Acoustic-articulatory correlations in a four-region model of the vocal tract: Theoretical bases and a comparison of two data sets

Mark Pennington

Abstract

A 27-tube frequency-domain vocal tract model (FDVT) was developed to calculate the formant frequencies F1–F4 and the quality or amplification factors Q1–Q4. The quality factors Q1–Q4 are the relative amplitudes of formants F1–F4. Four articulator regions are distinguished in the FDVT model: an 8-tube tongue root region, a 9-tube tongue body region (one quarter wavelength at the F2 formant frequency), a 6-tube blade region (one quarter wavelength at the F3 formant frequency), and a 4-tube lip region (one quarter wavelength at the F4 formant frequency). The vowel area functions of 10 speakers were previously investigated by the author (Pennington, 2011). More recently, Tiede (2013) gave the vowel area functions of another 12 speakers. To determine the degree of association between acoustic and articulatory parameters, correlation matrices are computed for the vowel system of each speaker. Then the correlation coefficients of the parameter pairs are averaged across the 10 speakers in the original data set and the 12 speakers in the Tiede data set. This study has two goals. The first is to compare the correlation results of the original and the Tiede vowel data sets. The second is to provide theoretical foundations for the observed acoustic-articulatory correlates. The following is a preview of the correlation results: (1) tongue root aperture (tongue root area divided by lip area) is inversely correlated with F1; (2) tongue body position (forebody area divided by hindbody area) is inversely correlated with F2; (3) tongue body aperture (tongue body area divided by lip area) is directly correlated with Q2; (4) blade position (foreblade area divided by hindblade area) is inversely correlated with F3; (5) blade aperture (blade area divided by lip area) is directly correlated with Q3; (6) lip length is inversely correlated with F4; (7) lip area is inversely correlated with Q4.

This paper is a completely revised and greatly expanded version of a 2011 article that appeared in volume 11 of the IULC Working Papers:

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

Contents

Part I. Background

1. Introduction

A key obstacle to the development of a comprehensive distinctive feature theory has been the long-standing uncertainty about the relation between speech gesture and acoustic outcome. Jakobson, Fant, and Halle (1952) identified correlates of production and perception for the twelve binary oppositions that they detected in the world's languages. However subsequent work on distinctive features has not continued the research into co-varying motor and acoustic parameters. Most emphasis has been devoted to the motor dimension (Chomsky & Halle, 1968; Clements & Hume, 1995) or exceptionally to the acoustic-perceptual dimension (Flemming, 2002). The principal reason for the uncertainty appears to be the lack of vocal tract studies in which acoustic-articulatory correlation coefficients are calculated and ranked. Fairbanks (1950) correlated vowel acoustic power with X-ray measures of channel diameter at three vocal tract locations (lips, incisors, tongue-palate). Nonetheless, before 1990, a general assessment of acoustic-articulatory correlations would have been almost impossible due to the scarcity of X-ray data. Complete vowel area functions of only 2 speakers were known up to that time (Fant, 1960; Mrayati & Guérin, 1976), which are clearly not enough to provide meaningful correlation averages across subjects. With the widespread adoption of MRI techniques after 1990, the vowel area functions of another 8 speakers became available for a total of 10 by 2006 (Table 1). More recently, Tiede (2013) obtained through MRI imaging the vowel area functions of 12 additional speakers, thereby more than doubling the overall number to 22. Six American English and six Japanese participants served as subjects with five men and one woman in each language group. In a previous article by the author (Pennington, 2011), acoustic-articulatory correlations were calculated using the 10-speaker data set shown in Table 1. A major goal of this paper is to compare the correlations of this earlier study with those of the 12-speaker data set newly provided by Tiede. The other central aim is to present the theoretical bases underlying the correlation results.

To obtain the acoustic-articulatory correlations, a 27-tube frequency-domain vocal tract model (FDVT) calculates eight acoustic parameters: the formant frequencies F1–F4 and the quality or amplification factors Q1–Q4. The quality factors Q1–Q4 are the relative amplitudes of formants F1–F4. The quality factor Q can be defined as the formant frequency F divided by its half-power bandwidth B. Current feature analyses of the speech segment discriminate the movements of four active articulators: the pharyngeal articulator or (i) the tongue root, and the three buccal articulators, (ii) the tongue body, (iii) the blade, and (iv) the lips (Halle & Stevens, 1991; Halle, Vaux, & Wolfe, 2000; Gussenhoven & Jacobs, 2017, pp. 76–79). Accordingly, the open-closed boundary conditions of a uniform pipe are applied to the vocal tract, partitioning it

Table 1
Seven X-ray and MRI studies with vowel area functions from ten speakers.

Study	Year	Method	Age	Sex	Language	Vowels
Fant	1960	X-ray	adult	male	Russian	6
Mrayati & 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 & Kasuya	1994	MRI	adult	male	Japanese	5
Yang & Kasuya	1994	MRI	adult	female	Japanese	5
Yang & 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

acoustically into 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 corresponding to a quarter wavelength at the fourth formant frequency.

Vowel area functions of a particular speaker are taken from published tables in the X-ray or 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 language-specific vowel system of the speaker. The correlation coefficients of the parameter pairs are then averaged across the 10 speakers in the original data set and the 12 speakers in the Tiede data set. An initial overview of the results is as follows:

1. Tongue root aperture (tongue root area divided by lip area) is inversely correlated with the F1 formant frequency.
2. Tongue body position (forebody area divided by hindbody area) is inversely correlated with the F2 formant frequency.
3. Tongue body aperture (tongue body area divided by lip area) is directly correlated with the Q2 quality factor.
4. Blade position (foreblade area divided by hindblade area) is inversely correlated with the F3 formant frequency.
5. Blade aperture (blade area divided by lip area) is directly correlated with the Q3 quality factor.
6. Lip length is inversely correlated with the F4 formant frequency.
7. Lip area is inversely correlated with the Q4 quality factor.

The term *aperture* is defined as the cross-sectional area of one of three articulator regions (tongue root, tongue body, or blade) divided by the cross-sectional area of the lip region. In earlier research (Pennington, 2011), a final summary gave the same statements as above except for 2. and 4. (p. 24). The two statements referred to tongue body and blade constrictions instead of tongue body and blade positions (cf. Sections 9.2.2–9.2.3 below). The summary was based on the original 10-speaker data set. In contrast, the overview

statements presented here are supported not only empirically by the original and Tiede data sets, but also by the theoretical findings.

The paper is broadly organized into a theoretical framework (Part II: Sections 3–5) followed by an experimental study (Part III: Sections 6–9), where the vowel acoustic-articulatory correlates are reviewed in light of the theoretical premises. Section 3 develops the perturbation analysis for the first three formant frequencies F1–F3. Section 5 shows that radiation bandwidth is the important determinant of the quality factors Q2–Q4.

2. Description and validation 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:

1. acoustic inductance $L = \rho/A$
2. series viscous loss $R_{visc} = (S/A^2)\sqrt{\omega\rho\mu/2}$
3. acoustic capacitance $C = A/\rho c^2$
4. parallel heat conduction loss $G_{heat} = S((\eta-1)/\rho c^2)\sqrt{\lambda\omega/2c_p\rho}$,

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: $\rho = 1.14\times10^{-3}$ gm/cm^3, $c = 3.5\times10^4$ cm/sec, $\eta = 1.4$, $\mu = 1.86\times10^{-4}$ dyne-sec/cm^2, $\lambda = 0.055\times10^{-3}$ cal/cm-sec-deg, $c_p = 0.24$ cal/gm-degree.

A mass-compliance-viscous loss (M-K-R_{wall}) series circuit in parallel with C and G_{heat} models the wall impedance Z_{wall}, where $M = 2.1$ gm, $K = 84.5\times10^3$ dyne/cm, and $R_{wall} = 800$ gm/sec. These per-unit-area parameters of the relaxed cheek were determined experimentally by Ishizaka, French, and Flanagan (1975).

When the radius of the lip opening is small compared with the radius of the head, the radiation impedance Z_{rad} comes close to that of a circular piston in a plane

$$Z_{rad} = \frac{\rho c}{A_0}\left[1 - \frac{J_1(2kr)}{kr} + j\frac{H_1(2kr)}{kr}\right],$$

where k is the angular wave number ω / c, $A_0 = \pi r^2$ the terminating lip area, and r the terminating lip radius (tube 1); J_1 and H_1 are first-order Bessel and Struve functions (Aarts & 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 & Guérin, 1976). For each frequency, the driving volume velocity at the closed glottis (tube 27) is set to 1 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. Since the power at a given frequency is proportional to the square of the volume velocity magnitude, the total power will be defined as

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

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 formant frequency F, as well as the low and high half-power frequencies f_1 and f_2 −3 dB below the peak formant frequency. Once the spectral bandwidth is determined $B = f_2 - f_1$, the dimensionless quality factor is found $Q = F / B$. Then the logarithmic quality factor is calculated log Q $= 20\log_{10}(Q)$. The log Q is a measure of the relative power or power gain at resonance (Kinsler & Frey, 1962, p. 195).

To test the validity of the FDVT model, the formant frequencies and bandwidths 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 & Espy-Wilson, 2004). As there are 27 tubes, the elementary tube is 2/3 cm long. The VTAR constants are set identical to the ones used in the FDVT program. VTAR provides an on-off switch for the radiation impedance; however, the method of computing the radiation impedance is not documented. The simulations yield typical differences of less than 1% between the FDVT and VTAR acoustic results (Appendix A). Thus the validity of the FDVT model is confirmed satisfactorily.

Part II. The theoretical framework

3. Articulator regions

3.1. Articulator regions: the perturbation analysis of formant frequencies F1–F3

3.1.1. The preliminary perturbation analysis

The first four formant frequencies of a lossless open-closed pipe of length *l* are

$$F_i = \frac{2i-1}{4} \times \frac{c}{l}$$ (Kinsler & Frey, 1962, p. 202),

where $1 \le i \le 4$ is the formant number and *c* the sound velocity. Thus the second F_2, third F_3, and fourth F_4 formant frequencies are odd multiples of the first formant frequency F_1. Since the wavelength or spatial period of formant frequency F_i is

$$\lambda_i = \frac{c}{F_i} = \frac{4l}{2i-1}$$ (Kinsler & Frey, 1962, p. 42),

the wavelengths of the first four formants are $4l$, $4l/3$, $4l/5$, $4l/7$. Seeing that the pipe length is $l = (2i-1) \times \lambda_i / 4$, then one quarter wavelength fits into the pipe at F_1, three quarter wavelengths at F_2, five quarter wavelengths at F_3, and seven quarter wavelengths at F_4 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. Stevens (1998, p. 140) presents the volume velocity and pressure envelopes for F_1, F_2, and F_3 while Fant (1960, p. 85) gives the volume velocity envelopes for F_1, F_2, F_3, F_4. Fig. 1 below shows the volume velocity envelopes for the F_2 and F_3 formant frequencies. For each formant F_i, there is a volume velocity maximum at the lips and a pressure maximum at the glottis.

Chiba and Kajiyama (1958, p. 151) discovered that there is 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:

> "The rules for relating formant frequency variations to localized constrictions or expansions within a single tube may be derived from simple impedance considerations. 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."

To summarize the Fant perturbation rules: (1) an area change at a volume velocity maximum has an effect on only the local inductance *L*; (2) an area change at the pressure maximum has an effect on only the local capacitance *C*. Remark that in an open-closed pipe, the quarter wavelength $\lambda_i / 4$ of formant number *i* is the smallest spatial interval containing both a volume velocity maximum and a pressure maximum. The two maxima are situated at either extremity of the quarter wavelength.

In order to develop the perturbation analysis, the equation for the resonance frequency of an idealized lumped-parameter pipe will be presented and then adapted to the analysis of formant frequencies. It is convenient to start with a horizontal mass-spring system where an oscillating mass *m* is attached to a spring of stiffness constant *s*. The resonance frequency of the mass-spring system is

$$f_0 = \frac{1}{2\pi}\sqrt{\frac{s}{m}}$$ (Kinsler & Frey, 1962, p. 3).

