Toward phonetically grounded distinctive features.
Part I: Acoustic-articulatory correlations in a four-region model of the vocal tract

Mark Pennington
mwpennin@indiana.edu

Abstract

The purpose of this paper is to correlate the four formant frequencies F1–F4 and quality factors Q1–Q4 with the positions, areas, or area ratios formed by the four active articulators: tongue root, tongue body, blade, and lips. The quality or amplification factor Q is the formant frequency F divided by the bandwidth B. Toward this end, a 27-tube frequency-domain vocal tract model (FDVT) is developed. Four articulator regions are delimited in the model: an 8-tube tongue root region, a 9-tube tongue body region (one-quarter wavelength of F2), a 6-tube blade region (one-quarter wavelength of F3), and a 4-tube lip region. Vowel area functions of ten speakers were taken from seven X-ray and MRI studies and fit to 27 equal-length tubes using cubic spline interpolation. Correlation matrices between the acoustic and articulatory parameters are calculated for the vowel system of each speaker. The coefficients of the parameter pairs are averaged across the ten speakers and the most highly correlated ones are ranked. Tongue root aperture (tongue root area normalized by lip area) is shown to be inversely correlated with F1 frequency. When the tongue body and blade constrictions move toward the lips through their respective one-quarter wavelengths, the F2 and F3 frequencies also shift higher. Tongue body aperture (tongue body area normalized by lip area) is directly correlated with Q2. Similarly, blade aperture (blade area normalized by lip area) is directly correlated with Q3. Lip protrusion displays an inverse correlation with F4 frequency. Lip aperture (lip area) has a moderate direct correlation with F1 and a weaker inverse correlation with Q4. In agreement with perturbation analysis, F1 frequency is found to be closely correlated with the tongue root area, but negligibly correlated with the tongue body area. Consequently, F1 distinctions among high, mid, and low vowels are made by varying the tongue root aperture and not, as is traditionally assumed, by raising or lowering the tongue body.

1. Introduction

There are only a few studies reporting anatomical measurements of the vocal tract in which acoustic-articulatory or articulatory-articulatory correlation coefficients are calculated and ranked. For example, Fairbanks [1950] correlated vowel intensities and X-ray measures of mouth opening at three locations. Ladefoged et al. [1972] found the correlations among jaw opening, tongue height, and tongue root advancement in an X-ray investigation. Using ultrasound scans, Stone et al. [2004] established correlations among five local segments of the tongue. Before 1990 a general assessment of acoustic-articulatory correlations would have been nearly impossible due to the scarcity of X-ray data. The vowel area functions of only two speakers were known up to that time, which are clearly not enough to provide meaningful correlation averages across subjects [Fant, 1960; Mrayati and Guérin, 1976]. With the development of MRI techniques after 1990, the vowel area functions of eight more speakers became available. As a consequence, it is now possible to achieve more robust correlation averages.

The goal of this paper is to correlate the first four formant frequencies F1–F4 and quality factors Q1–Q4 with the positions, areas, or area ratios formed by the four active articulators: tongue root, tongue body, blade, and lips. The quality or amplification factor

Q is defined as the formant frequency F divided by its bandwidth B. To achieve this objective, a 27-tube frequency-domain vocal tract model (FDVT) is developed (Section 2). The FDVT model is tested against a Matlab-based vocal tract model (VTAR) and the validity of the present model is confirmed (Section 3). Perturbation analysis of the open-closed uniform pipe is applied to the vocal tract, dividing it acoustically into four articulator regions (Sections 4.1–4.3). Each region is characterized by an active articulator: an 8-tube tongue root region, a 9-tube tongue body region corresponding to a quarter wavelength at the second formant frequency, a 6-tube blade region corresponding to a quarter wavelength at the third formant frequency, and a 4-tube lip region. The tubes of the lip (4) and tongue root (8) regions appear to be those remaining by default once the quarter-wavelength tubes of the blade (6) and the tongue body (9) are determined. The division of the vocal tract into four articulator regions is applicable to speech sounds because 1) the closed glottis condition is nearly always met and 2) the locations of the volume velocity maxima and minima of a lossy non-uniform pipe largely coincide with those of the uniform pipe [cf. Mrayati and Carré, 1976].

Vowel area functions of ten speakers were obtained from seven X-ray and MRI studies and fit to 27 equal-length tubes using cubic spline interpolation (Section 5). The articulatory parameters (positions, areas, area ratios) formed by the four articulators (tongue root, tongue body, blade, lips) are defined in Section 6, whereas the acoustic parameters, such as the formant frequencies (F1–F4), bandwidths (B1–B4), quality factors (Q1–Q4), and the overall power (PWR), are reviewed in Section 7. Correlation matrices between the acoustic and articulatory parameters are calculated for the vowel system of every speaker. The coefficients of the parameter pairs are then averaged across the ten speakers. The highest-ranking acoustic-articulatory correlations of each formant are presented and analyzed in Sections 8.1 through 8.3. A summary of the results is as follows (Section 8.4):

1. Tongue root aperture (tongue root area normalized by lip area) is inversely correlated with F1 frequency.
2. As the tongue body position (location of smallest constriction) moves toward the lips, the F2 frequency also shifts higher.
3. Tongue body aperture (tongue body area normalized by lip area) is directly correlated with Q2.
4. As the blade position (location of smallest constriction) moves toward the lips, the F3 frequency also shifts higher.
5. Blade aperture (blade area normalized by lip area) is directly correlated with Q3.
6. Lip position (sum of tube lengths in lip region) displays an inverse correlation with F4 frequency.
7. Lip aperture (lip area) has a moderate direct correlation with F1 frequency and a weaker inverse correlation with Q4.

The seven motor dimensions (positions, apertures) are similar to the oral tract variables proposed by Browman and Goldstein (1989), with the exception of tongue root aperture which was not discussed.

The elaboration of distinctive feature theory has been hindered by uncertainty about acoustic-articulatory relations, vowel height being a case in point (Section 9). Jones [1909, p. 10] related impressionistic or perceived vowel height to the vertical position of the highest point of the tongue. Subsequent X-ray measurements showed a rather weak correspondence between perceived vowel height and the vertical position of the tongue [Russell, 1928; Wood, 1975]. According to perturbation analysis, a narrower lip or wider tongue root area decreases F1 frequency while a wider lip or narrower tongue root area increases F1 frequency [Chiba and Kajiyama, 1958; Fant, 1960; Schroeder, 1967]. Nevertheless, the raising or lowering of the tongue body is still assumed to produce high or low sounds with correspondingly low and high F1 frequencies [Chomsky and Halle, 1968, pp. 304–305; Halle, 1992; Stevens, 1998, p. 250, 261]. In the absence of compelling data to the contrary, it is understandable that there has been some reluctance to abandon the traditional view of vowel height put forward by Jones. The present paper offers conclusive experimental evidence against the tongue body hypothesis. The average correlation coefficients between F1 frequency and the cross-sectional area of the tongue root, lip, and tongue body regions are respectively: –0.915, 0.723, and –0.154. As predicted by perturbation analysis, changes in tongue root, and to a lesser degree, lip area have a substantial effect on F1. On the other hand, the negligible correlation of –0.154 demonstrates that the raising or lowering of the tongue body has little influence on F1.

2. Description of the Frequency-Domain Vocal Tract model (FDVT)

A frequency-domain vocal tract model (FDVT) was developed by the author using Fortran 77. The vocal tract frequency response is computed by a transmission line lattice consisting of 27 single-tube T-sections [Fant, 1960, pp. 36–38]. Each T-section of length l is made up of two series circuits $a = Z \tanh(\Gamma / 2)$ and one shunt circuit $b = Z / \sinh \Gamma$, where the characteristic impedance is $Z = \sqrt{(R + j\omega L)/(G + j\omega C)}$ and the transfer constant is $\Gamma = l\sqrt{(R + j\omega L)(G + j\omega C)}$ [Fant, 1960, p. 28]. The per-unit-length analogous elements are:

a) acoustic mass $L = \rho / A$

b) viscous loss $R = (S / A^2)\sqrt{\omega\rho\mu / 2}$

c) acoustic compliance $C = A / \rho c^2$

d) heat conduction loss $G = S(\eta - 1 / \rho c^2)\sqrt{\lambda\omega / 2c_p\rho}$,

where the angular frequency $\omega = 2\pi f$, A is the tube area, S the tube circumference, ρ the air density, c the sound velocity, μ the viscosity coefficient, λ the coefficient of heat conduction, η the adiabatic constant, and c_p the specific heat of air at constant pressure. Flanagan [1972, pp. 28–35] reviews these analogous elements and gives the numerical value of the above constants: $\rho = 1.14 \times 10^{-3}\ \text{gm}/\text{cm}^3$, $c = 3.5 \times 10^4$ cm/sec, $\eta = 1.4$, $\mu = 1.86 \times 10^{-4}$ dyne-sec/cm^2, $\lambda = 0.055 \times 10^{-3}$ cal/cm-sec-deg, $c_p = 0.24$ cal/gm-degree.

A mass-compliance-viscous loss series circuit in parallel with C and G models the vocal tract wall impedance $Z_{wall} = [R_{wall} + j(\omega M - K / \omega)] / S$, where the resistance

$R_{wall} = 800\,\text{gm/sec}$, mass $M = 2.1\,\text{gm}$, and compliance $K = 84.5 \times 10^3\,\text{dyne/cm}$. These per-unit-area parameters of the relaxed cheek were determined experimentally by Ishizaka et al. [1975].

When the radius of the lip opening (tube 1) is small compared with the radius of the head, the radiation impedance Z_{rad} comes close to that of a circular piston mounted in an infinite plane baffle

$$Z_{rad} = \frac{\rho c}{\pi r^2}\left[1 - \frac{J_1(2kr)}{kr} + j\frac{H_1(2kr)}{kr}\right],$$

where k is the wave number ω / c and r the radius of tube 1; J_1 and H_1 are first-order Bessel and Struve functions [Aarts and Janssen, 2003]. The Fortran subroutines JY01A and STVH1 written by Zhang and Jin [1996, pp. 134–136, 347–348] are used to calculate the Bessel J_1 and Struve H_1 functions within the main Fortran program. Flanagan [1972, pp. 36–38] notes that for frequencies below 5 kHz and a lip opening area of 5 cm^2 or less (and hence $kr < 1$), the model of a piston in an infinite wall provides a fairly good approximation.

