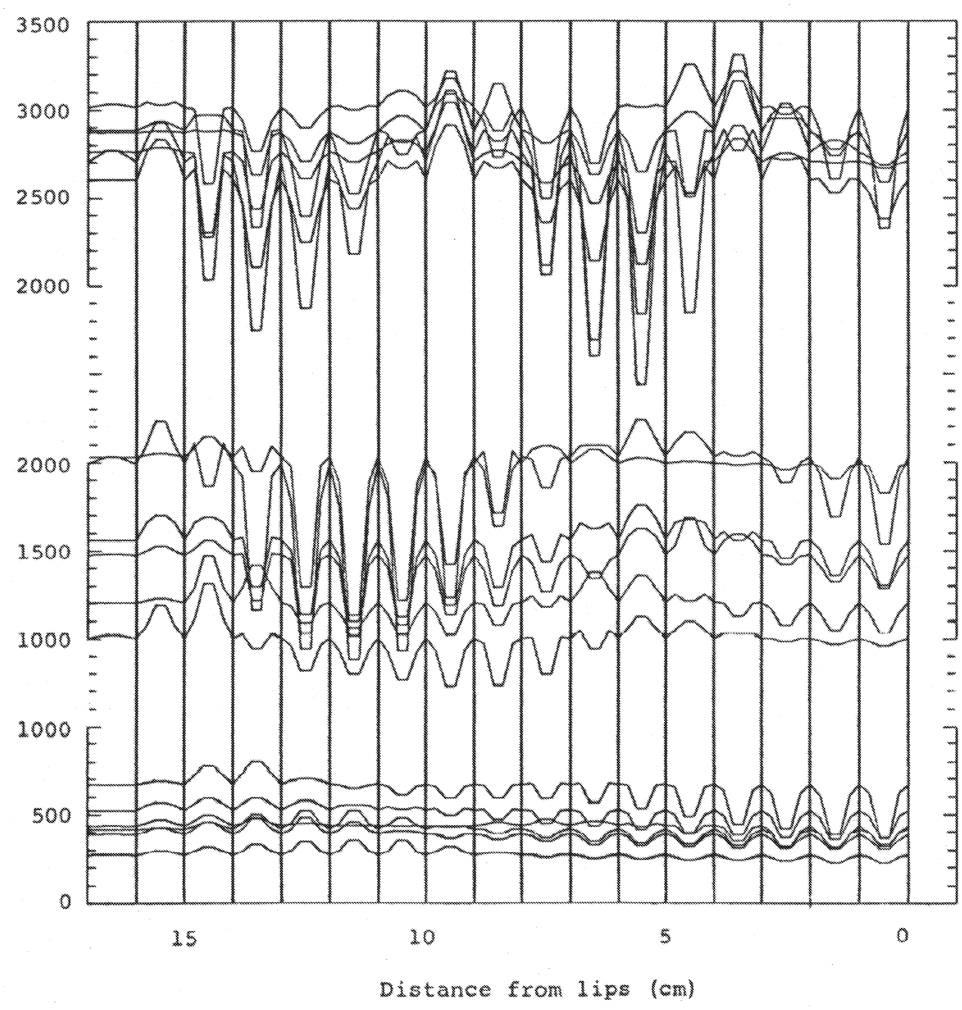


Figure 4. Formant frequencies superimposed: /V - constriction - V/.