A Helmholtz resonator consists of a small cylindrical opening (or neck) joined to an enclosure such as a cylindrical cavity (Kinsler & Frey, 1962, p. 186). The air in the neck acts like a mass, whereas the air in the cavity acts like a spring. Thus the Helmholtz resonator is the acoustical analog of the mass-spring system whose lumped mechanical parameters are mass and stiffness. The acoustic mass *M* and the acoustic stiffness *s* of the Helmholtz resonator are

$$M = Ll_L = \frac{\rho}{A_L} l_L \,;\; s = \frac{1}{Cl_C} = \frac{1}{(A_C / \rho c^2) l_C},$$

where ρ is the air density and *c* the sound velocity. The inductance area and length of the cylindrical neck are A_L and l_L; the per-unit-length acoustic inductance of the neck is *L*. The capacitance area and length of the cylindrical cavity are A_C and l_C; the per-unit-length acoustic capacitance of the cavity is *C*. Note that the acoustic stiffness *s* is the reciprocal of the product of the per-unit-length acoustic capacitance *C* and the capacitance length l_C. The resonance frequency of the cylindrical Helmholtz resonator is therefore (Kinsler & Frey, 1962, p. 193):

$$f_0 = \frac{1}{2\pi}\sqrt{\frac{s}{M}} = \frac{1}{2\pi}\sqrt{\frac{1}{Ll_L \times Cl_C}} = \frac{1}{2\pi}\sqrt{\frac{1}{(\rho / A_L) l_L \times (A_C / \rho c^2) l_C}}.$$

If the inductance and capacitance cylinders are of equal length and joined to form an idealized lumped-parameter pipe of length l, it follows that $l_L = l_C = l/2$. The equation for the resonance frequency becomes

$$f_0 = \frac{1}{2\pi}\sqrt{\frac{1}{(\rho / A_L)(l/2)\times(A_C/\rho c^2)(l/2)}} = \frac{1}{2\pi}\times\frac{2c}{l}\sqrt{\frac{A_L}{A_C}} = \frac{1}{\pi}\times\frac{c}{l}\sqrt{\frac{A_L}{A_C}}.$$

Hence the resonance frequency f_0 is inversely proportional to l when $\sqrt{A_L / A_C}$ is constant.

The equations for the pressure and volume velocity at position *x* and time *t* in a finite cylindrical pipe are

$$p(x,t) = [Ae^{-jkx} + Be^{jkx}]e^{j\omega t}$$

$$U(x,t) = \left(\frac{S}{\rho c}\right)[Ae^{-jkx} - Be^{jkx}]e^{j\omega t},$$

where A and B are complex amplitudes, S is the cross-sectional area, $\omega = 2\pi F$ the angular frequency, and $k = \omega / c = 2\pi F / c$ the angular wave number (Rossing & Fletcher, 2004, pp. 181–182). In this distributed-parameter pipe, pressure oscillations are expressed as the product of two sinusoidal functions, one indicating the spatial envelope of the pressure $Ae^{-jkx} + Be^{jkx}$, the other specifying the excursion of the oscillations as a function of time $e^{j\omega t}$. When the open-closed boundary conditions are imposed on the system (Rossing & Fletcher, p. 183), the first four formant frequencies and their quarter wavelengths are

$$F_i = \frac{2i-1}{4}\times\frac{c}{l};\ \lambda_i / 4 = \frac{l}{2i-1}\ \ 1 \le i \le 4,$$

as was given earlier. If one juxtaposes the equations for the F1 frequency of the distributed-parameter pipe and the resonance frequency of the idealized lumped-parameter pipe:

$$F_1 = \frac{1}{4}\times\frac{c}{l};\ f_0 = \frac{1}{\pi}\times\frac{c}{l}\sqrt{\frac{A_L}{A_C}},$$

the F1 frequency and the resonance frequency f_0 are seen to be quite similar when A_L equals A_C and hence when there is no area perturbation at the ends of the lumped-parameter pipe.

The F1, F2, F3, and F4 quarter wavelengths of the distributed-parameter pipe are l, $l/3$, $l/5$ and $l/7$. Inside the F1, F2, F3 and F4 quarter wavelengths, the sinusoidal time functions of the volume velocity and pressure are $e^{j2\pi(F_1)t}$, $e^{j2\pi(3F_1)t}$, $e^{j2\pi(5F_1)t}$ and $e^{j2\pi(7F_1)t}$ with frequencies F_1, $3F_1$, $5F_1$ and $7F_1$. By way of comparison, the resonance frequencies of lumped-parameter pipes with lengths l, $l/3$, $l/5$, and $l/7$ are f_0, $3f_0$, $5f_0$ and $7f_0$, given that f_0 is inversely proportional to l when all else is constant. Thus for the narrow purpose at hand, the lumped-parameter pipe of length l appears to be a suitable approximation of the distributed quarter wavelength $\lambda / 4$.

The following proportionality holds for the resonance frequency f_0:

$$f_0 = \frac{1}{\pi} \times \frac{c}{l} \sqrt{\frac{A_L}{A_C}} \propto \frac{1}{l} \sqrt{\frac{A_L}{A_C}}.$$

The acoustic inductance L is equal to ρ / A_L. As A_L becomes smaller, L increases and thereby lowers the resonance frequency, or vice versa. The acoustic capacitance C is equal to $A_C / \rho c^2$. As A_C becomes smaller, C decreases and thereby raises the resonance frequency, or vice versa.

The Fant perturbation rules state that:

1. An area change at a volume velocity maximum has an effect on only the local inductance L.
2. An area change at the pressure maximum has an effect on only the local capacitance C.

In view of the parallel drawn above between the idealized lumped-parameter pipe of length l and the distributed quarter wavelength $\lambda / 4$, one can hypothesize the following proportionality to hold for the formant frequency F_i:

$$F_i \propto \frac{1}{(\lambda_i / 4)} f\left(\frac{A_{U\,max}}{A_{p\,max}}\right),$$

where $A_{U\,max}$ is the area at the volume velocity maximum, $A_{p\,max}$ the area at the pressure maximum, and $f(A_{U\,max} / A_{p\,max})$ a dimensionless function of the area ratio $A_{U\,max} / A_{p\,max}$. Like the lumped-parameter pipe, one expects that (i) when $A_{U\,max}$ becomes smaller, L increases and lowers the formant frequency, or vice versa; (ii) when $A_{p\,max}$ becomes smaller, C decreases and raises the formant frequency, or vice versa. Fant advanced the same observations in the text cited above.

Fig. 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 (pressure maximum) 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 (pressure maximum) 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 (reduced $A_{U\,max}$) to the pressure maximum (reduced $A_{p\,max}$), the formant frequencies should shift from their lowest values to their highest values in agreement with the 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 (see also Badin, Perrier, Boë, & Abry, 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

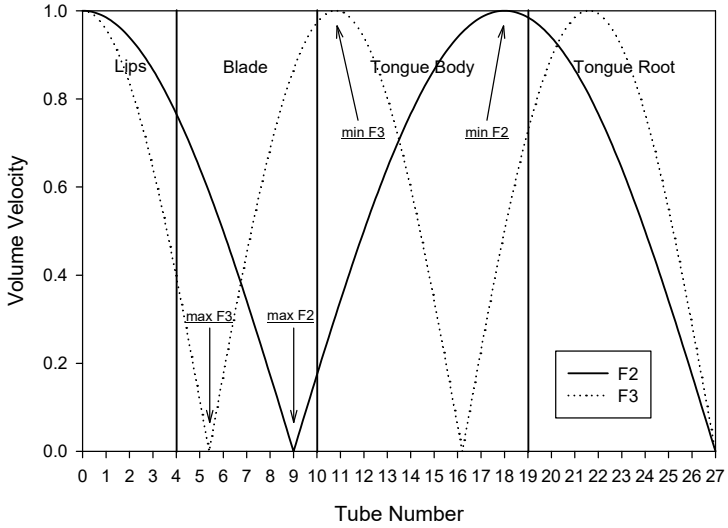

Fig. 1. Articulator regions and spatial volume velocity magnitudes of the second and third formants (F2, F3) in a lossless open-closed uniform pipe. The arrows indicate the constriction locations of the highest and lowest second formant frequencies max F2, min F2 as well as the highest and lowest third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

3.2 kHz. This is good 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 coextensive 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). Since the F4 quarter wavelength is 4/28 or 1/7 of the overall vocal tract length, the four tubes of the lip region fit into the F4 quarter wavelength almost exactly. 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 precise 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 5.33 cm in length (= 8/27 × 18 cm).

3.1.2. A test of the preliminary perturbation analysis

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 boundary 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 cm^2 (p. 49) and 0.25 cm^2 (p. 52). 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 and pressures of F1, F2, and F3 of eleven French vowels modeled as non-uniform pipes. Although the volume velocity and pressure values along the vocal tract often differed considerably among the synthetic vowels, the locations of the volume velocity and pressure maxima shifted little with respect to their locations in the uniform pipe designated as schwa [ə]. In sum, the partition of the vocal tract into the four articulator regions is valid for speech sounds because (a) the closed glottis boundary condition appears to be met and (b) the locations of the volume velocity and pressure maxima of a lossy non-uniform pipe largely coincide with those of the uniform pipe.

Because the formant frequency F_i is assumed to be proportional to $(\lambda_i / 4)^{-1} f(A_{U\,max} / A_{p\,max})$, a trading relation is expected between $A_{p\,max}$ and $A_{U\,max}$. Consequently, F_i should be identically raised whether $A_{p\,max}$ is halved or $A_{U\,max}$ is doubled; F_i should be identically lowered whether $A_{p\,max}$ is doubled or $A_{U\,max}$ is halved. To test for possible trading relations, a lossless FDVT model calculates the F1–F3 frequencies of open-closed pipes 18 cm long. The unmodified cross-sectional area of a pipe is 2 cm^2. Thus when the areas $A_{p\,max}$ and $A_{U\,max}$ are halved or doubled, the area either decreases to 1 cm^2 or increases to 4 cm^2. A summary of the area modifications for the F1–F3 quarter wavelengths and the corresponding formant frequencies is provided in

Table 2

Area modifications of the F1–F3 quarter wavelengths and the corresponding formant frequencies, frequency ratios, and octave changes per halving or doubling of the 1/3 quarter wavelength area. The full pipe length is 18 cm and the unmodified area is 2 cm^2.

F1 quarter wavelength (tubes 1–27)		
2 cm^2 uniform pipe (tubes 1–27)	486 Hz	
1 cm^2 hindtract (p max, tubes 19–27)	571 Hz	
4 cm^2 foretract (U max, tubes 1–9)	571 Hz	
geometric mean	571 Hz	
frequency ratio	571/486	1.1749
octave change		+0.2325
4 cm^2 hindtract (p max, tubes 19–27)	390 Hz	
1 cm^2 foretract (U max, tubes 1–9)	390 Hz	
geometric mean	390 Hz	
frequency ratio	390/486	0.8025
octave change		–0.3175

F2 quarter wavelength (tubes 11–19)		
2 cm^2 uniform pipe (tubes 1–27)	1458 Hz	
1 cm^2 forebody (p max, tubes 11–13)	1500 Hz	
4 cm^2 hindbody (U max, tubes 17–19)	1541 Hz	
geometric mean	1520 Hz	
frequency ratio	1520/1458	1.0425
octave change		+0.0601
4 cm^2 forebody (p max, tubes 11–13)	1364 Hz	
1 cm^2 hindbody (U max, tubes 17–19)	1330 Hz	
geometric mean	1347 Hz	
frequency ratio	1347/1458	0.9239
octave change		–0.1142

F3 quarter wavelength (tubes 5–10)		
2 cm^2 uniform pipe (tubes 1–27)	2430 Hz	
1 cm^2 foreblade (p max, tubes 5–6)	2510 Hz	
4 cm^2 hindblade (U max, tubes 9–10)	2454 Hz	
geometric mean	2482 Hz	
frequency ratio	2482/2430	1.0214
octave change		+0.0305
4 cm^2 foreblade (p max, tubes 5–6)	2283 Hz	
1 cm^2 hindblade (U max, tubes 9–10)	2329 Hz	
geometric mean	2306 Hz	
frequency ratio	2306/2430	0.9490
octave change		–0.0756

Table 2. In order to perform the evaluation, the question arises as to what fraction of the quarter wavelength should be assigned to the variable areas $A_{p\,max}$ and $A_{U\,max}$. In the previous section, the tongue root region was determined to be 8/27 of the F1 quarter wavelength which spans the entire vocal tract. The lip region, on the other hand, was found to be only 4/27 of the F1 quarter wavelength. The larger the quarter wavelength fraction, the more acoustic mass Ll_L and stiffness $1/Cl_C$ are available for modification by area change (cf. the inductance and capacitance lengths l_L and l_C in the cylindrical Helmholtz resonantor: $f_0 = (2\pi)^{-1}\sqrt{(Ll_L \times Cl_C)^{-1}}$). Hence 9/27 or 1/3 is the preferred

quarter wavelength fraction since it approximates the longer tongue root region and not the shorter lip region. As further evidence for a tongue root region of about 1/3 quarter wavelength, Section 3.2 below shows that the measured ratio between pharynx length and vocal tract length varies from 37.47% (men) to 25.39% (youngest children).