The loop volume velocities of the transmission line lattice are found by solving the system of simultaneous linear equations through matrix inversion [Mrayati and Guérin, 1976]. The driving volume velocity at the closed glottis (tube 27) is set to unity while the magnitude of the output volume velocity $U(0)$ is evaluated at the lip opening (tube 1). The frequency response of the system $20\log_{10} U(0)$ is calculated in 1 Hz steps from 12 Hz to 6502 Hz. The overall power is defined as

$$PWR = 10\log_{10}\sum_{i=12}^{6502} U_i^2(0).$$

After visual examination of the spectrum, the first four formant frequencies and their bandwidths are determined interactively by entering the most probable lower and upper bounds of the formants in Hertz. The program then searches within this range for the peak frequency f_p, as well as the low and high half-power frequencies f_1 and f_2 −3 dB below f_p. Once the spectral bandwidth of the formant is found $B = f_2 - f_1$, the logarithmic quality factor is calculated $\log Q = 20\log_{10}(f_p / B)$. The dimensionless $\log Q$ is a decibel measure of amplification at resonance [Kinsler and Frey, 1962, p. 195].

Another method of estimating bandwidths is through the use of the dissipated power/stored energy relation

$$B = \frac{dissipated\ power}{2\pi \times stored\ energy},$$

where B is the resonant frequency bandwidth in Hz [Fant and Pauli, 1975; Mrayati and Carré, 1976].

For a single tube, the kinetic energy is $E_{kin}(n) = L(n)[U^2(n) + U^2(n-1)]/2$, where $U^2(n)$ is the squared magnitude of the volume velocity of the loop nearer the glottis and $U^2(n-1)$ the squared magnitude of the volume velocity of the loop nearer the lips. In a similar manner, the potential energy is $E_{pot}(n) = C(n)p^2(n)$, where $p^2(n)$ is the squared magnitude of the pressure drop across the shunt elements of the tube. At resonance, the stored energy in the vocal tract is

$$E_{stor} = \sum_{n=1}^{27} E_{kin}(n) = \sum_{n=1}^{27} E_{pot}(n).$$

The power lost to viscosity in a single tube is $PR(n) = R(n)[U^2(n) + U^2(n-1)]/2$, whereas the power losses due to heat conduction and wall damping are $PG(n) = G(n)p^2(n)$ and $PG_{wall}(n) = G_{wall}(n)p^2(n)$. The power loss due to radiation is $PR_{rad} = R_{rad}U^2(0)$.

In view of the dissipated power/stored energy relation above, the formant bandwidth *B* may be expressed as

$$B = \frac{PR_{rad} + \sum_{n=1}^{27} PR(n) + \sum_{n=1}^{27} PG(n) + \sum_{n=1}^{27} PG_{wall}(n)}{2\pi \times E_{stor}},$$

or alternatively as

$$B_{sum} = B_{rad} + B_{visc} + B_{heat} + B_{wall},$$

where B_{rad}, B_{visc}, B_{heat}, B_{wall} are the bandwidths associated with power losses due to radiation, viscosity, heat conduction, and wall damping; B_{sum} is their sum.

The dissipated power/stored energy method of bandwidth determination indicates how much each type of power loss contributes to the total bandwidth. Thus it can serve to verify the spectral bandwidth measure. However because formant bandwidths calculated by this method are model-intrinsic and not directly observable in speech, only the spectral bandwidth $B = f_2 - f_1$ is suitable for determining the *Q* of actual speech signals.

3. Validation of the frequency-domain vocal tract model

In order to test the validity of the FDVT model, the formant frequencies F, bandwidths B, and amplitudes Amp ($= 20\log_{10} U(0)$) of a 4 cm^2 and a 1 cm^2 uniform pipe 18 cm in length are compared with those calculated by the Matlab-based VTAR program [Zhang and Espy-Wilson, 2004]. As there are 27 tubes, the elementary tube is 2/3 cm long. The VTAR constants associated with viscosity, heat conduction, and wall impedance are set identical to those of the FDVT program. VTAR provides an on-off switch for the radiation impedance, but the method of evaluating Z_{rad} is not documented.

The results of the simulations are presented in Tables 1.1 and 1.2. There is very close agreement between all the FDVT and VTAR acoustic parameters for both the 4 cm^2 pipe (Table 1.1) and the 1 cm^2 pipe (Table 1.2). For the FDVT model, a comparison between the spectral bandwidth B $(= f_2 - f_1)$ and the dissipated power/stored energy bandwidth B_{sum} demonstrates close agreement as well. Thus the validity of the FDVT model is confirmed satisfactorily. Tables 1.1 and 1.2 show that the radiation bandwidth B_{rad} grows progressively larger as the formant frequency increases, especially for the

Formant	f	Amp	B	B_{sum}	B_{rad}	B_{visc}	B_{heat}	B_{wall}
	Hz	dB	Hz	Hz	Hz	Hz	Hz	Hz
				FDVT 18 cm 4 cm^2 uniform pipe				
F1	493	30.98	15	17.11	3.23	4.48	1.99	7.40
F2	1398	24.80	33	36.17	24.34	7.55	3.35	0.93
F3	2324	18.12	70	73.98	59.59	9.74	4.32	0.34
F4	3259	13.53	115	116.70	99.88	11.53	5.12	0.17
				VTAR 18 cm 4 cm^{2}uniform pipe				
F1	493	30.98	17					
F2	1396	24.73	34					
F3	2316	17.94	71					
F4	3240	13.21	115					

Table 1.1. Acoustic parameters calculated by FDVT and VTAR models for a 4 cm^2 uniform pipe.

Formant	f	Amp	B	B_{sum}	B_{rad}	B_{visc}	B_{heat}	B_{wall}
	Hz	dB	Hz	Hz	Hz	Hz	Hz	Hz
				FDVT 18 cm 1 cm^2 uniform pipe				
F1	533	26.78	25	27.06	0.92	9.33	4.14	12.67
F2	1442	26.22	29	30.53	6.64	15.34	6.81	1.74
F3	2382	22.43	45	46.60	17.49	19.71	8.75	0.64
F4	3328	19.17	65	66.42	32.44	23.30	10.35	0.33
				VTAR 18 cm 1 cm^2 uniform pipe				
F1	533	26.78	25					
F2	1441	26.21	30					
F3	2379	22.39	46					
F4	3321	19.08	65					

Table 1.2. Acoustic parameters calculated by FDVT and VTAR models for a 1 cm^2 uniform pipe.

higher formants (F2–F4). The viscous and heat conduction bandwidths B_{visc} and B_{heat} likewise grow with increasing formant frequency, but far more slowly. These observations are in conformity with the analytical results for the uniform pipe, which indicate that B_{rad} is proportional to f^2 while B_{visc} and B_{heat} are proportional to $\sqrt{f}$ [Fant, 1960, p. 33, 307; Stevens, 1998, p. 155, 161]. Furthermore, Tables 1.1 and 1.2 show that within a given formant range, the radiation bandwidth B_{rad} is larger for the 4 cm^2 pipe than for the 1 cm^2 pipe whereas B_{visc} and B_{heat} are larger for the 1 cm^2 pipe than for the 4 cm^2 pipe. These relations are supported by the following analytical findings for the open-closed uniform pipe:

a) The radiation bandwidth B_{rad} is proportional to the pipe area A [Fant, 1960, p. 307; Stevens, 1998, p. 155].

b) The viscous bandwidth B_{visc} and heat conduction bandwidth B_{heat} are inversely proportional to A/S or $\sqrt{A}$ when the pipe is a cylinder of circumference S [Fant, 1960, p. 33; Stevens, 1998, p. 161].

4.1 Articulator regions in a uniform pipe model of the vocal tract

The volume velocity U and the pressure p both vary sinusoidally along the axis of a uniform pipe open at one end and closed at the other. The spatial volume velocity is a reciprocal function of the spatial pressure; hence the volume velocity minimum corresponds to a pressure maximum and vice versa. Chiba and Kajiyama [1958, p. 151] discovered that there was a systematic relationship between the location of a constriction in the vocal tract and the resulting changes in formant frequency:

> "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 current or of the excess pressure."

Further elaborating the perturbation analysis, Fant [1960, p. 86] states that:

> "A reduction of the tube cross-sectional area at the place of a volume velocity maximum is equal to the insertion of a lumped series inductance since the capacitance of the section may be neglected in view of the state of pressure minimum. If, on the other hand, a change in tube cross-sectional area is made near a volume velocity minimum, i.e., at the place of a pressure maximum, it is possible to disregard the distributed inductance at this place and take into consideration its capacitance only. The effect of the increased lumped inductance is to lower the resonance frequency and the effect of the decreased capacitance is to increase the resonance frequency."

As an illustration, Fant points out that the first formant (F1) has a volume velocity maximum at the lips and thus a constriction at the lips lowers the first formant frequency.

Based on the perturbation analysis of Fant and Pauli [1975], Story [2006] defines the tube sensitivity function as

$$S(n) = \frac{E_{kin}(n) - E_{pot}(n)}{E_{tot}},$$

where $E_{kin}(n)$ and $E_{pot}(n)$ are the kinetic and potential energies of the elementary tube n at a given formant frequency. The total energy of the N tubes of a pipe is

$$E_{tot} = \sum_{n=1}^{N} E_{kin}(n) + E_{pot}(n).$$

The general perturbation equation is

$$\frac{\Delta F}{F} = \sum_{n=1}^{N} S(n) \frac{\Delta A(n)}{A(n)},$$

where $\Delta A(n)/A(n)$ is the tube area change [Fant 1975; Mrayati and Carré, 1976]. A volume velocity maximum ($\max U$) is associated with a positive maximum of the sensitivity function ($\max S+$), and a pressure maximum ($\max p$) with a negative maximum of the sensitivity function ($\max S-$). Formant frequency and area increases are positive changes ($\Delta F/F+$, $\Delta A/A+$); formant frequency and area decreases are negative changes ($\Delta F/F-$, $\Delta A/A-$). Four combinations may be distinguished:

(a) $\max U$ ($\max S+$) × area decrease ($\Delta A/A-$) → formant decrease ($\Delta F/F-$)
(b) $\max U$ ($\max S+$) × area increase ($\Delta A/A+$) → formant increase ($\Delta F/F+$)
(c) $\max p$ ($\max S-$) × area decrease ($\Delta A/A-$) → formant increase ($\Delta F/F+$)
(d) $\max p$ ($\max S-$) × area increase ($\Delta A/A+$) → formant decrease ($\Delta F/F-$)

For example, a contraction (a) or an expansion (b) at the lips (volume velocity maximum) lowers or raises F1 respectively whereas a contraction (c) or an expansion (d) at the glottis (pressure maximum) raises or lowers F1 respectively.