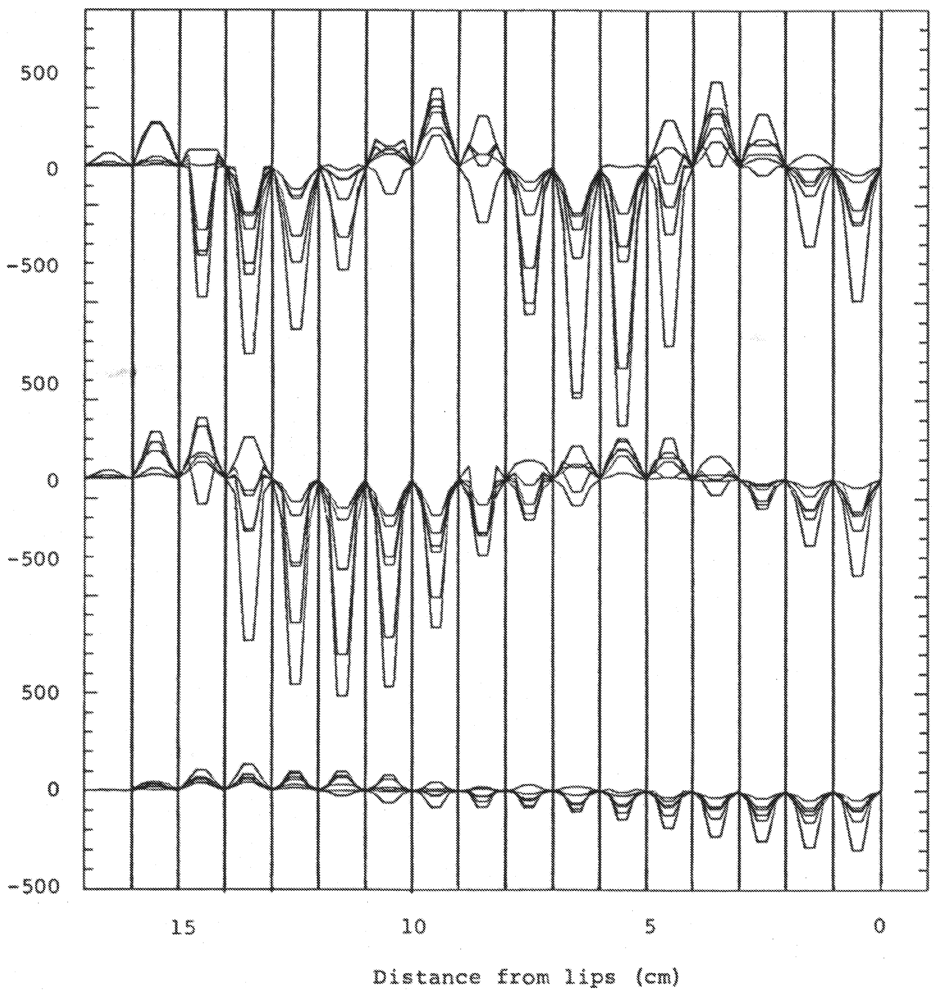


Figure 5. Formant transitions superimposed: /V - constriction - V/.

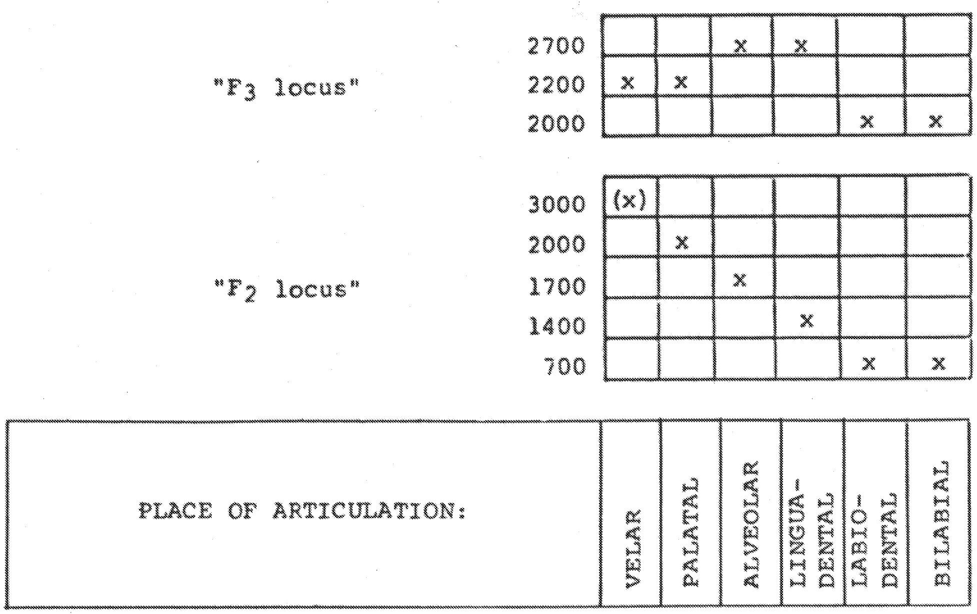


Figure 6. Formant loci, after Delattre.