The 1/3 quarter wavelengths for the 27-tube F1 quarter wavelength (No. 1–27), the 9-tube F2 quarter wavelength (No. 11–19), and the 6-tube F3 quarter wavelength (No. 5–10) are defined as:

(1) 9-tube F1 hindtract (No. 19–27) at $A_{p\,max}$

(2) 9-tube F1 foretract (No. 1–9) at $A_{U\,max}$

(3) 3-tube F2 forebody (No. 11–13) at $A_{p\,max}$

(4) 3-tube F2 hindbody (No. 17–19) at $A_{U\,max}$

(5) 2-tube F3 foreblade (No. 5–6) at $A_{p\,max}$

(6) 2-tube F3 hindblade (No. 9–10) at $A_{U\,max}$

With respect to the F1 resonance mode, $A_{p\,max}$ is closest to the glottis and $A_{U\,max}$ the lips, whereas at the F2–F3 resonance modes, $A_{p\,max}$ is closest to the lips and $A_{U\,max}$ the glottis.

In Table 2 the F1 frequency of the 2 cm^2 uniform pipe 18 cm in length is 486 Hz. When the hindtract area $A_{p\,max}$ is halved (2 to 1 cm^2) and the foretract area $A_{U\,max}$ is doubled (2 to 4 cm^2), the F1 frequencies are both raised to 571 Hz, . Moreover, when the hindtract area $A_{p\,max}$ is doubled (2 to 4 cm^2) and the foretract area $A_{U\,max}$ is halved (2 to 1 cm^2), the F1 frequencies are both lowered to 390 Hz. As the raised and lowered F1 frequencies are each identical, the proportionality

$$F_1 \propto \frac{1}{(\lambda_1 / 4)} f\left(\frac{A_{U\,max}}{A_{p\,max}}\right)$$

is exactly confirmed for the F1 frequency.

The raised and lowered F2 and F3 frequencies in Table 2 are not identical because the 9-tube tongue body region and the 6-tube blade region are not strictly coextensive with the F2 and F3 quarter wavelengths. Nevertheless, the formant frequency changes continue to conform to the above proportionality. When the area $A_{p\,max}$ of the forebody and foreblade is halved (2 to 1 cm^2) and the area $A_{U\,max}$ of the hindbody and hindblade is doubled (2 to 4 cm^2), the F2 and F3 frequencies are raised. When the area $A_{p\,max}$ of the forebody and foreblade is doubled (2 to 4 cm^2) and the area $A_{U\,max}$ of the hindbody and hindblade is halved (2 to 1 cm^2), the F2 and F3 frequencies are lowered. If the geometric mean $\sqrt{F_a F_b}$ of the two raised or the two lowered formant frequencies is taken, it can be divided by the corresponding formant frequency F_u of the 2 cm^2 uniform pipe to yield the frequency ratio $\sqrt{F_a F_b} / F_u$. The frequency ratio may then be converted into a positive or negative octave change $\log_2(\sqrt{F_a F_b} / F_u)$ per halving or doubling of the 1/3 quarter

wavelength area. For the F1, F2, and F3 formants, the positive octave changes are respectively +0.233, +0.060, and +0.031 per halving of $A_{p\,max}$ or doubling of $A_{U\,max}$; the negative octave changes are respectively –0.318, –0.114, and –0.076 per doubling of $A_{p\,max}$ or halving of $A_{U\,max}$. Because the trading relations between $A_{p\,max}$ and $A_{U\,max}$ are unambiguous, the assumption that the formant frequency F_i is proportional to $(\lambda_i / 4)^{-1} f(A_{U\,max} / A_{p\,max})$ appears to be correct.

3.1.3. Further development of the perturbation analysis and supporting data

In order to account for the numerical values of the positive and negative octave changes in Table 2, the frequency behavior of inductance and capacitance elements will be examined, first in an idealized lumped-parameter pipe with 2 equal-length tubes, then in a distributed-parameter pipe with 27 equal-length tubes. To facilitate the analysis, the following summarizes the frequency behavior of inductance and capacitance elements in electric circuit theory.

The inductive and capacitive impedances Z_L and Z_C of an inductor and a capacitor in a series electric circuit are

$$Z_L = X_L = \omega L\,;\; Z_C = X_C = \frac{1}{\omega C},$$

where X_L and X_C are inductive and capacitive reactances. The angular frequency ω is equal to $2\pi f$; L and C are the electrical inductance and capacitance (Boylestad, 2016, pp. 631–633). The inductive impedance Z_L rises with increasing angular frequency ω while the capacitive impedance Z_C falls with increasing ω. In a series resistor-inductor RL circuit, the amplitudes of the higher frequencies are attenuated if the output voltage is measured across the resistor with resistance R. When $X_L = R$, the low-pass frequency response of the RL circuit is

$$\omega_L = \frac{R}{L},$$

where ω_L is the angular cutoff frequency at which the output voltage is $1/\sqrt{2}$ or –3 dB below the low-frequency maximum of the spectrum (Boylestad, pp. 982–986).

In a series resistor-capacitor RC circuit, the amplitudes of the lower frequencies are reduced if the output voltage is taken across the resistor with resistance R. When $X_C = R$, the high-pass frequency response of the RC circuit is

$$\omega_C = \frac{1}{RC},$$

where ω_C is the angular cutoff frequency at which the output voltage is $1/\sqrt{2}$ or –3 dB below the high-frequency maximum of the spectrum (Boylestad, pp. 987–989).

In a series resistor-inductor-capacitor RLC circuit, the amplitudes of both the higher and lower frequencies are attenuated if the output voltage is taken across the

resistor. When $X_L = X_C = R$ so that $\omega L = 1/\omega C$ and $\omega^2 = 1/LC$, the band-pass frequency response of the RLC circuit becomes

$$\omega_{LC} = \sqrt{\frac{1}{LC}},$$

where the angular resonance frequency ω_{LC} is at the peak of the spectrum under the condition that the resistance *R* is small (Boylestad, pp. 923–924; Kinsler & Frey, 1962, pp. 18–19). Observe that ω_{LC} is also the geometric mean of the low-pass –3 dB cutoff ω_L at the higher frequency side of the resonance curve and the high-pass –3 dB cutoff ω_C at the lower frequency side of the resonance curve (Boylestad, p. 931):

$$\omega_{LC} = (\omega_L)^{1/2}(\omega_C)^{1/2} = \left(\frac{R}{L}\right)^{1/2}\left(\frac{1}{RC}\right)^{1/2} = \left(\frac{1}{LC}\right)^{1/2}.$$

Given that $\omega_{LC} = (\omega_L \omega_C)^{1/2}$, a doubling of ω_L can be exactly offset by a halving of ω_C, and vice versa.

When the electrical inductance *L* and capacitance *C* are set equal to their acoustical analogs, one has

$$L = \frac{\rho l_L}{A_L};\ C = \frac{A_C l_C}{\rho c^2}\ \text{(Kinsler \& Frey, 1962, p. 191),}$$

where ρ is the air density and *c* the sound velocity. A_L and l_L are the area and length of the acoustic inductance tube; A_C and l_C are the area and length of the acoustic capacitance tube. The angular low-pass cutoff frequency ω_L and the angular high-pass cutoff frequency ω_C are therefore:

$$\omega_L = \frac{R}{L} = \frac{R}{\left(\frac{\rho l_L}{A_L}\right)} = \frac{RA_L}{\rho l_L};\ \omega_C = \frac{1}{RC} = \frac{1}{R\left(\frac{A_C l_C}{\rho c^2}\right)} = \frac{\rho c^2}{RA_C l_C},$$

where the acoustical resistance *R* replaces the electrical resistance. If all else remains constant, the low-pass cutoff frequency ω_L is directly proportional to the inductance area A_L while the high-pass cutoff frequency ω_C is inversely proportional to the capacitance area A_C. When the inductance tube and the capacitance tube are of equal length and united to form a pipe of length *l*, then $l_L = l_C = l/2$. Hence the geometric mean ω_{LC} of the low-pass cutoff ω_L and the high-pass cutoff ω_C is

$$\omega_{LC} = (\omega_L)^{1/2}(\omega_C)^{1/2} = \left(\frac{RA_L}{\rho(l/2)}\right)^{1/2}\left(\frac{\rho c^2}{RA_C(l/2}\right)^{1/2} = \left(\frac{4c^2 A_L}{l^2 A_C}\right)^{1/2} = \frac{2c}{l}\left(A_L^{+1}\right)^{1/2}\left(A_C^{-1}\right)^{1/2}.$$

Thus, as was likewise shown in Section 3.1.1, the following f_0 proportionality holds for an idealized lumped-parameter pipe consisting of one tube with inductance area A_L and one tube with capacitance area A_C:

$$f_0 = \frac{\omega_{LC}}{2\pi} \propto \frac{1}{l}\left(A_L^{+1}\right)^{1/2}\left(A_C^{-1}\right)^{1/2}.$$

The product $\left(A_L^{+1}\right)^{1/2}\left(A_C^{-1}\right)^{1/2}$ is the geometric mean of the area of the inductance tube and the reciprocal of the area of the capacitance tube.

The geometric mean of the inductance area and the reciprocal of the capacitance area is of major importance in determining not only the resonance frequency f_0 as seen above, but also the formant frequency F_i as will be demonstrated shortly. Therefore some properties of the geometric mean are reviewed. The geometric mean a^{GM} may be defined as the product:

$$a_{1,N}^{GM} = \prod_{i=1}^{N} a_i^{1/N},$$

where the index *i* ranges over the sequence 1 to *N* for each variable a_i and *N* is the total number of variables. A fraction of the geometric mean $a_{1,N}^{GM}$ is the partial geometric mean:

$$a_{j,k}^{PGM} = \prod_{i=j}^{k} a_i^{1/N},$$

where the indices *j* and *k* are the first and last indices of a substring extracted from the sequence 1 to *N*. A geometric mean can therefore be expressed as the product of two or more partial geometric means. For example,

$$a_{1,N}^{GM} = \prod_{i=1}^{N} a_i^{1/N} = a_{1,k}^{PGM} a_{k+1,N}^{PGM} = \left(\prod_{i=1}^{k} a_i^{1/N}\right)\left(\prod_{i=k+1}^{N} a_i^{1/N}\right).$$

If the numerical values of all the variables in a partial geometric mean are equal to one another, then the exponents of the variables can be added because the base is the same:

$$a_{j,k}^{PGM} = \prod_{i=j}^{k} a_i^{1/N} = a_{j,k}^{(k-j+1)/N} = a_{j,k}^{n/N}, \text{ only if } a_j = \cdots = a_k = a_{j,k}.$$

Thus for a partial geometric mean, the parameter *n* in the fractional exponent n/N characterizes the number of variables having the same value.

As was stated in Section 3.1.1, the Fant perturbation rules for the acoustic reactance elements *L* and *C* of an open-closed distributed-parameter pipe are:

1. An area change at a volume velocity maximum U_{max} has an effect on only the local inductance *L*.
2. An area change at the pressure maximum p_{max} has an effect on only the local capacitance *C*.

The two rules can be illustrated by examining the kinetic energy *KE* and the potential energy *PE* at the extremities of the quarter wavelength $\lambda_i/4$. The kinetic and potential energies may be defined as (cf. Fant & Pauli, 1975):

$$KE = \frac{1}{2} L \Delta l U^2;\ PE = \frac{1}{2} C \Delta l p^2.$$

The per-unit-length acoustic inductance and capacitance are *L* and *C*. The terminating length Δl at the ends of the quarter wavelength is assumed to be short such that the volume velocity *U* and pressure *p* are nearly constant over length Δl. The terminating cylindrical area is $A_{\Delta l}$. Substituting $\rho / A_{\Delta l}$ for *L* and $A_{\Delta l} / \rho c^2$ for *C* (Section 3.1.1), one obtains

$$KE = \frac{1}{2}\frac{\rho}{A_{\Delta l}}\Delta l U^2;\ PE = \frac{1}{2}\frac{A_{\Delta l}}{\rho c^2}\Delta l p^2.$$

It was pointed out at the beginning of Section 3.1.1 that the volume velocity is a reciprocal function of the pressure. Because the volume velocity maximum U_{max} corresponds to a pressure minimum p_{min}, the stored potential energy *PE* in the capacitance $A_{\Delta l} / \rho c^2$ is negligibly small so that the influence of the capacitive area $A_{\Delta l}$ is also small. Hence a change in area $A_{\Delta l}$ at the volume velocity maximum U_{max} has an effect on only the inductance $\rho / A_{\Delta l}$ and its large stored kinetic energy *KE*. Conversely, because the pressure maximum p_{max} corresponds to a volume velocity minimum U_{min}, the stored kinetic energy *KE* in the inductance $\rho / A_{\Delta l}$ is negligibly small so that the influence of the inductive area $A_{\Delta l}$ is also small. Therefore a change in area $A_{\Delta l}$ at the pressure maximum p_{max} has an effect on only the capacitance $A_{\Delta l} / \rho c^2$ and its large stored potential energy *PE*.