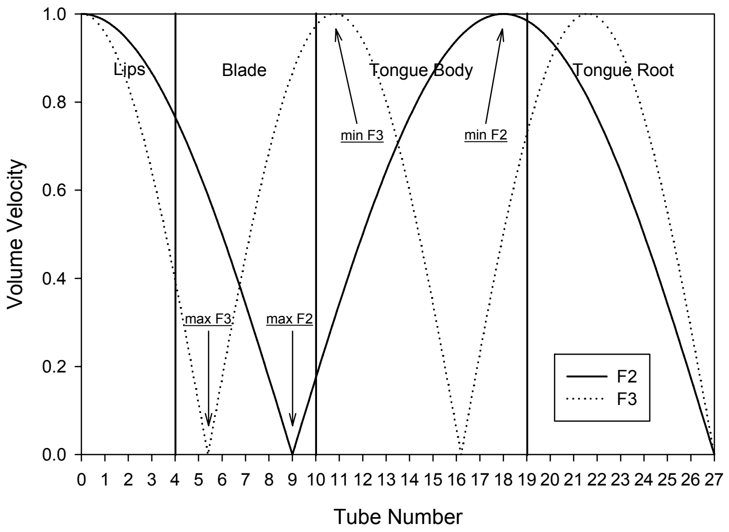

Figure 1. Articulator regions and spatial volume velocity magnitudes of the second and third formants (F2, F3) in a lossless open-closed uniform pipe. The arrows indicate the constriction locations of the highest and lowest second formant frequencies max F2, min F2 as well as the highest and lowest third formant frequencies max F3, min F3.

Figure 1 displays the spatial volume velocity magnitudes of the second and third formants in a lossless open-closed uniform pipe. The second formant (F2) shows a volume velocity maximum at one-third the vocal tract length from the glottis and a volume velocity minimum at two-thirds the overall length from the glottis. Likewise, the third formant (F3) shows a volume velocity maximum at three-fifths the vocal tract length from the glottis and a volume velocity minimum at four-fifths the overall length from the glottis. As the constriction locations of F2 and F3 move toward the lips from the volume velocity maximum (pressure minimum) to the volume velocity minimum (pressure maximum), the formant frequencies should shift from their lowest values to their highest values in agreement with perturbation analysis. An examination of Fant's nomograms [1960, Figures 1.4-9 and 1.4-11] reveals that F2 and F3 are raised as expected when the constrictions are moved forward through their respective one-quarter wavelengths [cf. also Badin et al., 1990]. The range of formant frequencies in single-constriction nomograms is not unlike that of actual speech. Fant found second formants from 0.5 to 2.5 kHz and third formants from 1.8 to 3.2 kHz. This is strong evidence that a constriction whose position is varied between two limits—one-third and two-thirds the overall vocal tract length from the glottis—can generate the F2 frequencies observed in speech. In a similar manner, a constriction whose position is varied between three-fifths and four-fifths the total length from the glottis can readily produce representative F3 frequencies. These two quarter-wavelength regions are anatomically co-extensive with the tongue body and the blade, respectively.

The vocal tract model of 27 equal-length tubes is partitioned into four regions, each of which corresponds to an active articulator: the lips, the blade, the tongue body, and the tongue root. The lip region consists of 4 tubes (No. 1–4), the blade region 6 tubes

(No. 5–10), the tongue body region 9 tubes (No. 11–19), and the tongue root region 8 tubes (No. 20–27). It is useful for the ensuing discussion to introduce a convenient reference length of 18 cm for the 27-tube vocal tract. Each tube is therefore 0.67 cm long. For the third formant, the one-quarter wavelength between three-fifths and four-fifths 18 cm from the glottis is 3.60 cm (= 1/5 × 18 cm). This F3 quarter wavelength is modeled by a six-tube blade region 4 cm long (= 6/27 × 18 cm). If instead the interval were approximated by five tubes 3.33 cm long (= 5/27 × 18 cm), the blade would most likely be too short. Keating [1991, p. 31], for example, gives an upper bound of 3 to 4 cm for blade length. The entire six-tube blade region is then advanced from the F3 volume velocity maximum in order to avoid excessive overlap with the F2 volume velocity minimum. As a result, the lip region consists of four tubes and is 2.67 cm long (= 4/27 × 18 cm). For the second formant, the one-quarter wavelength between one-third and two-thirds 18 cm from the glottis is 6 cm (= 1/3 × 18 cm). The F2 quarter wavelength is modeled by the exact interval—a nine-tube tongue body region 6 cm in length (= 9/27 × 18 cm). However the entire tongue body region is shifted one tube back from the F2 volume velocity minimum so that the posterior blade remains close to the F3 volume velocity maximum. Hence the tongue root region is made up of eight tubes rather than nine and is 5.33 cm long (= 8/27 × 18 cm). The elementary tubes of the lip (4) and tongue root (8) regions appear to be those remaining by default once the quarter-wavelength tubes of the blade (6) and the tongue body (9) are determined.

4.2 Articulator regions in a non-uniform pipe model of the vocal tract

The above partition of the vocal tract into four articulator regions presupposes that the locations of the volume velocity maxima and minima of speech sounds do not deviate markedly from those of the lossless uniform pipe open at one end and closed at the glottis. The closed glottis condition holds when the glottal area is appreciably smaller than the area of the open end, thereby sustaining the odd resonance modes at frequencies on the order of three (F2), five (F3), and seven (F4) times the fundamental resonance frequency (F1) of the vocal tract. A peak glottal area from 0.05 to 0.2 cm^2 is typical of adult voiced sounds whereas glottal areas between 0.1 and 0.4 cm^2 are characteristic of voiceless sounds [Stevens, 1998, p. 35]. Using an analysis-by-synthesis procedure, Badin [1989] found the resonance frequencies of [ʃ] to be in the neighborhood of 430 Hz (F1), 1750 Hz (F2), 2680 Hz (F3), and 3200 Hz (F4) with the glottal areas set at both 0.1 and 0.25 cm^2. Clearly, the pattern of odd resonance frequencies is preserved even for the relatively large glottal opening of 0.25 cm^2. On the basis of X-ray area functions, Mrayati and Carré [1976] computed the damped volume velocities of F1, F2, and F3 of eleven French vowels modeled as non-uniform pipes. Although the volume velocity magnitudes along the vocal tract often differed considerably among the synthetic vowels, the locations of the volume velocity maxima and minima shifted little with respect to their locations in the uniform pipe. In sum, the partition of the vocal tract into the four articulator regions appears to be valid for speech sounds because a) the closed glottis condition is nearly always met and b) the locations of the volume velocity maxima and minima of a lossy non-uniform pipe largely coincide with those of the uniform pipe in Figure 1.

4.3 Articulator regions: boundaries

Determination of the anatomical boundaries of the four oral articulators (lips, blade, tongue body, tongue root) has been a long-standing problem in phonetic science. The positions of the three quasi-independently controllable lingual articulators (blade, tongue body, tongue root) are traditionally defined with reference to a passive articulator along the roof of the mouth and pharynx, where "it is normally assumed that the sound at a named place of articulation is made by the articulator lying opposite the place of articulation…" [Handbook of the IPA, 1999, p. 8]. For example, the IPA chart distinguishes three blade positions (passive articulators: front teeth, alveolar ridge, postalveolar area) and three tongue body positions (passive articulators: hard palate, velum, uvula). Among the lingual articulators, the only clear anatomical boundary is the edge of the uvula separating the tongue body from the tongue root. There is no corresponding sharp boundary between the postalveolar area and the hard palate that separates the blade from the tongue body.

In the preceding discussion it was shown that the vocal tract and the open-closed uniform pipe possess a similar spatial distribution of odd resonance modes. Moreover, the blade region spans approximately an F3 quarter wavelength, and the tongue body region an F2 quarter wavelength. It seems reasonable to conclude, therefore, that the blade and tongue body regions are delimited by the acoustic partitions illustrated in Figure 1, and not by the anatomical particularities of the vocal tract opposite the active articulators. As a rough analogy, consider the complete stopping of an unfretted violin string. It is of no acoustic importance which finger shortens the length of the string. The pitch of the note is regulated by varying the target location on the fingerboard. Likewise, the passive articulators forming the constrictions of the blade (dental, alveolar, postalveolar) and the tongue body (palatal, velar, uvular) are not acoustically pertinent in themselves. The F3 and F2 frequencies are controlled by changing the target locations of the lingual constrictions relative to the volume velocity maxima and minima along the vocal tract.

Because the edge of the uvula creates a distinct border dividing the tongue body and tongue root, it is possible to compare two ratios of tongue root length to total vocal tract length, one anatomical, the other acoustical. Fitch and Giedd [1999] used MRI imaging to estimate certain dimensions of the vocal tract, including the mean length of the pharynx (from the uvula to the glottis) and the total vocal tract. The ratios of pharynx to vocal tract length, expressed as percentages, are as follows:

men	19–25	37.47%
women	19–25	32.04%
children	11–12	32.42%
children	7–8	29.69%
children	2–4	25.39%

With the exception of men and the youngest children, the MRI ratios of pharynx to vocal tract length match fairly well the acoustical ratio of 8/27 or 29.63% obtained in Section

Study	Year	Method	Age	Sex	Language	Vowels
Fant	1960	X-ray	adult	male	Russian	6
Mrayati and Guérin	1976	X-ray	adult	male	French	11
Baer et al., TB	1991	MRI	adult	male	English	4
Baer et al., PN	1991	MRI	adult	male	English	4
Yang and Kasuya	1994	MRI	adult	male	Japanese	5
Yang and Kasuya	1994	MRI	adult	female	Japanese	5
Yang and Kasuya	1994	MRI	11	male	Japanese	5
Story et al.	1996	MRI	adult	male	English	10
Story et al.	1998	MRI	adult	female	English	10
Takemoto et al.	2006	MRI	adult	male	Japanese	5

Table 2. Seven X-ray and MRI studies with vowel area functions from ten speakers.

4.1. Thus for women and older children, the acoustic partition between the tongue body and the tongue root tends to coincide with the edge of the uvula. On the other hand, the acoustic partition is somewhat anterior to the uvular edge for children 2 to 4 years old and is significantly posterior to it for men. Although an acoustic partition may coincide with a natural anatomical boundary, as is observed for women and older children, the results for men and the youngest children show that no such relationship need exist [see Turner et al., 2009, p. 2379 who reach the same conclusion].

An eight-region vocal tract simulation has been proposed which is likewise founded on perturbation analysis [Mrayati et al., 1988; Carré, 2004]. However, it can not constitute a realistic model of speech production because the four phonologically relevant articulators (lips, blade, tongue body, tongue root) are not taken into account as such [cf. the remarks by Boë and Perrier, 1990, p. 227].

5. Vowel area functions

Seven X-ray and MRI papers containing tabulated vowel area functions are listed in Table 2, for a total of ten speakers. The vowel area functions were fit to 27 equal-length tubes by means of cubic spline interpolation. The Matlab command *spline* performed the operation after which the original and the interpolated area functions were compared graphically. Occasionally the interpolated function would overshoot or undershoot the original function. When this occurred, the outlier was replaced by the nearest original value. For the first and second formants, the original and the interpolated area functions always showed negligible frequency differences when both functions were calculated by the Matlab-based VTAR program. On the other hand, the third and fourth formants often displayed a good deal of sensitivity to deviations from the original area function.

6. Articulatory parameters

Using the interpolated area functions, the FVDT program calculates the eleven articulatory parameters presented in Table 3 (see Figure 1 for the tube numbers).

1) The lip length parameter $L(lip)$ is the sum of tube lengths in the lip region from tube 1 to 4 (lip protrusion).

Length L	Minimum Area Index $I_{\min A}$	Mean Area $\overline{A}$	Area Ratio
	$I_{\min A}(root)$	$\log_2 \overline{A}(root)$	$\log_2 \overline{A}(root) / \overline{A}(lip)$
	$I_{\min A}(body)$	$\log_2 \overline{A}(body)$	$\log_2 \overline{A}(body) / \overline{A}(lip)$
	$I_{\min A}(blade)$	$\log_2 \overline{A}(blade)$	$\log_2 \overline{A}(blade) / \overline{A}(lip)$
$L(lip)$		$\log_2 \overline{A}(lip)$	

Table 3. The eleven articulatory parameters. $L(lip)$ is the sum of tube lengths in the lip region (lip protrusion). The minimum area index $I_{\min A}$ is the integer index of the tube with the smallest constriction in a given articulator region.

Power	First Formant	Second Formant	Third Formant	Fourth Formant
PWR (dB)	$\log_2 F1$	$\log_2 F2$	$\log_2 F3$	$\log_2 F4$
	$B1$ (Hz)	$B2$ (Hz)	$B3$ (Hz)	$B4$ (Hz)
	$20\log_{10} Q1$ (dB)	$20\log_{10} Q2$ (dB)	$20\log_{10} Q3$ (dB)	$20\log_{10} Q4$ (dB)

Table 4. The thirteen acoustic parameters.

2) The minimum area index $I_{\min A}$ is the integer index of the tube with the smallest constriction in a given articulator region. Consequently, $I_{\min A}(blade)$ indicates the blade position, $I_{\min A}(body)$ the tongue body position, and $I_{\min A}(root)$ the tongue root position. The blade range is $6 \le I_{\min A}(blade) \le 1$ (from tube 5 to 10), the tongue body range is $9 \le I_{\min A}(body) \le 1$ (from tube 11 to 19), and the tongue root range is $8 \le I_{\min A}(root) \le 1$ (from tube 20 to 27).

3) The mean area $\overline{A}$ is the length-weighted mean of the tube areas $A(n)$ in a given articulator region

$$\overline{A} = \frac{\sum_{n}^{m} l(n)A(n)}{\sum_{n}^{m} l(n)}$$

The tube numbers for the mean lip area $\overline{A}(lip)$ are $n = 1$, $m = 4$, those of the mean blade area $\overline{A}(blade)$ $n = 5$, $m = 10$, those of the mean body area $\overline{A}(body)$ $n = 11$, $m = 19$, and those of the mean root area $\overline{A}(root)$ $n = 20$, $m = 27$. When the tube length l remains constant, the mean area $\overline{A}$ is simply the sum of the tube areas $A(n)$ in the articulator region divided by the number of tubes. The mean area $\overline{A}$ is converted to its base-2 logarithm $\log_2 \overline{A}$. Hence an area doubling leads to a unit increase in $\log_2 \overline{A}$.

First Formant and Power Correlations: ranked means and standard deviations

Acoustic Parameter	Rank	Articulatory Parameter	mean	s.d.
$\log F1$	1	$\log\overline{A}(root)/\overline{A}(lip)$	–0.935	0.063
$\log F1$	2	$\log\overline{A}(root)$	–0.915	0.071
$\log F1$	3	$\log\overline{A}(lip)$	0.723	0.224
$\log F1$	4	$I_{\min A}(root)$	0.545	0.213
$\log Q1$	1	$\log\overline{A}(root)/\overline{A}(lip)$	–0.849	0.121
$\log Q1$	2	$\log\overline{A}(root)$	–0.841	0.110
$\log Q1$	3	$\log\overline{A}(blade)$	0.687	0.180
$B1$	1	$\overline{A}(root)$	0.418	0.403
$B1$	2	$\overline{A}(blade)$	–0.300	0.521
$B1$	3	$\overline{A}(blade)/\overline{A}(lip)$	–0.293	0.479
PWR	1	$\log\overline{A}(root)/\overline{A}(lip)$	–0.844	0.097
PWR	2	$\log\overline{A}(root)$	–0.768	0.113
PWR	3	$\log\overline{A}(lip)$	0.710	0.201

Table 5.1. First formant and power correlations. The acoustic-articulatory correlations of each speaker's vowel system are calculated, then the cross-speaker mean and its standard deviation (s.d.) is found and ranked (N = 10).

4) The three area ratios are the mean areas of the blade, tongue body, and tongue root regions normalized by the mean lip area: $\overline{A}(blade)/\overline{A}(lip)$, $\overline{A}(body)/\overline{A}(lip)$, and $\overline{A}(root)/\overline{A}(lip)$. These area ratios are likewise transformed into their base-2 logarithms.

7. Acoustic parameters

The thirteen acoustic parameters are shown in Table 4. The first four formant frequencies (F1–F4), bandwidths (B1–B4), quality factors (Q1–Q4), and the overall power (PWR) are determined as set forth in Section 2. The formant frequency F is converted to $\log_2 F$ in order to account for the log frequency (octave) scales of pitch and formant perception [Miller, 1989]. As mentioned earlier, log Q ($= 20\log_{10} Q$) is a decibel measure of amplification at resonance.

8.1 Acoustic-articulatory correlations: first formant and power

To estimate the strength of association between the articulatory and acoustic parameters in Tables 3 and 4, Pearson correlation matrices are calculated for the vowel system of each speaker. Then the coefficients of the parameter pairs are averaged across the ten speakers.

In Table 5.1, the mean coefficients of the first formant and power parameters are ranked according to size. The frequency parameter log F1 exhibits the largest correlation coefficient with the lip-normalized area ratio $\log\overline{A}(root)/\overline{A}(lip)$ (r = –0.935), followed by $\log \overline{A}(root)$ (r = –0.915) and then $\log \overline{A}(lip)$ (r = 0.723). The tongue root area $\log \overline{A}(root)$ manifestly dominates the area ratio $\log \overline{A}(root) / \overline{A}(lip)$. The trading relation between the lip and tongue root areas—as indicated by the ratio $\log \overline{A}(root) / \overline{A}(lip)$—conforms to perturbation analysis since a wider lip or a narrower tongue root area increases F1 while a narrower lip or a wider tongue root area decreases F1.

The quality factor log Q1 patterns after log F1, the largest coefficient being associated with $\log \overline{A}(root) / \overline{A}(lip)$, and the next largest with $\log \overline{A}(root)$. The close correspondence between log Q1 and log F1 is consistent with the magnitude of the transfer function at resonance

$$|H(F)| = Q = \frac{F}{B},$$

which predicts a linear relation between Q and F on the assumption that B remains constant [Fant, 1960, p. 54]. To detect a possible linear increase of log Q1 with log F1, a linear regression analysis is applied to each vowel system. The mean slope is found to be 7.62 dB/octave, not greatly different from the linear slope of 6 dB/octave. Because wall losses diminish with increasing frequency, B1 becomes smaller as F1 increases [Flanagan, 1972, p. 69]. Thus the observed mean slope grows somewhat faster than 6 dB/octave.

The overall power PWR also behaves similarly to log F1, with the largest coefficient corresponding to $\log \overline{A}(root) / \overline{A}(lip)$, the second largest to $\log \overline{A}(root)$, and the third largest to $\log \overline{A}(lip)$. The clear parallel between PWR and log F1 suggests that most of the intrinsic power of vowels is concentrated in the first formant and is therefore nearly equal to $\log|H(F1)|$ $(= 20\log_{10}|H(F1)|)$, or equivalently $\log Q1$ $(= 20\log_{10} Q1)$. A regression analysis of PWR and log F1 yields a mean slope of 8.17 dB/octave, which is reasonably comparable to the slope of log Q1 and log F1 above.

Fairbanks [1950] correlated the average intensities of eleven vowels uttered by ten Midwestern speakers with X-ray measurements of the same vowels produced by a separate speaker from the Middle West. The correlation coefficients were 0.95 for the incisors, 0.65 for the lips, and 0.36 for the tongue-palate. Despite methodological differences and the absence of tongue root measures, the results confirm the direct involvement of the lip region (incisors, lips) in determining the intrinsic intensity of vowels.