In the idealized lumped-parameter pipe, each tube of air is always associated with the same type of reactance element, either an acoustic inductance *L* or an acoustic capacitance *C*. In the distributed-parameter pipe, the degree to which a tube of air acts like *L* or *C* depends on the location of the tube relative to the volume velocity and pressure maxima of the quarter wavelength $\lambda_t / 4$. As demonstrated in the preceding paragraph, when its area is perturbed, a tube of air functions like an acoustic inductance *L* at the volume velocity maximum and like an acoustic capacitance *C* at the pressure maximum.

Four area types are distinguished for a distributed-parameter pipe with 27 equal-length tubes:

1. The inductive area type $A^{+1}_{U\,max}$ or simply A^{+1}. It is indicated by a tube area with the positive exponent +1, cf. A^{+1}_L in $f_0 \propto l^{-1}\left(A^{+1}_L\right)^{1/2}\left(A^{-1}_C\right)^{1/2}$. An inductive area tube lies inside the quarter wavelength $\lambda_t / 4$ at the volume velocity maximum.
2. The capacitive area type $A^{-1}_{p\,max}$ or simply A^{-1}. It is indicated by a tube area with the negative exponent –1, cf. A^{-1}_C in $f_0 \propto l^{-1}\left(A^{+1}_L\right)^{1/2}\left(A^{-1}_C\right)^{1/2}$. A capacitive area tube lies inside the quarter wavelength $\lambda_t / 4$ at the pressure maximum.
3. The neutral area type A^0 (where $A^0 = 1$). It is indicated by a tube area with the exponent 0. A neutral area tube lies inside the quarter wavelength $\lambda_t / 4$ but

neither at a volume velocity maximum nor at a pressure maximum. Therefore A^0 has no effect on the formant frequency F_i.

4. The constant area type A^c. It is designated by a tube area with the superscript *c*. A constant area tube lies outside the selected quarter wavelength $\lambda_i / 4$. As A^c is a constant, it can be dropped from the formant frequency proportionality because the proportionality relation is invariant with respect to multiplication by a constant.

In the preceding Section 3.1.2, it was determined that the 1/3 quarter wavelengths for the F1 formant frequency are:

(1) 9-tube F1 hindtract at $A_{p\,max}$ (tubes 19–27)

(2) 9-tube F1 foretract at $A_{U\,max}$ (tubes 1–9)

Since the resonance frequency f_0 is directly proportional to the geometric mean of a sequence of 2 area types $\left(A_L^{+1}\right)^{1/2}\left(A_C^{-1}\right)^{1/2}$, the F1 frequency is predicted to be directly proportional to the geometric mean of the following sequence of 27 area types (cf. $f_0 \propto l^{-1}\left(A_L^{+1}\right)^{1/2}\left(A_C^{-1}\right)^{1/2}$ and $F_i \propto (\lambda_i / 4)^{-1} f(A_{U\,max} / A_{p\,max})$ in Section 3.1.2):

$$F_1 \propto \frac{1}{\lambda_1 / 4}(A_1^{+1})^{1/27}(A_2^{+1})^{1/27}(A_3^{+1})^{1/27}(A_4^{+1})^{1/27}(A_5^{+1})^{1/27}(A_6^{+1})^{1/27}(A_7^{+1})^{1/27}(A_8^{+1})^{1/27}(A_9^{+1})^{1/27} \times$$
$$(A_{10}^{0})^{1/27}(A_{11}^{0})^{1/27}(A_{12}^{0})^{1/27}(A_{13}^{0})^{1/27}(A_{14}^{0})^{1/27}(A_{15}^{0})^{1/27}(A_{16}^{0})^{1/27}(A_{17}^{0})^{1/27}(A_{18}^{0})^{1/27} \times$$
$$(A_{19}^{-1})^{1/27}(A_{20}^{-1})^{1/27}(A_{21}^{-1})^{1/27}(A_{22}^{-1})^{1/27}(A_{23}^{-1})^{1/27}(A_{24}^{-1})^{1/27}(A_{25}^{-1})^{1/27}(A_{26}^{-1})^{1/27}(A_{27}^{-1})^{1/27}.$$

Recall that when the variables of a partial geometric mean are all equal, their exponents can be summed:

$$(A_1^{+1})^{1/27}(A_2^{+1})^{1/27}(A_3^{+1})^{1/27}(A_4^{+1})^{1/27}(A_5^{+1})^{1/27}(A_6^{+1})^{1/27}(A_7^{+1})^{1/27}(A_8^{+1})^{1/27}(A_9^{+1})^{1/27} = (A_{U\,max}^{+1})^{9/27}$$
$$(A_{10}^{0})^{1/27}(A_{11}^{0})^{1/27}(A_{12}^{0})^{1/27}(A_{13}^{0})^{1/27}(A_{14}^{0})^{1/27}(A_{15}^{0})^{1/27}(A_{16}^{0})^{1/27}(A_{17}^{0})^{1/27}(A_{18}^{0})^{1/27} = (A^0)^{9/27}$$
$$(A_{19}^{-1})^{1/27}(A_{20}^{-1})^{1/27}(A_{21}^{-1})^{1/27}(A_{22}^{-1})^{1/27}(A_{23}^{-1})^{1/27}(A_{24}^{-1})^{1/27}(A_{25}^{-1})^{1/27}(A_{26}^{-1})^{1/27}(A_{27}^{-1})^{1/27} = (A_{p\,max}^{-1})^{9/27}$$

The F_1 proportionality is then

$$F_1 \propto \frac{1}{\lambda_1 / 4}(A_{U\,max}^{+1})^{9/27} \times (A^0)^{9/27} \times (A_{p\,max}^{-1})^{9/27}$$

$$\text{or } F_1 \propto \frac{1}{\lambda_1 / 4}\left(\frac{A_{U\,max}}{A_{p\,max}}\right)^{9/27},$$

since $(A^0)^{9/27} = 1$. The term $(A_{U\,max} / A_{p\,max})^{9/27}$ constitutes a partial geometric mean because it is a fraction of the geometric mean. When either the hindtract area $A_{p\,max}$ is halved or the foretract area $A_{U\,max}$ is doubled, F_1 increases by a factor of $2^{9/27}$ or a positive change of +1/3 (+0.333) octave ($= \log_2(2^{1/3})$). When either the hindtract area $A_{p\,max}$ is doubled or the foretract area $A_{U\,max}$ is halved, F_1 decreases by a factor of $2^{-9/27}$ ($= (2^{-1})^{9/27}$) or a negative change of –1/3 (–0.333) octave ($= \log_2(2^{-1/3})$). For the F1

frequency in Table 2, the positive octave change is +0.233 per halving of $A_{p\,max}$ or doubling of $A_{U\,max}$; the negative octave change is –0.318 per doubling of $A_{p\,max}$ or halving of $A_{U\,max}$. Thus the F1 frequency results given by the theoretical perturbation analysis match fairly well those calculated by lossless FDVT models in Table 2.

The 1/3 quarter wavelengths for the F2 formant frequency are (Section 3.1.2):

(1) 3-tube F2 forebody at $A_{p\,max}$ (tubes 11–13)

(2) 3-tube F2 hindbody at $A_{U\,max}$ (tubes 17–19)

The F2 frequency is proportional to the geometric mean of the following sequence of 27 area types:

$$F_2 \propto \frac{1}{\lambda_2 / 4}(A_1^c)^{1/27}(A_2^c)^{1/27}(A_3^c)^{1/27}(A_4^c)^{1/27}(A_5^c)^{1/27}(A_6^c)^{1/27}(A_7^c)^{1/27}(A_8^c)^{1/27}(A_9^c)^{1/27} \times$$
$$(A_{10}^c)^{1/27}(A_{11}^{-1})^{1/27}(A_{12}^{-1})^{1/27}(A_{13}^{-1})^{1/27}(A_{14}^0)^{1/27}(A_{15}^0)^{1/27}(A_{16}^0)^{1/27}(A_{17}^{+1})^{1/27}(A_{18}^{+1})^{1/27} \times$$
$$(A_{19}^{+1})^{1/27}(A_{20}^c)^{1/27}(A_{21}^c)^{1/27}(A_{22}^c)^{1/27}(A_{23}^c)^{1/27}(A_{24}^c)^{1/27}(A_{25}^c)^{1/27}(A_{26}^c)^{1/27}(A_{27}^c)^{1/27}.$$

The areas of every constant area type A^c are set equal to one another. Summing the exponents of similar variables, one obtains

$$(A_1^c)^{1/27}(A_2^c)^{1/27}(A_3^c)^{1/27}(A_4^c)^{1/27}(A_5^c)^{1/27}(A_6^c)^{1/27}(A_7^c)^{1/27}(A_8^c)^{1/27}(A_9^c)^{1/27}(A_{10}^c)^{1/27} = (A^c)^{10/27}$$
$$(A_{11}^{-1})^{1/27}(A_{12}^{-1})^{1/27}(A_{13}^{-1})^{1/27} = (A_{p\,max}^{-1})^{3/27}$$
$$(A_{14}^0)^{1/27}(A_{15}^0)^{1/27}(A_{16}^0)^{1/27} = (A^0)^{3/27}$$
$$(A_{17}^{+1})^{1/27}(A_{18}^{+1})^{1/27}(A_{19}^{+1})^{1/27} = (A_{U\,max}^{+1})^{3/27}$$
$$(A_{20}^c)^{1/27}(A_{21}^c)^{1/27}(A_{22}^c)^{1/27}(A_{23}^c)^{1/27}(A_{24}^c)^{1/27}(A_{25}^c)^{1/27}(A_{26}^c)^{1/27}(A_{27}^c)^{1/27} = (A^c)^{8/27}$$

The F_2 proportionality is then

$$F_2 \propto \frac{1}{\lambda_2 / 4}(A^c)^{10/27} \times (A_{p\,max}^{-1})^{3/27} \times (A^0)^{3/27} \times (A_{U\,max}^{+1})^{3/27} \times (A^c)^{8/27}.$$

Because the relation of proportionality is invariant with respect to multiplication by a constant, the constant partial geometric means $(A^c)^{10/27}$, $(A^c)^{8/27}$, and $(A^0)^{3/27}$ $(=1)$ can be dropped to yield:

$$F_2 \propto \frac{1}{\lambda_2 / 4}(A_{p\,max}^{-1})^{3/27} \times (A_{U\,max}^{+1})^{3/27}$$

$$\text{or } F_2 \propto \frac{1}{\lambda_2 / 4}\left(\frac{A_{U\,max}}{A_{p\,max}}\right)^{3/27}.$$

When either the forebody area $A_{p\,max}$ is halved or the hindbody area $A_{U\,max}$ is doubled, F_2 increases by a factor of $2^{3/27}$ or a positive change of +1/9 (+0.111) octave $(=\log_2(2^{1/9}))$. When either the forebody area $A_{p\,max}$ is doubled or the hindbody area $A_{U\,max}$ is halved, F_2 decreases by a factor of $2^{-3/27}$ or a negative change of –1/9 (–0.111) octave $(=\log_2(2^{-1/9}))$. For the F2 frequency in Table 2, the positive octave change is +0.060 per halving of $A_{p\,max}$ or doubling of $A_{U\,max}$; the negative octave change is –0.114 per

doubling of $A_{p\,max}$ or halving of $A_{U\,max}$. Note the near equality between the theoretical and experimental negative octave changes: –0.111 vs. –0.114.

At the F2 resonance mode, the 27-tube pipe comprises exactly three quarter wavelengths 9 tubes long. Hence for the F2 formant frequency, there are six 1/3 quarter wavelengths each of which is 3 tubes long. Thus it is possible to check the validity of the perturbation analysis using a more complex test case than the F2 forebody and hindbody which span only two 1/3 quarter wavelengths. The six 1/3 quarter wavelengths are:

(1) 3-tube F2 segment at $A_{U\,max}$ (tubes 1–3)

(2) 3-tube F2 segment at $A_{p\,max}$ (tubes 7–9)

(3) 3-tube F2 segment at $A_{p\,max}$ (tubes 10–12)

(4) 3-tube F2 segment at $A_{U\,max}$ (tubes 16–18)

(5) 3-tube F2 segment at $A_{U\,max}$ (tubes 19–21)

(6) 3-tube F2 segment at $A_{p\,max}$ (tubes 25–27)

The three 3-tube $A_{p\,max}$ segments are discontinuous because (6) is spatially separated from (2) and (3); the three 3-tube $A_{U\,max}$ segments are discontinuous because (1) is spatially separated from (4) and (5). The three $A_{p\,max}$ segments are assumed to have the same area while the three $A_{U\,max}$ segments likewise have the same area. The F2 frequency is proportional to the geometric mean of the following sequence of 27 area types:

$$F_2 \propto \frac{1}{\lambda_2 / 4}(A_1^{+1})^{1/27}(A_2^{+1})^{1/27}(A_3^{+1})^{1/27}(A_4^{0})^{1/27}(A_5^{0})^{1/27}(A_6^{0})^{1/27}(A_7^{-1})^{1/27}(A_8^{-1})^{1/27}(A_9^{-1})^{1/27} \times$$
$$(A_{10}^{-1})^{1/27}(A_{11}^{-1})^{1/27}(A_{12}^{-1})^{1/27}(A_{13}^{0})^{1/27}(A_{14}^{0})^{1/27}(A_{15}^{0})^{1/27}(A_{16}^{+1})^{1/27}(A_{17}^{+1})^{1/27}(A_{18}^{+1})^{1/27} \times$$
$$(A_{19}^{+1})^{1/27}(A_{20}^{+1})^{1/27}(A_{21}^{+1})^{1/27}(A_{22}^{0})^{1/27}(A_{23}^{0})^{1/27}(A_{24}^{0})^{1/27}(A_{25}^{-1})^{1/27}(A_{26}^{-1})^{1/27}(A_{27}^{-1})^{1/27}.$$

Adding the exponents of similar variables, one finds

$$(A_1^{+1})^{1/27}(A_2^{+1})^{1/27}(A_3^{+1})^{1/27} = (A_{U\,max}^{+1})^{3/27}$$
$$(A_4^{0})^{1/27}(A_5^{0})^{1/27}(A_6^{0})^{1/27} = (A^{0})^{3/27}$$
$$(A_7^{-1})^{1/27}(A_8^{-1})^{1/27}(A_9^{-1})^{1/27} = (A_{p\,max}^{-1})^{3/27}$$
$$(A_{10}^{-1})^{1/27}(A_{11}^{-1})^{1/27}(A_{12}^{-1})^{1/27} = (A_{p\,max}^{-1})^{3/27}$$
$$(A_{13}^{0})^{1/27}(A_{14}^{0})^{1/27}(A_{15}^{0})^{1/27} = (A^{0})^{3/27}$$
$$(A_{16}^{+1})^{1/27}(A_{17}^{+1})^{1/27}(A_{18}^{+1})^{1/27} = (A_{U\,max}^{+1})^{3/27}$$
$$(A_{19}^{+1})^{1/27}(A_{20}^{+1})^{1/27}(A_{21}^{+1})^{1/27} = (A_{U\,max}^{+1})^{3/27}$$
$$(A_{22}^{0})^{1/27}(A_{23}^{0})^{1/27}(A_{24}^{0})^{1/27} = (A^{0})^{3/27}$$
$$(A_{25}^{-1})^{1/27}(A_{26}^{-1})^{1/27}(A_{27}^{-1})^{1/27} = (A_{p\,max}^{-1})^{3/27}$$

For the 3-tube $A_{U\,max}$ segments and 3-tube $A_{p\,max}$ segments , the F_2 proportionality is then

$$F_2 \propto \frac{1}{\lambda_2 / 4}(A_{U\,max}^{+1})^{3/27} \times (A^{0})^{3/27} \times (A_{p\,max}^{-1})^{3/27} \times$$

$$(A_{p\,max}^{-1})^{3/27} \times (A^0)^{3/27} \times (A_{U\,max}^{+1})^{3/27} \times$$
$$(A_{U\,max}^{+1})^{3/27} \times (A^0)^{3/27} \times (A_{p\,max}^{-1})^{3/27}.$$

Adding the exponents of like variables once again, the F_2 proportionality becomes

$$F_2 \propto \frac{1}{\lambda_2 / 4} (A_{U\,max}^{+1})^{9/27} \times (A^0)^{9/27} \times (A_{p\,max}^{-1})^{9/27}$$

$$\text{or } F_2 \propto \frac{1}{\lambda_2 / 4} \left(\frac{A_{U\,max}}{A_{p\,max}} \right)^{9/27},$$

since $(A^0)^{9/27} = 1$. When the area of the three 3-tube $A_{p\,max}$ segments is halved or the area of the three 3-tube $A_{U\,max}$ segments is doubled, F_2 increases by a factor of $2^{9/27}$ or a positive change of +1/3 (+0.333) octave ($= \log_2(2^{1/3})$). When the area of the three 3-tube $A_{p\,max}$ segments is doubled or the area of the three 3-tube $A_{U\,max}$ segments is halved, F_2 decreases by a factor of $2^{-9/27}$ or a negative change of –1/3 (–0.333) octave ($= \log_2(2^{-1/3})$).

To verify the correctness of these theoretical F2 octave changes, a lossless FDVT model computes the F2 frequency of an open-closed pipes 18 cm long with a default area of 2 cm^2 as in Section 3.1.2. The F2 frequency of the 2 cm^2 uniform pipe is 1458 Hz. When the area of the three $A_{p\,max}$ segments is halved (2 to 1 cm^2) and the area of the three $A_{U\,max}$ segments is doubled (2 to 4 cm^2), the F2 frequencies are both raised to 1714 Hz. Furthermore, when the area of the three $A_{p\,max}$ segments is doubled (2 to 4 cm^2) and the area of the three $A_{U\,max}$ segments is halved (2 to 1 cm^2), the F2 frequencies are both lowered to 1171 Hz. Thus the raised and lowered F2 frequencies are each identical. The positive octave change is +0.233 per halving of $A_{p\,max}$ or doubling of $A_{U\,max}$ ($\log_2(1714/1458) = +0.2334$); the negative octave change is –0.316 per doubling of $A_{p\,max}$ or halving of $A_{U\,max}$ ($\log_2(1171/1458) = -0.3162$). Like the F1 formant with the same partial geometric mean $(A_{U\,max} / A_{p\,max})^{9/27}$, the theoretical F2 octave changes correspond fairly well to those calculated by the FDVT models: +0.333 vs. +0.233 and –0.333 vs. –0.316. The perturbation analysis can therefore account for the F2 octave changes of a pipe in which the area of three discontinuous $A_{p\,max}$ segments and the area of three discontinuous $A_{U\,max}$ segments are varied independently. Given the complexity of the test case, perturbation analysis using the partial geometric mean appears to be the only plausible way of explaining these results.

The 1/3 quarter wavelengths for the F3 formant frequency are (Section 3.1.2):

(1) The 2-tube F3 foreblade at $A_{p\,max}$ (tubes 5–6)

(2) The 2-tube F3 hindblade at $A_{U\,max}$ (tubes 9–10)

The F3 frequency is proportional to the geometric mean of the following sequence of 27 area types:

$$F_3 \propto \frac{1}{\lambda_3 / 4}(A_1^c)^{1/27}(A_2^c)^{1/27}(A_3^c)^{1/27}(A_4^c)^{1/27}(A_5^{-1})^{1/27}(A_6^{-1})^{1/27}(A_7^0)^{1/27}(A_8^0)^{1/27}(A_9^{+1})^{1/27} \times$$
$$(A_{10}^{+1})^{1/27}(A_{11}^c)^{1/27}(A_{12}^c)^{1/27}(A_{13}^c)^{1/27}(A_{14}^c)^{1/27}(A_{15}^c)^{1/27}(A_{16}^0)^{1/27}(A_{17}^c)^{1/27}(A_{18}^c)^{1/27} \times$$
$$(A_{19}^c)^{1/27}(A_{20}^c)^{1/27}(A_{21}^c)^{1/27}(A_{22}^c)^{1/27}(A_{23}^c)^{1/27}(A_{24}^c)^{1/27}(A_{25}^c)^{1/27}(A_{26}^c)^{1/27}(A_{27}^c)^{1/27}.$$

The areas of each constant area type A^c are set equal to one another. Summing the exponents of similar variables, one has

$$(A_1^c)^{1/27}(A_2^c)^{1/27}(A_3^c)^{1/27}(A_4^c)^{1/27} = (A^c)^{4/27}$$
$$(A_5^{-1})^{1/27}(A_6^{-1})^{1/27} = (A_{p\,max}^{-1})^{2/27}$$
$$(A_7^0)^{1/27}(A_8^0)^{1/27} = (A^0)^{2/27}$$
$$(A_9^{+1})^{1/27}(A_{10}^{+1})^{1/27} = (A_{U\,max}^{+1})^{2/27}$$
$$(A_{11}^c)^{1/27}(A_{12}^c)^{1/27}(A_{13}^c)^{1/27}(A_{14}^c)^{1/27}(A_{15}^c)^{1/27}(A_{16}^c)^{1/27}(A_{17}^c)^{1/27}(A_{18}^c)^{1/27}(A_{19}^c)^{1/27} \times$$
$$(A_{20}^c)^{1/27}(A_{21}^c)^{1/27}(A_{22}^c)^{1/27}(A_{23}^c)^{1/27}(A_{24}^c)^{1/27}(A_{25}^c)^{1/27}(A_{26}^c)^{1/27}(A_{27}^c)^{1/27} = (A^c)^{17/27}$$

The F_3 proportionality is then

$$F_3 \propto \frac{1}{\lambda_3 / 4}(A^c)^{4/27} \times (A_{p\,max}^{-1})^{2/27} \times (A^0)^{2/27} \times (A_{U\,max}^{+1})^{2/27} \times (A^c)^{17/27}.$$

The constant partial geometric means $(A^c)^{4/27}$, $(A^0)^{2/27}$ $(=1)$, and $(A^c)^{17/27}$ may be dropped from the formant frequency proportionality to give:

$$F_3 \propto \frac{1}{\lambda_3 / 4}(A_{p\,max}^{-1})^{2/27} \times (A_{U\,max}^{+1})^{2/27}$$

$$\text{or } F_3 \propto \frac{1}{\lambda_3 / 4}\left(\frac{A_{U\,max}}{A_{p\,max}}\right)^{2/27}.$$

When either the foreblade area $A_{p\,max}$ is halved or the hindblade area $A_{U\,max}$ is doubled, F_3 increases by a factor of $2^{2/27}$ or a positive change of +2/27 (+0.074) octave $(=\log_2(2^{2/27}))$. When either the foreblade area $A_{p\,max}$ is doubled or the hindblade area $A_{U\,max}$ is halved, F_3 decreases by a factor of $2^{-2/27}$ or a negative change of –2/27 (–0.074) octave $(=\log_2(2^{-2/27}))$. For the F3 frequency in Table 2, the positive octave change is +0.031 per halving of $A_{p\,max}$ or doubling of $A_{U\,max}$; the negative octave change is –0.076 per doubling of $A_{p\,max}$ or halving of $A_{U\,max}$.

Summarizing the theoretical and experimental comparisons above, it is clear that the octave changes obtained through the perturbation analysis are often close to those calculated by the FDVT models in Table 2. In particular, the negative octave changes are nearly identical seeing that the theoretical values are F1: –0.333, F2: –0.111, F3: –0.074 and the experimental values are F1: –0.318, F2: –0.114, F3: –0.076. As for the positive octave changes, the theoretical values are F1: +0.333, F2: +0.111, F3: +0.074 while the experimental values are F1: +0.233, F2: +0.060, F3: +0.031. Although the theoretical and experimental values mostly agree for the positive F1 octave changes: +0.333 vs. +0.233,

the experimental values for the positive F2 and F3 octave changes are substantially smaller than the theoretical ones. For an explanation of the smaller positive F2 and F3 octave changes, see Appendix C. The perturbation analysis can also model the F2 octave changes of an open-closed pipe in which the area of three discontinuous $A_{p\,max}$ segments and the area of three discontinuous $A_{U\,max}$ segments are varied separately. Like the F1 octave changes with the same partial geometric mean $(A_{U\,max} / A_{p\,max})^{9/27}$, the theoretical and experimental F2 octave changes are +0.333 vs. +0.233 and –0.333 vs. –0.316. In this complex test case, perturbation analysis using the partial geometric mean appears to be the only suitable way to account for the observed results.

As the number of tubes *n* in the fractional exponent $n/27$ of the partial geometric mean $(A_{U\,max} / A_{p\,max})^{n/27}$ grows smaller, identical area changes of $A_{p\,max}$ and $A_{U\,max}$ have progressively less effect on formant frequency F_i. For example, the fractional exponents $9/27$, $3/27$, $2/27$ produce octave changes of ± 0.333, ± 0.111, ± 0.074 for each doubling or halving of the area ratio $A_{U\,max} / A_{p\,max}$. When *n* is small, the partial geometric mean is a small fraction of the geometric mean of the inductive and capacitive area types $A_{U\,max}^{+1}$ and $A_{p\,max}^{-1}$. Hence there is relatively less inductance and capacitance available for modification by area change. Consequently, F_i displays narrower frequency shifts with smaller *n*.