The relatively small coefficients of B1 show that bandwidth is a poor acoustic cue for signaling articulatory relationships. Poor results are also seen for B2, and to a lesser extent, B3 and B4 (Tables 5.2–5.4). Hence raw bandwidth does not represent a useful acoustic parameter.

To conclude this section, observe that log F1 is positively correlated (r = 0.545) with the tongue root position $I_{\min A}(root)$. Thus as F1 frequency increases, the smallest constriction tends to be located higher in the pharynx. The tendency for low vowels to be accompanied by a high tongue root position, and conversely, was also noted during the course of the data analysis.

Second Formant Correlations: ranked means and standard deviations				
Acoustic		Articulatory		
Parameter	Rank	Parameter	mean	s.d.
log *F2*	1	$I_{\min A}(body)$	0.617	0.280
log *F2*	2	$L(lip)$	–0.531	0.462
log *F2*	3	$I_{\min A}(blade)$	0.393	0.626
log *Q2*	1	$\log \overline{A}(body)/\overline{A}(lip)$	0.607	0.294
log *Q2*	2	$\log \overline{A}(lip)$	–0.580	0.235
log *Q2*	3	$\log \overline{A}(lip)/\overline{A}(root)$	–0.415	0.380
B2	1	$\overline{A}(blade)/\overline{A}(lip)$	–0.624	0.304
B2	2	$\overline{A}(lip)$	0.388	0.312
B2	3	$\overline{A}(blade)$	–0.352	0.420

Table 5.2. Second formant correlations.

Third Formant Correlations: ranked means and standard deviations				
Acoustic		Articulatory		
Parameter	Rank	Parameter	mean	s.d.
log *F3*	1	$L(lip)$	–0.439	0.438
log *F3*	2	$I_{\min A}(blade)$	0.400	0.314
log *F3*	3	$I_{\min A}(body)$	0.021	0.492
log *Q3*	1	$\log \overline{A}(blade)/\overline{A}(lip)$	0.611	0.231
log *Q3*	2	$\log \overline{A}(lip)$	–0.499	0.324
log *Q3*	3	$\log \overline{A}(body)$	–0.447	0.308
B3	1	$\overline{A}(blade)/\overline{A}(lip)$	–0.546	0.176
B3	2	$\overline{A}(body)$	0.523	0.284
B3	3	$\overline{A}(lip)$	0.417	0.470

Table 5.3. Third formant correlations.

8.2 Acoustic-articulatory correlations: second and third formants

The mean correlation coefficients for the second and third formants are presented in Tables 5.2 and 5.3. Recall from Section 6 that $I_{\min A}(body)$ and $I_{\min A}(blade)$ are the tongue body and blade indices of the tube with the smallest constriction in the articulator region. The tongue body and blade ranges are $9 \le I_{\min A}(body) \le 1$ and $6 \le I_{\min A}(blade) \le 1$, with the larger integer near the lips and the smaller near the glottis. Summarizing the known correspondences between formant frequencies and articulator positions, Fant [1960, p. 26] stated that a very low or high F2 formant frequency indicates either a retracted articulation or a palatal position of the tongue. As expected, there is a positive correlation (r = 0.617) between log F2 and the tongue body position $I_{\min A}(body)$. He also observed that a very low or high F3 formant frequency signals either a retroflex modification or a prepalatal/dental articulation. Accordingly, there is a positive correlation (r = 0.400) between log F3 and the blade position $I_{\min A}(blade)$. Thus when $I_{\min A}(body)$ and $I_{\min A}(blade)$ increase from 1 to 9 and from 1 to 6, so do the F2 and F3

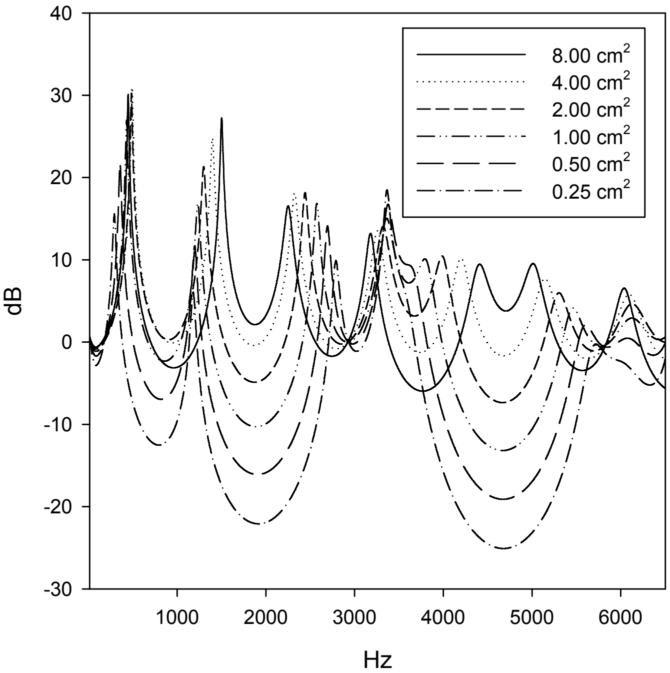

Figure 2.1. Spectra of a 4 cm^2 uniform pipe 18 cm in length, with the tongue body region (tubes 11–19) varying from 8 cm^2 to 0.25 cm^2.

frequencies in agreement with the perturbation analysis of Section 4.1. Note, however, that log F3 and lip length $L(lip)$ achieve a correlation coefficient similar in magnitude (r = –0.439) to that between log F3 and blade position, indicating that lip protrusion may lower F3 as much as blade advancement raises it. Lip protrusion likewise lowers F2, though to a lesser degree on account of the larger correlation between log F2 and tongue body position $I_{\min A}(body)$. Therefore in order to achieve their F3 and F2 targets, the blade and tongue body positions must compensate for changes in the lip length $L(lip)$.

The quality factors log Q2 and log Q3 display the largest correlations with the respective lip-normalized area ratios $\log \overline{A}(body)/\overline{A}(lip)$ and $\log \overline{A}(blade)/\overline{A}(lip)$. To determine how log Q2 and log Q3 change with the area ratios, regression analyses are performed. The mean log Q2 slope is 2.54 dB per doubling of $\overline{A}(body)/\overline{A}(lip)$ while the mean log Q3 slope is 3.85 dB per doubling of $\overline{A}(blade)/\overline{A}(lip)$.

The validity of these observations is tested by means of a simple two-pipe model. The areas of the tongue body or blade region are varied in an otherwise uniform pipe 4 cm^2 in area and 18 cm long. The six areas of the tongue body pipe (tubes 11–19) or the blade pipe (tubes 5–10) are the following: 8 cm^2, 4 cm^2, 2 cm^2, 1 cm^2, 0.5 cm^2, and 0.25 cm^2. The frequency responses resulting from the area variations of the tongue body and blade pipes are plotted in Figures 2.1 and 2.2. As the area of the lip region remains fixed at 4 cm^2, both spectra demonstrate the expected lowering of the formant intensities $\log|H(F2)|$ and $\log|H(F3)|$ with increasingly smaller articulator areas. However the third formant in Figure 2.2 shows a dramatic decline by comparison to the second

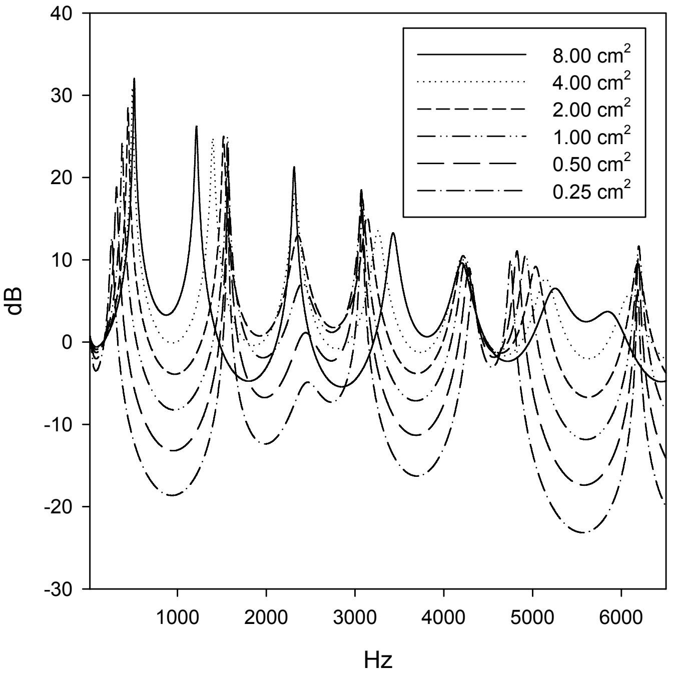

Figure 2.2. Spectra of a 4 cm^2 uniform pipe 18 cm in length, with the blade region (tubes 5–10) varying from 8 cm^2 to 0.25 cm^2.

formant in Figure 2.1. In addition, the attenuation of the third formant is accompanied by strikingly few changes in the other formants even when the blade area is only 0.25 cm^2. To investigate the two-pipe model further, regression slopes for the second and third formants are calculated. The log Q2 slope is 1.62 dB per doubling of $\overline{A}(body)/\overline{A}(lip)$ and the log Q3 slope is 4.45 dB per doubling of $\overline{A}(blade)/\overline{A}(lip)$. The steeper slope of log Q3 is in accordance with the wide range of third formant intensities seen in Figure 2.2. As pointed out in Section 3, radiation bandwidth is proportional to the area of a uniform pipe. This suggests that radiation loss is a plausible mechanism for the behavior of Q2 and Q3. To determine whether the slopes of log Q2 and log Q3 are mainly due to radiation damping, the analysis of the two-pipe model is conducted without radiation loss. The log Q2 slope is then 0.68 dB per doubling of $\overline{A}(body)/\overline{A}(lip)$ while the log Q3 slope is 0.19 dB per doubling of $\overline{A}(blade)/\overline{A}(lip)$. In view of these negligible values, both slopes must be governed by radiation damping.

8.3 Acoustic-articulatory correlations: fourth formant

The mean correlation coefficients for the fourth formant are given in Table 5.4. The frequency parameter log F4 and lip length $L(lip)$ yield the correlation coefficient with the largest magnitude (r = –0.388), the negative sign indicating that F4 frequency decreases as lip length increases and vice versa. Changes in lip length also cause inversely proportional shifts of the other formant frequencies, F2 and F3 in particular (cf. Tables 5.2–5.3). However, unlike them, the F4 frequency is nearly free from the confounding effects of the lingual articulators. For example, the correlation coefficients

Fourth Formant Correlations: ranked means and standard deviations

Acoustic		Articulatory		
Parameter	Rank	Parameter	mean	s.d.
$\log F4$	1	$L(lip)$	–0.388	0.293
$\log F4$	2	$I_{\min A}(body)$	0.165	0.409
$\log F4$	3	$I_{\min A}(blade)$	0.010	0.422
$\log Q4$	1	$\log \overline{A}(lip)$	–0.619	0.252
$\log Q4$	2	$\log \overline{A}(lip)/\overline{A}(root)$	–0.584	0.259
$\log Q4$	3	$\log \overline{A}(blade)/\overline{A}(lip)$	0.400	0.375
$B4$	1	$\overline{A}(lip)$	0.633	0.359
$B4$	2	$\overline{A}(lip)/\overline{A}(root)$	0.528	0.411
$B4$	3	$\overline{A}(blade)/\overline{A}(lip)$	–0.388	0.165

Table 5.4. Fourth formant correlations.

of log F4 with the positions of the tongue body $I_{\min A}(body)$ and blade $I_{\min A}(blade)$ are only 0.165 and 0.01 (both s.d. > 0.4). As a consequence, the F4 frequency furnishes the optimal cue for the lip length parameter $L(lip)$.

The range of lip protrusion can be tentatively estimated by subtracting the mean of the shortest vocal tracts (vowels with drawn lips) from the mean of the longest vocal tracts (vowels with protruded lips) for the seven men in Table 2. The mean of the shortest vocal tracts is 16.43 cm, that of the longest vocal tracts is 18.64 cm. Thus the adult male range of presumed lip protrusion is on average 2.21 cm (s.d. 0.82). However the simplifying assumption that lip protrusion is responsible for all changes in vocal tract length is not warranted. Raising and lowering the larynx can also shorten and lengthen the vocal tract [Ewan and Krones, 1974]. In vocal tract X-ray tracings collected from twelve languages, Wood (1979) observed that the larynx tends to be lower for protruded [u o y ø]-like vowels than for drawn [ɨ ɤ i ɛ]-like vowels. In his own X-ray investigation of two male speakers (Southern British English and Egyptian Arabic), he found a maximum difference of about 1 cm between protruded and drawn vowels. Hoole and Kroos [1998] point out that there can be substantial interspeaker variability in larynx height. Two of their three male subjects showed a maximum difference of 0.7–1.0 cm between German protruded and drawn vowels, yet the other displayed only a 0.2 cm difference. Riordan [1977, Fig. 5] obtained a similarly small maximum difference in two of four French men. Remark that the correlation between log F4 and $L(lip)$ (the sum of tube lengths in the lip region) is exactly the same as the correlation between log F4 and the summed tube lengths of any other region—including the entire length of the vocal tract. This reflects the fact that the correlation coefficient is invariant under linear transformations such as proportional length changes. Because F4 frequency is identically correlated with both lip length $L(lip)$ and total vocal tract length, larynx height must also have an acoustic effect. Nevertheless, lip protrusion will be considered the main determinant of vocal tract length since the lips constitute the most visible active articulator [for a review of the lips in visual speech perception, see Rosenblum, 2008; cf.

also Chuenwattanapranithi et al., 2008 for the signaling function of lip protrusion from an ethological perspective].

Although log F4 and log F3 attain strong relative correlations with lip length $L(lip)$ and blade position $I_{\min A}(blade)$, respectively, the absolute values are modest in themselves ($\mathrm{r} \approx |0.4|$). These moderate correlations are unanticipated because: a) F4, like the other formant frequencies, is inversely proportional to the length of a uniform pipe; b) F3 and blade position should show the same degree of association as F2 and tongue body position in light of perturbation analysis. In Section 5 it was pointed out that the F3 and F4 frequencies often exhibit high sensitivity to small deviations of the vowel area function whereas the F1 and F2 frequencies do not. As a result, the large-scale changes of the area function needed to control F1 and F2 may also give rise to unpredictable shifts in F3 and F4. Extrinsic noise sources include the original measurement error and the cubic spline interpolation of the present study.

The quality factor log Q4 reveals an inverse correlation (r = –0.619) with the lip area $\log \overline{A}(lip)$. The mean of the log Q4 regression slope is –5.67 dB per doubling of $\overline{A}(lip)$. This value approaches the slope of –6 dB per doubling of area that results from the strict proportionality between radiation bandwidth and the area of a uniform pipe (Section 3). Hence the slope of log Q4 is governed by radiation damping like the quality factors Q2 and Q3 in Section 8.2.

8.4 Acoustic-articulatory correlations: summary and discussion

It is convenient to characterize an articulatory configuration by its motor dimension (position, aperture) as opposed to rather unwieldy areas or area ratios. Except for tongue root aperture which they did not discuss, the remaining motor dimensions are analogous to the six oral tract variables proposed by Browman and Goldstein [1989]. Paired with the corresponding motor dimension, the most strongly correlated acoustic and articulatory parameters are the following:

1. Tongue root aperture $\log \overline{A}(root)/\overline{A}(lip) \;\propto \log 1/F1$ (the tongue root area normalized by lip area is inversely correlated with F1 frequency).
2. Tongue body position $I_{\min A}(body) \;\propto \log F2$ (as the location of the smallest tongue body constriction moves toward the lips, F2 frequency also shifts higher).
3. Tongue body aperture $\log \overline{A}(body)/\overline{A}(lip) \;\propto \log Q2$ (the tongue body area normalized by lip area is directly correlated with Q2).
4. Blade position $I_{\min A}(blade) \;\propto \log F3$ (as the location of the smallest blade constriction moves toward the lips, F3 frequency also shifts higher).
5. Blade aperture $\log \overline{A}(blade)/\overline{A}(lip) \;\propto \log Q3$ (the blade area normalized by lip area is directly correlated with Q3).
6. Lip position $L(lip) \;\propto \log 1/F4$ (the sum of tube lengths in the lip region displays an inverse correlation with F4 frequency).

7. Lip aperture $\log \overline{A}(lip) \propto \log F1, \propto \log 1/Q4$ (the lip area has a moderate direct correlation with F1 frequency and a weaker inverse correlation with Q4).

The first formant frequency shows the largest correlation (r = –0.935) with the tongue root aperture $\log \overline{A}(root)/\overline{A}(lip)$. The tongue root area $\overline{A}(root)$ is the dominant term in the area ratio because the correlation coefficient of log F1 and the tongue root area (r = –0.915) nearly equals the coefficient of $\log \overline{A}(root)/\overline{A}(lip)$ itself. Furthermore, the correlation magnitude of the tongue root area is notably larger than the coefficient of log F1 and lip area (r = 0.723). It was mentioned at the end of Section 8.1 that low vowels favor a high tongue root position and vice versa. This is evidence that the entire tongue root region participates in the control of F1 frequency. The greater weight of the tongue root region relative to the lip region follows from the perturbation analysis of the first formant. To illustrate, the F1 of an unmodified 4 cm^2 uniform pipe 18 cm long is 493 Hz. When the area of the four-tube lip region is decreased to half the uniform value (2 cm^2), then F1 is lowered to 439 Hz. Also when the areas of the four tubes nearest the glottis are increased to double the uniform value (8 cm^2), F1 is similarly lowered to 439 Hz. The results are in agreement with the perturbation analysis outlined in Section 4.1. But when the area of the entire eight-tube tongue root region is increased to double its uniform value (8 cm^2), then F1 is reduced even further to 407 Hz. The additional frequency drop occurs because the mean area of the tongue root region with eight expanded tubes is larger than the mean area with four expanded tubes in the vicinity of the F1 pressure maximum. Thus for a given area perturbation, the eight-tube tongue root region has a greater influence on the first formant than the four-tube lip region.

The tongue body position $I_{\min A}(body)$ and the blade position $I_{\min A}(blade)$ are associated with log F2 and log F3. However to attain their F2 and F3 targets, the tongue body and blade positions must compensate for the varying lip length $L(lip)$ which also affects F2 and especially F3 (cf. Section 8.2 above). The tongue body aperture $\log \overline{A}(body)/\overline{A}(lip)$ and the blade aperture $\log \overline{A}(blade)/\overline{A}(lip)$ correspond respectively to log Q2 and log Q3. To achieve their Q2 and Q3 targets, the tongue body and blade apertures must similarly counterbalance changes in lip area $\overline{A}(lip)$ because this term appears in the denominators of $\log \overline{A}(body)/\overline{A}(lip)$ and $\log \overline{A}(blade)/\overline{A}(lip)$. In general then, the positions and apertures of the tongue body and the blade need to be adjusted for both lip position and aperture in order to reach specific acoustic goals.

The lip position $L(lip)$ is best correlated with log F4 as discussed in Section 8.3. The lip aperture $\log \overline{A}(lip)$ has a moderate direct correlation with log F1 (r = 0.723) and a weaker inverse correlation with log Q4 (r = –0.619). To estimate the combined effect of the two acoustic correlates, a multiple linear regression analysis is carried out using the dependent variable $\log \overline{A}(lip)$ and the two independent variables log F1 and log Q4. The combined mean coefficient is 0.843, which represents a net improvement over log F1 and log Q4 alone.