In Section 3.1.1, it was hypothesized that the formant frequency proportionality is

$$F_i \propto \frac{1}{(\lambda_i / 4)} f\left(\frac{A_{U\,max}}{A_{p\,max}}\right),$$

where $A_{U\,max}$ is the area at the volume velocity maximum, $A_{p\,max}$ the area at the pressure maximum, and $f(A_{U\,max} / A_{p\,max})$ a dimensionless function of the area ratio $A_{U\,max} / A_{p\,max}$. The proportionality is based on (a) an idealized model of the lumped-parameter pipe and (b) Fant's rules for impedance perturbations in a distributed-parameter pipe. The function

$$f\left(\frac{A_{U\,max}}{A_{p\,max}}\right)$$

was not further specified. In this section, it was found that for the open-closed pipe, the function is equal to the partial geometric mean:

$$f\left(\frac{A_{U\,max}}{A_{p\,max}}\right) = \left(\frac{A_{U\,max}}{A_{p\,max}}\right)^{n/N}.$$

The area ratio $A_{U\,max} / A_{p\,max}$ is raised to the fractional exponent n/N. The parameter *n* is the number of tubes spanning the 1/3 quarter wavelengths of formant frequency F_i and $N = 27$, the total number of equal-length tubes. The general proportionality for formant frequency F_i can therefore be stated as

$$F_i \propto \frac{1}{(\lambda_i / 4)} \left(\frac{A_{U\,max}}{A_{p\,max}} \right)^{a/b},$$

where the fractional exponent a / b is the simplification or the experimentally determined value of n / N. For instance, if $n / N = 9 / 27$, then one has the reduced fraction $a / b = 1 / 3$ after simplification. The formant frequency proportionality demonstrates that the perturbation analysis can be expressed rather concisely when three conditions are fulfilled: (i) the area of every tube in the 1/3 quarter wavelength at the volume velocity maximum is equal to $A_{U\,max}$; (ii) the area of every tube in the 1/3 quarter wavelength at the pressure maximum is equal to $A_{p\,max}$; (iii) the area of every tube outside the selected quarter wavelength is equal to a constant (see Fant, 1975, Stevens, 1998, pp. 148–152, McGowan, 2018, and Section 3.3 below for a related formulation of perturbation analysis cast in terms of kinetic and potential energies).

3.2. Articulator regions: anatomical boundaries

Determination of the anatomical boundaries of the four active articulators (lips, blade, tongue body, tongue root) has been a long-standing problem in phonetic science. The locations 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 International Phonetic Association, 1999, p. 8). For example, the IPA chart distinguishes three blade locations (passive articulators: front teeth, alveolar ridge, postalveolar area) and three tongue body locations (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 Fig. 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 and pressure maxima 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 y.o.	37.47%
women	19–25 y.o.	32.04%
children	11–12 y.o.	32.42%
children	7–8 y.o.	29.69%
children	2–4 y.o.	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% found in Section 3.1.2. 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, Walters, Monaghan, & Patterson, 2009, p. 2379, who reach the same conclusion).

3.3. Articulator regions: the concept of acoustic partitioning

On the basis of sensitivity functions, Mrayati, Carré, and Guérin introduced a proposal in 1988 for the acoustic partitioning of the vocal tract—the concept that has inspired the present work. The sensitivity S of formant frequency F_i to a change in vocal tract area is defined as the difference between the kinetic energy and potential energy at the point of area change divided by the total energy: $(KE - PE)/TE$ (p. 259). The authors noted that the sensitivity functions of an open-closed uniform pipe are antisymmetric as well as sinusoidal and thus "have an odd number of zero-crossings, namely 1, 3, and 5 for the first three formants (p. 261)". They then distinguished eight vocal tract regions whose boundaries are set at the zero-crossings of the sensitivity functions. A recent exposition of the proposal (Carré, Divenyi, & Mrayati, 2017, p. 106) designates these regions as R_1 glottal, R_3 pharyngeal, R_4 uvular, R_5 velar, R_6 coronal, R_8 labial, whereas R_2 (lower pharynx) and R_7 (teeth) are "considered as intermediate regions the acoustic effects of which on the formant are small" (p. 104). In an extensive critical review Boë and Perrier (1990, p. 227) observed that the proposal lacks articulatory constraints that could reduce the number of parameters controlling the eight regions (but

see the procedure in Carré, Divenyi, & Mrayati, 2017, p. 91, which addresses this shortcoming for the case of tongue body movement).

The eight-region partition of the vocal tract relies on sensitivity functions. The articulator-region partition elaborated in this study is based on quarter wavelengths. The example of coronals can briefly illustrate how the articulator-region partition yields a more accurate analysis than the eight-region partition. In the eight-region proposal, the lengths of the coronal R_6 region and the intermediate R_7 (teeth) region are respectively $2l/15$ and $l/15$, where l is the length of an open-closed uniform pipe (Mrayati, Carré, & Guérin, 1988, p. 265; Carré, Divenyi, & Mrayati, 2017, p. 81). If the regions R_6 and R_7 are joined, the resulting region is $3l/15$ long or one-fifth the pipe length l. Inspection of Fig. 1 shows that, for the third formant (F3), the length of the combined region equals one quarter wavelength: a volume velocity minimum (pressure maximum) at four-fifths l from the closed end of the pipe and a volume velocity maximum at three-fifths l from the closed end. Because a constriction near the pressure maximum at four-fifths l leads to a significant F3 increase, the acoustic effect of the intermediate R_7 (teeth) region is not at all negligible for the third formant (cf. the F3 frequency of the foreblade pressure maximum in Table 2). This is inconsistent with Carré, Divenyi, and Mrayati's statement on the relative unimportance of the intermediate region R_7. Consequently, the whole of their R_6 and R_7 regions should be designated as a coronal or blade region instead of R_6 alone.

4. Formant bandwidth: introduction

To examine the role of formant bandwidths in acoustic-articulatory relations, it is helpful to consider the quality or amplification factor Q of a series resistor-inductor-capacitor RLC electric circuit at the resonance frequency f_0. The term was first defined as the ratio of inductive reactance to resistance $Q = 2\pi f_0 L / R$, where $2\pi f_0 L$ is the inductive reactance X_L and R is the resistance (Green, 1955). At resonance frequency f_0 the inductive reactance X_L equals the capacitive reactance $X_L = X_C$, where $X_C = 1/2\pi f_0 C$ (Boylestad, 2016, p. 924). Hence the term also came to mean the ratio of capacitive reactance to resistance $Q = X_C / R = 1/2\pi f_0 CR$, given that the driving input voltage E is multiplied by Q across both reactive elements L and C: $EQ = V_L = V_C$ (Boylestad, pp. 926–927). Another way of defining the quality factor is the quotient formed between the resonance frequency and the half-power bandwidth: $Q = f_0 / B$, where $B = f_2 - f_1$; f_1 and f_2 are the low and high cutoff frequencies at which the output is $1/\sqrt{2}$ or –3 dB below the peak of the spectrum. If $Q \geq 10$, the resonance shape is symmetrical with respect to the resonance frequency f_0 (Boylestad, pp. 929–931). As the bandwidth B equals $f_0 / Q = f_0 R / 2\pi f_0 L = R / 2\pi L$, it may be more broadly expressed as

$$B = \frac{\frac{1}{2} R I_{\max}^2}{2\pi \times \frac{1}{2} L I_{\max}^2} = \frac{\text{average dissipated power}}{2\pi \times \text{stored energy}},$$

since in a series circuit the same current $I_{\max}$ flows through each of the elements R and L.

Fant and Pauli (1975), followed by Mrayati and Carré (1976), applied this general bandwidth expression to a resonance mode F in their transmission line models of the vocal tract. The kinetic and potential energies of tube *n* with length $l(n)$ are

$$KE(n) = \frac{1}{2} L(n) l(n) U^2(n);\ PE(n) = \frac{1}{2} C(n) l(n) p^2(n),$$

where $U(n)$, $p(n)$, $L(n)$, $C(n)$ are the volume velocity, pressure, per-unit-length acoustic inductance, and per-unit-length acoustic capacitance of tube *n*. At resonance the kinetic and potential energy of the *N* tubes of the pipe are equal:

$$\sum_{n=1}^{N} KE(n) = \sum_{n=1}^{N} PE(n) = \frac{1}{2} TE,$$

where *TE* is the total stored energy. The power lost to viscosity R_{visc} in a single tube *n* is $P_{visc}(n) = R_{visc}(n) l(n) U^2(n)$, whereas the power losses due to heat conduction G_{heat} and wall damping G_{wall} are $P_{heat}(n) = G_{heat}(n) l(n) p^2(n)$ and $P_{wall}(n) = G_{wall}(n) l(n) p^2(n)$. The power loss due to radiation is $P_{rad} = R_{rad} U_0^2$, where U_0 is the output volume velocity and R_{rad} the radiation resistance (Section 2):

$$R_{rad} = \frac{\rho c}{A_0} \left[1 - \frac{J_1(2kr)}{kr} \right].$$

The terminating lip area is $A_0 = \pi r^2$; the terminating lip radius is *r*. The radiation resistance R_{rad} takes the form of the characteristic acoustic resistance of air in a pipe $R_{char} = \rho c / A_0$ multiplied by a dimensionless radiation resistance (Flanagan, 1972, p. 36). In view of the general power-energy relation above, the bandwidth B at formant frequency F can therefore be written as

$$B = \frac{P_{rad} + \sum_{n=1}^{N} P_{visc}(n) + \sum_{n=1}^{N} P_{heat}(n) + \sum_{n=1}^{N} P_{wall}(n)}{2\pi \times TE},$$

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. When calculated by the 27-tube FDVT model, the individual bandwidths of a 4 cm^2 and a 1 cm^2 uniform pipe 18 cm long are presented in Appendix A. Remark the very close agreement between the two ways of determining overall bandwidth: (i) the measured half-power

bandwidth $B = f_2 - f_1$ and (ii) the power-energy bandwidth B_{sum} introduced by Fant and Pauli. The radiation bandwidth B_{rad} is the largest contributor to the overall bandwidth especially for the wider 4 cm^2 pipe. Moreover, B_{rad} grows progressively larger as the formant frequency increases. The viscous and heat conduction bandwidths B_{visc} and B_{heat} make the next largest contribution especially for the narrower 1 cm^2 pipe. Both B_{visc} and B_{heat} grow with increasing formant frequency, but more slowly compared to the radiation bandwidth B_{rad}. The heat conduction bandwidth B_{heat} is about half the viscous bandwidth B_{visc}. The wall damping bandwidth B_{wall} remains rather small for F1 and is negligible for the higher formants (F2–F4).

In a uniform pipe the power losses and stored energy of each of the *N* tubes are the same. Accordingly, the viscous bandwidth for the uniform pipe is

$$B_{visc} = \frac{\frac{1}{2}P_{visc}}{2\pi \times \frac{1}{2}TE} = \frac{\frac{1}{2}R_{visc}lU^2}{2\pi \times \frac{1}{2}LlU^2} = \frac{R_{visc}}{2\pi \times L} = \frac{R_{visc}}{2\pi \times (\rho / A)},$$

where the stored energy in the denominator equals the kinetic energy $KE = LlU^2 / 2$; A and l are the area and length of the pipe. Likewise, the heat conduction bandwidth is

$$B_{heat} = \frac{\frac{1}{2}P_{heat}}{2\pi \times \frac{1}{2}TE} = \frac{\frac{1}{2}G_{heat}lp^2}{2\pi \times \frac{1}{2}Clp^2} = \frac{G_{heat}}{2\pi \times C} = \frac{G_{heat}}{2\pi \times (A / \rho c^2)},$$

where the stored energy equals the potential energy $PE = Clp^2 / 2$. Calculation of the viscous and heat conduction bandwidths using these two equations yields the following results for the 1 cm^2 pipe (the values between parentheses are taken from Appendix A for comparison): $B2_{visc}$ 15.34 (15.34), $B3_{visc}$ 19.71 (19.71), $B4_{visc}$ 23.30 (23.30); $B2_{heat}$ 6.81 (6.81), $B3_{heat}$ 8.75 (8.75), $B4_{heat}$ 10.35 (10.35). The results are therefore identical to those found by the FDVT transmission line model, as expected. Stevens (1998, p. 161) observed that both bandwidths B_{visc} and B_{heat} are proportional to the square root of the formant frequency ($B \propto \sqrt{F}$) as well as proportional to S / A or $1 / \sqrt{A}$ when the pipe is a cylinder of circumference S ($B \propto 1 / \sqrt{A}$).