9. Vowel height

In 1867 A. M. Bell introduced a vowel classification consisting of three primary height distinctions (*high*, *mid*, *low*) and three primary frontness distinctions (*front*, *central*, *back*) [for a review, see Catford, 1981]. Three secondary height distinctions (*raised*, *plain*, *lowered*) and three secondary frontness distinctions (*advanced*, *plain*, *retracted*) were also put forward. Hence as Bell pointed out, there may be up to nine vertical and nine horizontal degrees of phonetic differentiation [1867, p. 16]. For more clarity, the equivalent modern terms have been substituted for Bell's: central ('mixed'), raised ('higher'), plain ('normal'), lowered ('lower'), advanced ('outer'), retracted ('inner'). Two traditions have emerged from this classification. The International Phonetic Association retains the primary frontness terms (*front*, *central*, *back*) as well as the secondary diacritics (*raised*, *lowered*, *advanced*, *retracted*), but substitutes the terms close and open for high and low. The American tradition, on the other hand, follows Bell's terminology more closely. Pullum and Ladusaw [1996, xxxiii–iv] mention that both the IPA and American traditions can accommodate at least seven distinct vowel heights, with a two-way split of the high and low vowels, and a three-way split of the mid vowel. For example, a recent IPA chart [Handbook of the IPA, 1999] distinguishes 1) close, 2) near-close, 3) close-mid, 4) mid, 5) open-mid, 6) near-open, 7) open. In his survey of 317 phonological systems Maddieson [1984, p. 204] characterizes the vowel height categories as 1) high, 2) lowered high, 3) higher mid, 4) mid, 5) lower mid, 6) raised low, 7) low.

Using a pattern playback procedure, Delattre et al. [1952] found that vowel height is inversely related to F1 frequency and that vowel frontness is directly related to F2 frequency. Miller [1953] varied the formant frequency, amplitude, and fundamental frequency of synthetic vowels and asked subjects to identify each sound as one of eleven American English vowels. Consistent identification patterns were obtained for vowels with fixed F1 and F2 frequencies but fairly arbitrary formant amplitudes. Furthermore, an increase in F0 often produced a perceptual shift toward more open vowels characterized by a higher F1. Later studies revealed only a modest influence of F0 on vowel identification—in contrast to the robust effect of F1 frequency [Hoemeke and Diehl, 1994, and references therein]. When a harmonic coincides with the resonance frequency, an increase in F0 should yield a higher peak F1 frequency. This behavior is compatible with a model of speech perception based on spectral peaks [cf. Hillenbrand et al., 2006]. In their auditory model Lindblom et al. [2009] observed that the peak (or dominant) F1 frequency follows the strongest source harmonic(s) and not the filter resonance frequency.

As summarized in Section 8.4, the tongue root aperture (or lip-normalized tongue root area) is best correlated with log F1 (r = –0.935). Accordingly, vowel height distinctions are made by varying chiefly the tongue root aperture. Thus for greater terminological precision, the primary and secondary vertical labels: *high*, *mid*, *low*, *raised*, *plain*, *lowered* should be replaced by the equivalent primary and secondary tongue root terms: *expanded*, *neutral*, *contracted*, *advanced*, *plain*, *retracted*. Observe, however, that lip aperture is moderately correlated with log F1 (r = 0.723), which somewhat justifies the traditional vertical labels on the condition that they refer to the height of the lower lip. The primary term *expanded* (*high*) is adapted from Lindau [1979]; its opposing primary term *contracted* (*low*) is notionally identical to Perkell's *constricted pharynx* [1971]. The secondary term *advanced tongue root* (*raised*) is currently in widespread use, its opposing secondary term *retracted tongue root* (*lowered*) appreciably less so [Vaux, 1999]. The terms *root-advanced* and *root-unadvanced* were first employed by Stewart [1967] to distinguish between raised and unraised vowels in the cross-height harmony system of Akan. In later phonological work, the binary feature [Advanced Tongue Root] or [ATR] became the conventional indicator of the general contrast between raised and nonraised vowels [see Trigo, 1991 for a historical overview].

Hellwag presented the earliest known vowel triangle in 1781 [p. 41]. He observed the mandible (together with the tongue and lower lip) to be fully abducted for [a] but minimally abducted for [i u]. Lindau [1975, p. 30] points out that the tongue root area becomes smaller with mandible lowering:

> "Since the effective hinge of the mandible is above and behind the tongue, lowering the mandible will necessarily lower the front part of the tongue. It will also cause the back part of the tongue to retract, so there is less space for variation of the sagittal cross-dimension of the pharynx for low vowels than for non-low vowels."

The biomechanical linkage of the pharynx and the jaw should lead to a quasi-inverse relationship between the tongue root and lip areas. The mean correlation of –0.507 between the articulatory parameters $\overline{A}(root)$ and $\overline{A}(lip)$ supports this hypothesis (s.d. 0.241, N = 10). It is apparent though that mandible lowering represents just one of several strategies to control tongue root area. When the jaw is fixed by a rigid bite block, for example, typical vowel F1 frequencies continue to be produced as a result of motor compensation [Gay et al., 1981]. Hence the perceived vowel height may poorly reflect the actual abduction of the jaw or lower lip. Inconsistent correspondence between perceived height and the degree of jaw opening is reported for normal unconstrained vowels as well [Ladefoged et al., 1972].

By analogy with the tri-partition of front, central, and back vowels, Bell [1867, p. 16] suggested "a corresponding tri-partition of the aperture between the tongue and the palate, according to the 'high,' 'mid,' or 'low' position of the tongue." Fischer-Jørgensen [1985] attributes the association of vowel height with the highest point of the tongue to a British tradition introduced by Daniel Jones. For instance, Jones [1909, p. 10] wrote:

> "Vowels are thus classified as front, mixed, and back, according to the horizontal position of the highest point of the tongue. They may also be classified according to the vertical position of the highest point of the tongue."

Continuing the tradition, Chomsky and Halle specifically identify the tongue body as the principal determiner of vowel height [1968, pp. 304–305]:

> "HIGH–NONHIGH. High sounds are produced by raising the body of the tongue above the level that it occupies in the neutral position; nonhigh sounds are produced without such a raising of the tongue body.
>
> LOW–NONLOW. Low sounds are produced by lowering the body of the tongue below the level that it occupies in the neutral position; nonlow sounds are produced without such a lowering of the body of the tongue."

The same definition of HIGH and LOW is maintained in more recent accounts of tongue body features [Halle, 1992; Stevens, 1998, p. 250].

The perturbation analysis of the second formant gives support to Jones' statement relating vowel frontness and the horizontal position of the highest point of the tongue. The second resonance mode displays a volume velocity maximum at one-third and a volume velocity minimum at two-thirds the overall length from the glottis. This quarter-wavelength interval is approximately co-extensive with the tongue body (Figure 1). As the constriction formed by the tongue body and palate moves toward the lips from the volume velocity maximum (back vowel) to the volume velocity minimum (front vowel), the F2 frequency should shift from its lowest to highest value (Section 4.1). The theoretical expectation is confirmed since the tongue body position $I_{\min A}(body)$ is best correlated with log F2.

On the other hand, the perturbation analysis of the first formant does not support Jones' statement relating vowel height and the vertical position of the highest point of the tongue. The first resonance mode (F1) exhibits a volume velocity maximum (pressure minimum) at the lips and a volume velocity minimum (pressure maximum) at the glottis. Therefore the F1 formant frequency should be quite sensitive to area changes in the terminal lip and tongue root regions, but much less so in the medial tongue body region. The very low mean correlation of –0.154 between log F1 and the tongue body area $\log \overline{A}(body)$ shows that, as predicted, area changes in the tongue body region have little effect on F1 frequency. In addition, the equally low mean correlation of 0.217 between the articulatory parameters $\overline{A}(root)$ and $\overline{A}(body)$ indicates that the tongue root and the tongue body areas vary almost independently of each other (s.d. 0.571, N = 10).

Wood [1975] observed that early X-ray studies, such as Russell [1928], had already raised many doubts about the hypothesized relation between vowel height and the vertical position of the tongue. To pursue the question further, he examined the vertical tongue positions in X-ray tracings from fifteen languages and found frequent height inversions among the high and mid as well as the mid and low vowels. Given the evident limitations of the hypothesis, Wood concluded that its continued acceptance in the language sciences was due more to the lack of a satisfactory alternative than to its phonetic groundedness. Perturbation analysis, however, does provide a framework for a better understanding of acoustic-motor relationships. The mean correlations between log F1 and the areas of the four articulator regions are, from the largest absolute value to the smallest (N = 10):

1	$\log \overline{A}(root)$	–0.915 (s.d. 0.071)
2	$\log \overline{A}(lip)$	0.723 (s.d. 0.224)
3	$\log \overline{A}(blade)$	0.720 (s.d. 0.137)
4	$\log \overline{A}(body)$	–0.154 (s.d. 0.532)

The perturbation analysis of the first formant correctly accounts for the observed rank order since the terminal articulator regions (tongue root, lip) show larger coefficient magnitudes than the more medial ones (blade, tongue body). Recall also from Section 8.4

that for a given area perturbation the 8-tube tongue root region has a greater effect on the first formant than the 4-tube lip region.

10. Summary of the main findings

A 27-tube vocal tract model (FDVT) is developed to determine eight acoustic parameters: the formant frequencies F1–F4 and the quality or amplification factors Q1–Q4. Four articulator regions are delimited in the FDVT model, each characterized by an active articulator: the 8-tube tongue root region, the 9-tube tongue body region corresponding to a quarter wavelength at the second formant frequency, the 6-tube blade region corresponding to a quarter wavelength at the third formant frequency, and the 4-tube lip region. The vowel area functions of ten speakers were obtained from seven X-ray and MRI investigations and fit to the 27 equal-length tubes by means of cubic spline interpolation. Correlation matrices between the acoustic and articulatory parameters are calculated for the vowel system of each speaker. The coefficients of the parameter pairs are then averaged across the ten speakers. The results of the study show that (Section 8.4):

1. Tongue root aperture (tongue root area normalized by lip area) is inversely correlated with F1 frequency.
2. As the tongue body position (location of smallest constriction) moves toward the lips, the F2 frequency also shifts higher.
3. Tongue body aperture (tongue body area normalized by lip area) is directly correlated with Q2.
4. As the blade position (location of smallest constriction) moves toward the lips, the F3 frequency also shifts higher.
5. Blade aperture (blade area normalized by lip area) is directly correlated with Q3.
6. Lip position (sum of tube lengths in lip region) displays an inverse correlation with F4 frequency.
7. Lip aperture (lip area) has a moderate direct correlation with F1 frequency and a weaker inverse correlation with Q4.

Furthermore, it is shown in Sections 8.2 and 8.3 that the dominant mechanism governing the quality factor of the higher formants (Q2–Q4) is radiation damping. The above findings clearly demonstrate that the first four formant frequencies and quality factors are critical acoustic parameters in speech.