5. Radiation bandwidth

5.1. Radiation bandwidth: kinetic energy radiation bandwith B_{rad}^{KE}

The more important radiation bandwidth B_{rad} can be estimated for a uniform pipe if one assumes that the average dissipated power lost to radiation $P_{rad} / 2 = R_{rad}U_0^2 / 2$ can be approximated as the total power $R_{rad}U^2$ generated over the length *l* of the pipe:

$$B_{rad}^{KE} = \frac{\frac{1}{2}P_{rad}}{2\pi \times \frac{1}{2}TE} = \frac{\frac{1}{2}R_{rad}U_0^2}{2\pi \times \frac{1}{2}LlU^2} \approx \frac{R_{rad}U^2}{2\pi \times \frac{1}{2}LlU^2} = \frac{\frac{\rho c}{A_0}\left[1 - \frac{J_1(2kr)}{kr}\right]}{2\pi \times \frac{1}{2}(\rho / A_{VT})l},$$

where the stored energy in the denominator equals the kinetic energy $KE = LlU^2/2$ and A_{VT} is the area of the vocal tract pipe. This equation for radiation bandwidth produces the following results for the 4 cm^2 pipe (again the values between parentheses are from Appendix A): $B2_{rad}$ 24.49 (24.34), $B3_{rad}$ 66.10 (59.59), $B4_{rad}$ 125.42 (99.88). It is clear that as the formant frequency rises, the analytical radiation bandwidths increasingly exceed those of the FDVT transmission line model. A second approximation of B_{rad}^{KE} is obtained by substituting $(kr)^2/2$ for the dimensionless radiation resistance:

$$B_{rad}^{KE} \approx \frac{\frac{\rho c}{A_0}\left[1 - \frac{J_1(2kr)}{kr}\right]}{2\pi \times \frac{1}{2}(\rho / A_{VT})l} \approx \frac{\frac{\rho c}{A_0}\left[\frac{(kr)^2}{2}\right]}{2\pi \times \frac{1}{2}(\rho / A_{VT})l} = \frac{\frac{\rho c}{A_0}\left[\frac{2\pi F^2 A_0}{c^2}\right]}{2\pi \times \frac{1}{2}(\rho / A_{VT})l},$$

when $kr \ll 1$ (Flanagan, 1972, p. 36). The approximate dimensionless radiation resistance $(kr)^2/2$ in turn equals $2\pi F^2 A_0 / c^2$ where F is the formant frequency (Stevens, 1998, p. 153). The two terminating lip areas A_0 therefore cancel each other in the radiation resistance term R_{rad} of the generated power (cf. Dunn, 1950, p. 743 who gives $R_{rad} = \rho\omega^2 / 2\pi c$, where $\omega = 2\pi F$). Hence the vocal tract area A_{VT} is the only area variable. The results of this second approximation for the 4 cm^2 pipe are: $B2_{rad}$ 24.82, $B3_{rad}$ 68.58, $B4_{rad}$ 134.87, which are fairly comparable to those of the preceding first approximation. When the expression for the kinetic energy radiation bandwidth is further reduced, it becomes

$$B_{rad}^{KE} \approx \frac{2F^2 A_{VT}}{cl}.$$

Fant (1960, p. 307) and Stevens (1998, p. 155) both give a similar formula. Because the radiation bandwidth is proportional to the square of the formant frequency F^2, its contribution to overall bandwidth is relatively small for the F1 frequency (Appendix A). If the formant frequency F is fixed, the kinetic energy radiation quality factor is

$$Q_{rad}^{KE} \approx \frac{F}{(2F^2 A_{VT} / cl)} = \frac{cl}{2FA_{VT}},$$

since $Q = F / B$. The quality factor Q_{rad}^{KE} is then inversely proportional to the vocal tract area A_{VT} with all else remaining the same:

$$Q_{rad}^{KE} \propto \frac{1}{A_{VT}}.$$

Taking the logarithm of both sides of the proportionality, one obtains log Q_{rad}^{KE} $(= 20\log_{10} Q_{rad}^{KE}) \propto 20\log_{10}(1/A_{VT})$. Thus log Q_{rad}^{KE} decreases –6 dB per doubling of the vocal tract area A_{VT}.

5.2. Radiation bandwidth: potential energy radiation bandwith B_{rad}^{PE}

At resonance the kinetic and potential energies are equal: $LlU^2/2 = Clp^2/2$. The general expression for the radiation bandwidth B_{rad} is therefore:

$$B_{rad} \approx \frac{R_{char}U^2\left[\frac{2\pi F^2 A_0}{c^2}\right]}{2\pi \times \frac{1}{2}LlU^2} = \frac{G_{char}p^2\left[\frac{2\pi F^2 A_0}{c^2}\right]}{2\pi \times \frac{1}{2}Clp^2},$$

where the approximate dimensionless radiation resistance is $2\pi F^2 A_0/c^2$ as in the previous section. The stored energy in the denominator of the first quotient is the kinetic energy $KE = LlU^2/2$. The stored energy in the denominator of the second quotient is the potential energy $PE = Clp^2/2$. The characteristic acoustic resistance is $R_{char} = \rho c/A_0$; the characteristic acoustic conductance is $G_{char} = A_0/\rho c$. The products $R_{char}U^2$ and $G_{char}p^2$ are each equal to the power of a progressive plane wave in a pipe of area A_0 (see Kinsler & Frey, 1962, p. 121). Hence the generated power in the numerator of the first quotient is equal to the generated power in the numerator of the second quotient.

The potential energy radiation bandwidth is

$$B_{rad}^{PE} \approx \frac{G_{rad}p^2}{2\pi \times \frac{1}{2}Clp^2} = \frac{G_{char}\left[\frac{2\pi F^2 A_0}{c^2}\right]p^2}{2\pi \times \frac{1}{2}Clp^2} = \frac{\frac{A_0}{\rho c}\left[\frac{2\pi F^2 A_0}{c^2}\right]}{2\pi \times \frac{1}{2}(A_{VT}/\rho c^2)l}.$$

Unlike for B_{rad}^{KE}, the two terminating lip areas A_0 do not cancel each other in the radiation conductance term G_{rad} of the generated power. Note also that the terminating lip area A_0 has an infinitesimal thickness in this expression for B_{rad}^{PE} since A_0 is the vibrating surface of a circular piston lying in a plane (Kinsler & Frey, 1962, p. 166). Simplifying the final term, the potential energy radiation bandwidth becomes

$$B_{rad}^{PE} \approx \frac{2F^2A_0^2}{clA_{VT}}.$$

If the formant frequency *F* is fixed, the potential energy radiation quality factor is

$$Q_{rad}^{PE} \approx \frac{F}{(2F^2A_0^2/clA_{VT})} = \frac{clA_{VT}}{2FA_0^2}.$$

The quality factor Q_{rad}^{PE} is directly proportional to the vocal tract area A_{VT} and inversely proportional to the square of the terminating lip area A_0^2 with all else the same:

$$Q_{rad}^{PE} \propto \frac{A_{VT}}{A_0^2}.$$

It follows that log Q_{rad}^{PE} $(= 20\log_{10} Q_{rad}^{PE}) \propto 20\log_{10}(A_{VT} / A_0^2)$. Hence log Q_{rad}^{PE} increases +6 dB per doubling of the vocal tract area A_{VT} and decreases –12 dB per doubling of the terminating lip area A_0 when the other area variable is held constant.

5.3. Radiation bandwidth: volume velocity and pressure types of quarter wavelength

As was pointed out above, at resonance the general expression for the radiation bandwidth B_{rad} is

$$B_{rad} = \frac{\text{generated power}}{2\pi \times KE} = \frac{\text{generated power}}{2\pi \times PE}.$$

The stored energy in the first quotient is the kinetic energy $KE = LlU^2 / 2$; the stored energy in the second quotient is the potential energy $PE = Clp^2 / 2$. The equation indicates that when the length *l* of the vocal tract pipe is divided by 3, 5, or 7 to obtain the F2, F3, or F4 quarter wavelength, the denominator that includes the stored energy term is reduced by the same factor. The generated power in the numerator is also expected to decline in a similar way since the power in a pipe one quarter wavelength long is equal to 1/3, 1/5, or 1/7 of the total power generated over the length of the vocal tract pipe. Thus the equation for radiation bandwidth B_{rad} is hypothesized to remain approximately valid when the vocal tract area A_{VT} is replaced by the quarter wavelength area $A_{\lambda/4}$. Consequently, the following near equalities are assumed to hold:

$$Q_{rad}^{KE} \propto \frac{1}{A_{VT}} \approx \frac{1}{A_{\lambda/4}};\ Q_{rad}^{PE} \propto \frac{A_{VT}}{A_0^2} \approx \frac{A_{\lambda/4}}{A_0^2}.$$

A satisfactory account of radiation bandwidth will require two types of quarter wavelength to be distinguished (i) the volume velocity type of quarter wavelength with a volume velocity maximum at the lip-oriented edge; (ii) the pressure type of quarter wavelength with a pressure maximum at the lip-oriented edge. As will be shown below, there is a notable difference in radiation bandwidth behavior between the volume velocity and pressure types of quarter wavelength (see Section 3.1.3 for the equations $KE = L\Delta lU^2 / 2$ and $PE = C\Delta lp^2 / 2$):

1. The volume velocity type of quarter wavelength has a volume velocity maximum U_{max} at the lip-oriented edge. The corresponding kinetic energy $KE = L\Delta lU_{max}^2 / 2$ also has its maximum value at the lip-oriented edge. The volume velocity type of quarter wavelength is associated with only the kinetic energy radiation bandwidth B_{rad}^{KE}. Since $Q_{rad}^{KE} \propto 1 / A_{A_{\lambda/4}}$, the quality factor Q_{rad}^{KE} decreases –6 dB per doubling of the quarter wavelength area $A_{\lambda/4}$.

2. The pressure type of quarter wavelength has a pressure maximum p_{max} at the lip-oriented edge. The corresponding potential energy $PE = C\Delta l p_{max}^2 / 2$ also has its maximum value at the lip-oriented edge. The pressure type of quarter wavelength is associated with only the potential energy radiation bandwidth B_{rad}^{PE}. Since $Q_{rad}^{PE} \propto A_{\lambda/4} / A_0^2$, the quality factor $\mathrm{Q}_{\mathrm{rad}}^{\mathrm{PE}}$ increases +6 dB per doubling of the quarter wavelength area $A_{\lambda/4}$ and decreases –12 dB per doubling of the terminating lip area A_0.

In order to verify numerically rules 1 and 2 for the volume velocity and pressure types, two-pipe FDVT models are implemented which calculate log Q using radiation loss only. Accordingly, losses due to viscosity, heat conduction, and wall damping are omitted for the moment.

5.4. Radiation bandwidth: two-pipe models

5.4.1. The reference vocal tract

Before proceeding to the two-pipe models, where the area of one articulator region is variable while the three other regions are constant, it is of use to determine the length and cross-sectional area of the reference vocal tract. As described below in Section 6, the vowel area functions from the 10-speaker data set in Table 1 and the 12-speaker data set in Tiede (2013) are fit to 27 equal-length tubes using cubic spline interpolation. The interpolated vowel area functions from both data sets are pooled into adult male and female groups, with a total of 115 vowels in the former (N = 17) and 29 in the latter (N = 4). The average vocal tract length of the men's group is 17.58 cm (s.d. 1.14), that of the women's group 14.33 cm (s.d. 1.22). By comparison, Goldstein (1980, p. 186) found slightly smaller vocal tract lengths of 16.93 cm and 14.09 cm for men and women 21 years old. The average vocal tract area of the men's group is 2.72 cm^2 (s.d. 1.17) while that of the women's group is 1.95 cm^2 (s.d. 0.73). For present purposes, the reference vocal tract may then be conveniently defined as a uniform pipe 17 cm in length and 2 cm^2 in cross-sectional area. Remark in addition that the ratio between the women's and men's mean vocal tract lengths is 14.33/17.58 or 0.815, which comes close to $2^{-1/4}$ or 0.841. It is clear from Section 3.1.1 that formant frequencies are inversely proportional to vocal tract length $F \propto l^{-1}$. Therefore if the vocal tract length l is shortened by the factor $2^{-1/4}$, formant frequencies are raised 1/4 octave given that $F \propto (2^{-1/4} \times l)^{-1} = 2^{1/4} \times l^{-1}$. The man-to-woman scale factor of 1/4 octave for formant frequencies (18.9%) has been reported in several studies (see Pennington, 2017, Section 3.2).

5.4.2. Variable area lip pipe—constant area non-lip pipe

The FDVT model computes the log Q values of a variable area 4-tube lip pipe (tubes 1–4) and a constant area 23-tube non-lip pipe (tubes 5–27) that includes the three

other articulator regions (6-tube blade, 9-tube tongue body, 8-tube tongue root). The five areas of the variable lip pipe chosen for evaluation are 0.5, 1, 2, 4, 8 cm^2. The area of the constant non-lip pipe is set at 2 cm^2, the cross-sectional area of the reference vocal tract. The combined length of the lip and non-lip pipes is 17 cm, the length of the reference vocal tract. Linear regression is applied with log Q values being the dependent variable and base-2 logarithms of the five lip areas (0.5, 1, 2, 4, 8 cm^2) being the independent variable. The base-2 logarithm is taken so that each doubling of the area results in a unit increment (–1, 0, 1, 2, 3).