Acknowledgements

I wish to thank Stuart Davis, Samuel Obeng, Robert Port, and Charles Watson for reading the first draft of this report and giving me their valuable comments and criticisms. I am especially indebted to Kenneth de Jong who provided me with insightful observations and numerous detailed suggestions for improvement. Any remaining errors of fact or interpretation are of course solely my own.

References

Aarts, R.M.; Janssen, A.J.E.M.: Approximation of the Struve function H_1 occurring in impedance calculations. J. acoust. Soc. Am. *113:* 2635–2637 (2003).

Badin, P.: Acoustics of voiceless fricatives: production theory and data. Q. Prog. Status Rep., Speech Transm. Lab., R. Inst. Technol., Stockh., No. 30, pp. 33–55 (1989).

Badin, P.; Perrier, P.; Boë, L.J.; Abry, C.: Vocalic nomograms: acoustic and articulatory considerations upon formant convergences. J. acoust. Soc. Am. *87:* 1290–1300 (1990).

Baer, T.; Gore, J.C.; Gracco, L.C.; Nye, P.W.: Analysis of vocal tract shape and dimensions using magnetic resonance imaging: vowels. J. acoust. Soc. Am. *90:* 799–828 (1991).

Bell, A.M.: Visible speech (Simkin, Marshall & Co., London 1867).

Boë, L.J.; Perrier, P.: Comments on “Distinctive regions and modes: a new theory of speech production” by M. Mrayati, R. Carré, and B. Guérin. Speech Commun. *9:* 217–230 (1990).

Browman, C.P.; Goldstein, L.: Articulatory gestures as phonological units. Phonology *6:* 201–251 (1989).

Carré, R.: From an acoustic tube to speech production. Speech Commun. *42:* 227–240 (2004).

Catford, J.C.: Observations on the recent history of vowel classification; in Asher, Henderson, Towards a history of phonetics, pp. 19–32 (University Press, Edinburgh 1981).

Chomsky, N.; Halle, M.: The sound pattern of English (Harper & Row, New York 1968).

Chiba, T.; Kajiyama, M.: The vowel, its nature and structure (Phonetic Society of Japan, Tokyo 1958).

Chuenwattanapranithi, S.; Xu, Y.; Thipakorn, B.; Maneewongvatana. S.: Encoding emotions in speech with the size code. A perceptual investigation. Phonetica *65:* 210–230 (2008).

Delattre, P.; Liberman, A.M.; Cooper, F.S.; Gerstman, L.J.: An experimental study of the acoustic determinants of vowel color; observations on one- and two-formant vowels synthesized from spectrographic patterns. Word *8:* 195–210 (1952).

Ewan, W.G.; Krones, R.: Measuring larynx movement using the thyroumbrometer. J. Phonet. *2:* 327–335 (1974).

Fairbanks, G.: A physiological correlative of vowel intensity. Speech Monogr. *17:* 390–395 (1950).

Fant, G.: Acoustic theory of speech production (Mouton, The Hague 1960).

Fant, G.: Vocal-tract area and length perturbations. Q. Prog. Status Rep., Speech Transm. Lab., R. Inst. Technol., Stockh., No. 16, pp. 1–14 (1975).

Fant, G.; Pauli, S.: Spatial characteristics of vocal tract resonance modes; in Proceedings of the Speech Communication Seminar, pp. 121–132 (Almqvist & Wiksell, Stockholm 1975).

Fischer-Jørgensen, E.: Some basic vowel features, their articulatory correlates, and their explanatory power in phonology; in Fromkin, Phonetic linguistics, pp. 79–99 (Academic Press, Orlando 1985).

Fitch, W.T.; Giedd, J.: Morphology and development of the human vocal tract: a study using magnetic resonance imaging. J. acoust. Soc. Am. *106:* 1511–1522 (1999).

Flanagan, J.L.: Speech analysis, synthesis and perception (Springer Verlag, Berlin 1972).

Gay, T.; Lindblom, B.; Lubker, J.: Production of bite-block vowels: acoustic equivalence by selective compensation. J. acoust. Soc. Am. *69:* 802–810 (1981).

Halle, M.: Phonological features; in Bright, International Encyclopedia of Linguistics, pp. 207–212 (Oxford University Press, New York 1992).

Handbook of the International Phonetic Association (Cambridge University Press, 1999).

Hellwag, C.F.: Dissertatio de formatione loquelae (Verlag von Gebr. Henninger, Heilbronn 1781/1886).

Hillenbrand, J.M.; Houde, R.A.; Gayvert, R.T.: Speech perception based on spectral peaks versus spectral shape. J. acoust. Soc. Am. *119:* 4041–4054 (2006).

Hoemeke, K.A.; Diehl, R.L.: Perception of vowel height: the role of F_1–F_0 distance. J. acoust. Soc. Am. *96:* 661–674 (1994).

Hoole, P.; Kroos, C.: Control of larynx height in vowel production. Proc. 5th Int. Conf. Spoken Lang. Process. (ICSLP '98), 2, pp. 531–534 (1998).

Ishizaka, K.; French, J.C.; Flanagan, J.L.: Direct determination of vocal tract wall impedance. IEEE Trans. Acoust. Speech Signal Process. *23:* 370–373 (1975).

Jones, D.: The pronunciation of English (Cambridge University Press, 1909).

Keating, P.: Coronal places of articulation; in Paradis, Prunet, Phonetics and Phonology 2. The special status of coronals: internal and external evidence, pp. 29–48 (Academic Press, San Diego 1991).

Kinsler, L.E.; Frey, A.R.: Fundamentals of Acoustics (John Wiley & Sons, New York 1962).

Ladefoged, P.; DeClerk, J.; Lindau, M.; Papçun, G.: An auditory-motor theory of speech production. UCLA Working Papers Phonet. *22:* 48–75 (1972).

Lindau, M.: [Features] for vowels. UCLA Working Papers Phonet. *30* (1975).

Lindau, M.: The feature expanded. J. Phonet. *7:* 163–176 (1979).

Lindblom, B.; Diehl, R.; Creeger, C.: Do ‘Dominant Frequencies’ explain the listener’s response to formant and spectrum shape variations? Speech Commun. *51:* 622–629 (2009).

Maddieson, I.: Patterns of sounds (Cambridge University Press, 1984).

Miller, R.L.: Auditory tests with synthetic vowels. J. acoust. Soc. Am. *25:* 114–121 (1953).

Miller, J.D.: Auditory-perceptual interpretation of the vowel. J. acoust. Soc. Am. *85:* 2114–2134 (1989).

Mrayati, M.; Carré, R.: Relations entre la forme du conduit vocal et les caractéristiques acoustiques des voyelles françaises. Phonetica *33:* 285–306 (1976).

Mrayati, M.; Guérin, B.: Étude des caractéristiques acoustiques des voyelles orales françaises par simulation du conduit vocal avec pertes. Rev. d’Acoustique *36:* 18–32 (1976).

Mrayati, M.; Carré, R.; Guérin, B.: Distinctive regions and modes: a new theory of speech production. Speech Commun. *7:* 257–286 (1988).

Perkell, J.S.: Physiology of speech production: a preliminary study of two suggested revisions of the features specifying vowels. Quart. Prog. Report, Res. Lab. Electron. MIT *102:* 123–139 (1971).

Pullum, G.K.; Ladusaw, W.A.: Phonetic symbol guide (University of Chicago Press, 1996).

Riordan, C.J.: Control of vocal-tract length in speech. J. acoust. Soc. Am. *62:* 998–1002 (1977).

Rosenblum, L.D.: Speech perception as a multimodal phenomenon. Curr. Dir. Psychol. Sci. *17:* 405–409 (2008).

Russell, G.O.: The vowel (Ohio State University Press, 1928).

Schroeder, M.R.: Determination of the geometry of the human vocal tract by acoustic measurements. J. acoust. Soc. Am. *41:* 1002–1010 (1967).

Stevens, K.N.: Acoustic phonetics (MIT Press, 1998).

Stewart, J.M.: Tongue root position in Akan vowel harmony. Phonetica *16:* 185–204 (1967).

Stone, M.; Epstein, M.A.; Iskarous, K.: Functional segments in tongue movement. Clin. Linguist. Phon. *18:* 507–521 (2004).

Story, B.H.: Technique for “tuning” vocal tract area functions based on acoustic sensitivity functions. J. acoust. Soc. Am. *119:* 715–718 (2006).

Story, B.H.; Titze, I.R.; Hoffman, E.A.: Vocal tract area functions from magnetic resonance imaging. J. acoust. Soc. Am. *100:* 537–554 (1996).

Story, B.H.; Titze, I.R.; Hoffman, E.A.: Vocal tract area functions for an adult female speaker based on volumetric imaging. J. acoust. Soc. Am. *104:* 471–487 (1998).

Takemoto, H.; Honda, K.; Masaki, S.; Shimada, Y.; Fujimoto, I.: Measurement of temporal changes in vocal tract area function from 3D cine-MRI data. J. acoust. Soc. Am. *119:* 1037–1049 (2006).

Trigo, L.: On pharynx-larynx interactions. Phonology *8:* 113–136 (1991).

Turner, R.E.; Walters, T.C.; Monaghan, J.J.M.; Patterson, R.D.: A statistical, formant-pattern model for segregating vowel type and vocal-tract length in developmental formant data. J. acoust. Soc. Am. *125:* 2374–2386 (2009).

Vaux, B.: A note on pharyngeal features. Harvard Working Papers Ling. *7:* 39–63 (1999).

Wood, S.: The weaknesses of the tongue-arching model of vowel articulation. Working Papers *11:* 55–107 (Dept. of Linguistics, Lund 1975).

Wood, S.: A radiographic analysis of constriction location for vowels. J. Phonet. *7:* 25–43 (1979).

Yang, C.S.; Kasuya, H.: Accurate measurement of vocal tract shapes from magnetic resonance images of child, female, and male subjects. Proc. 3rd Int. Conf. Spoken Lang. Process. (ICSLP '94), pp. 623–626 (1994).

Zhang, S.; Jin, J.: Computation of special functions (Wiley Interscience, New York 1996).

Zhang, Z.; Espy-Wilson, C.Y.: A vocal-tract model of American English /l/. J. acoust. Soc. Am. *115:* 1274–1280 (2004).