At the F4 resonance mode the 4-tube lip region (tubes 1–4) nearly spans a quarter wavelength one-seventh (4/28) the length *l* of the 27-tube vocal tract. This F4 quarter wavelength is a volume velocity type of quarter wavelength with a volume velocity maximum at the lip-oriented edge (see Section 3.1.1 and Fant, 1960, p. 85). In Section 5.3, rule 1 for the volume velocity type states that $Q_{rad}^{KE} \propto 1/A_{A_{\lambda/4}}$. As a result, the quality factor $Q4_{rad}^{KE}$ should decrease –6 dB per doubling of the lip area $A_{\lambda/4}$. The calculated log Q4 regression slope is –7.48 dB per doubling of the 4-tube lip area. Hence the rule for the volume velocity type describes the log Q4 regression behavior reasonably well. Observe also that the power-energy equation for radiation bandwidth B_{rad} in Section 5.3 appears to hold when the vocal tract area A_{VT} is replaced by the quarter wavelength lip area $A_{\lambda/4}$. Thus the near equality $Q_{rad}^{KE} \propto 1/A_{VT} \approx 1/A_{\lambda/4}$ is verified.

In the FDVT model (Fig. 1), the 6-tube blade region spans the F3 quarter wavelength that has a pressure maximum (volume velocity minimum) at four-fifths the vocal tract length *l* from the glottis and a pressure minimum (volume velocity maximum) at three-fifths *l* (tubes 5–10); the 9-tube tongue body region spans the F2 quarter wavelength that has a pressure maximum (volume velocity minimum) at two-thirds the vocal tract length *l* from the glottis and a pressure minimum (volume velocity maximum) at one-third *l* (tubes 11–19). Thus the F3 and F2 quarter wavelengths are a pressure type of quarter wavelength with a pressure maximum at the lip-oriented edge. Rule 2 for the pressure type in Section 5.3 states that $Q_{rad}^{PE} \propto A_{\lambda/4}/A_0^2$, in which Q_{rad}^{PE} increases +6 dB per doubling of the quarter wavelength area $A_{\lambda/4}$ and decreases –12 dB per doubling of the terminating lip area A_0. The log Q3 and log Q2 regression slopes are respectively –10.13 dB and –6.90 dB per doubling of the 4-tube lip area. The log Q3 regression slope is fairly similar to that given for the terminating lip area A_0 (–10.13 dB vs. –12 dB per doubling of lip area), thereby establishing the relevance of rule 2 for the pressure type of quarter wavelength. Although the terminating lip area A_0 is the surface of an idealized circular piston without thickness, the proportionality $Q_{rad}^{PE} \propto A_{\lambda/4}/A_0^2$ remains mostly valid when A_0 is replaced by the area of the 4-tube lip region. In the ratio $A_{\lambda/4}/A_0^2$, the square of the terminating lip area A_0 grows substantially faster than the non-squared

quarter wavelength area $A_{\lambda/4}$. As a consequence, it is essential to keep the area of the lip region fixed when the area of the blade or tongue body region is varied.

5.4.3. Constant area lip pipe—variable area non-lip pipe

The FDVT model calculates the log Q values of three variable area pipes: (i) a 9-tube tongue body pipe (tubes 11–19), (ii) a 6-tube blade pipe (tubes 5–10), and (iii) a 6-tube blade-shifted pipe (tubes 11–16) in which the blade region is shifted back one F3 quarter wavelength toward the glottis. The default area of the other tubes is 2 cm^2. Therefore the lip pipe area is always fixed at 2 cm^2. The total length of the model is 17 cm. Linear regression is performed with log Q values being the dependent variable and the base-2 logarithms of the five variable areas (0.5, 1, 2, 4, 8 cm^2) being the independent variable.

As was covered in the previous section, the 9-tube tongue body region spans the F2 quarter wavelength with a pressure maximum at two-thirds the vocal tract length *l* from the glottis and a pressure minimum at one-third *l* (tubes 11–19); the 6-tube blade region spans the F3 quarter wavelength with a pressure maximum at four-fifths the vocal tract length *l* from the glottis and a pressure minimum at three-fifths *l* (tubes 5–10). Thus the F2 and F3 quarter wavelengths are a pressure type of quarter wavelength with a pressure maximum at the lip-oriented edge. Hence rule 2 for the pressure type $Q_{rad}^{PE} \propto A_{\lambda/4} / A_0^2$ applies where Q_{rad}^{PE} increases +6 dB per doubling of the quarter wavelength area $A_{\lambda/4}$ and decreases –12 dB per doubling of the terminating lip area A_0. The log Q2 regression slope is +4.13 dB per doubling of tongue body area whereas the log Q3 regression slope is +7.82 dB per doubling of blade area. The average of the log Q2 and log Q3 regression slopes for the tongue body and blade pipes (= 5.98 dB) approaches the predicted Q_{rad}^{PE} increase of +6 dB per doubling of quarter wavelength area $A_{\lambda/4}$ when the lip pipe area is fixed.

For each tube *n* of the 27-tube FDVT model, decibel (dB) measures of volume velocity (= $10\log_{10} U^2(n)$) and pressure (= $10\log_{10} p^2(n)$) are plotted in Fig. 2 and Fig. 3 employing radiation loss only. In Fig. 2 the volume velocity and pressure data for the F2 resonance mode are shown for the five tongue body areas; in Fig. 3 these data for the F3 resonance mode are given for the five blade areas. As always in this paper, the volume velocity source at the glottis equals 1 for each frequency. Because the driving volume velocity input is 1 at the F2 and F3 resonance frequencies, the dB measure of volume velocity and pressure in tube *n* represents a transfer function or the ratio between output and input. To provide some perspective on the transfer function, one notes that in an RLC circuit at resonance frequency f_0, the quality or amplification factor Q is equal to the transfer function V_C / E, the ratio between output capacitor voltage V_C and input voltage E (Section 4). The volume velocity or pressure measures assessed over the entire set of 27 tubes represent a spatial transfer function at the respective formant frequency. There

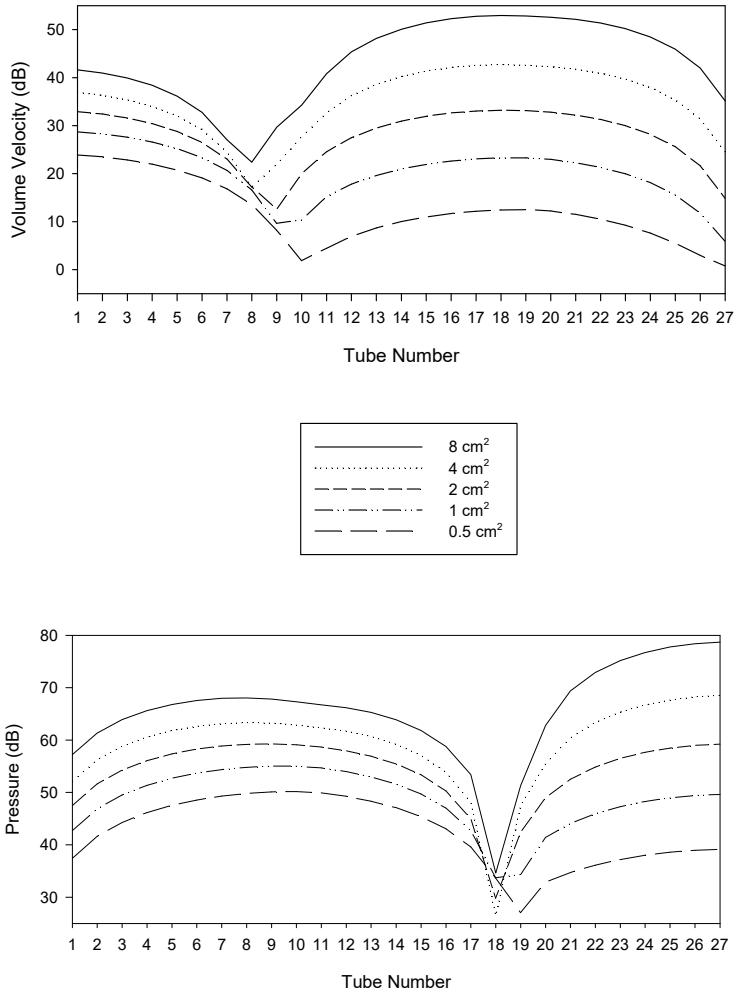

Fig. 2. The F2 spatial transfer functions of volume velocity and pressure for five tongue body areas using radiation loss only. The cross-sectional areas of the 9-tube tongue body pipe (tubes 11–19) vary from 0.5 to 8 cm^2 while the rest of the 17 cm pipe is constant at 2 cm^2.

exists a close similarity between the F2 volume velocity envelope in Fig. 1 and the five F2 volume velocity curves with different tongue body areas in Fig. 2. The comparison reveals that the general shape of the F2 spatial transfer function is constrained by the boundary conditions of the open-closed uniform pipe. A strong resemblance also exists between the F3 volume velocity envelope in Fig. 1 and the five F3 volume velocity curves in Fig. 3. The expected reciprocal relationship between volume velocity minima and pressure maxima is seen in Fig. 2 and Fig. 3 as well. The F2 and F3 spatial transfer functions of the vowels in Mrayati and Carré (1976) are likewise constrained by the open-closed boundary conditions.

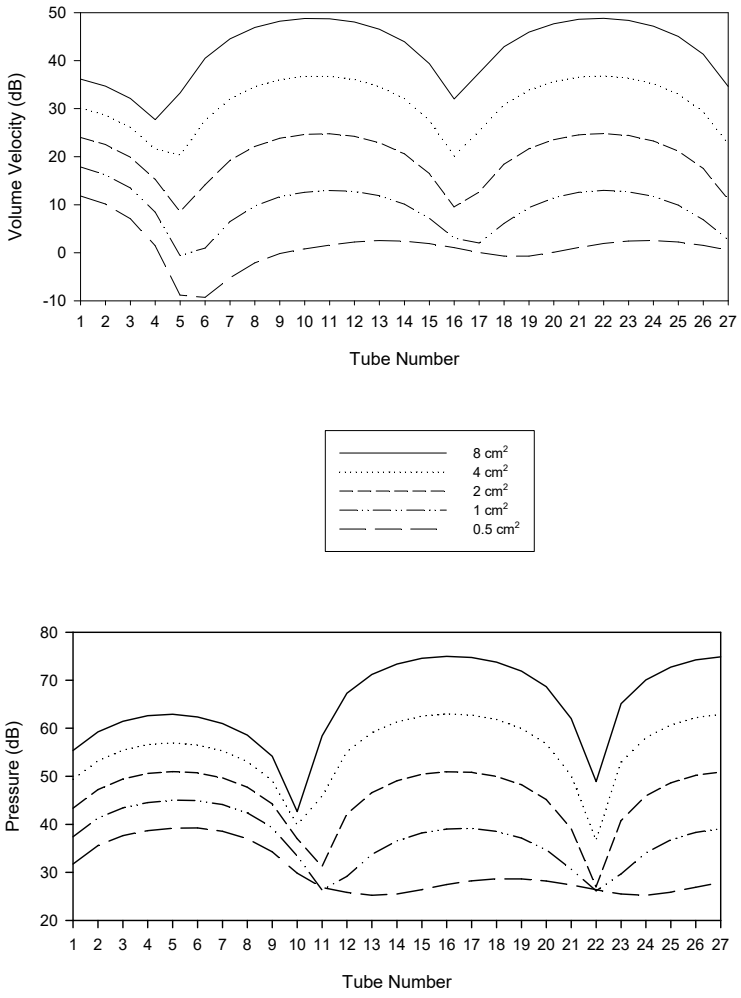

Fig. 3. The F3 spatial transfer functions of volume velocity and pressure for five blade areas using radiation loss only. The cross-sectional areas of the 6-tube blade pipe (tubes 5–10) vary from 0.5 to 8 cm^2 while the rest of the 17 cm pipe is constant at 2 cm^2.

It is clear from the above presentation that the general shape of the F2 and F3 spatial transfer functions is constrained by the boundary conditions of the open-closed pipe. In Fig. 2 it is equally apparent that there is a regular rise of F2 volume velocity and pressure levels along the 27-tube vocal tract as the tongue body area increases. In Fig 3 there is a similar rise of F3 volume velocity and pressure levels along the vocal tract as the blade area increases. Thus the F2 and F3 spatial transfer functions are not only restricted in shape by the boundary conditions, but their dB levels display a collective shift upward over the whole vocal tract as the tongue body or blade area becomes larger. When a change in quarter wavelength area $A_{\lambda/4}$, and hence Q_{rad}^{PE}, brings about dB level shifts of the spatial transfer functions in Fig. 2 and Fig. 3, the dB shifts visibly propagate beyond the edges of the selected F2 or F3 quarter wavelength.

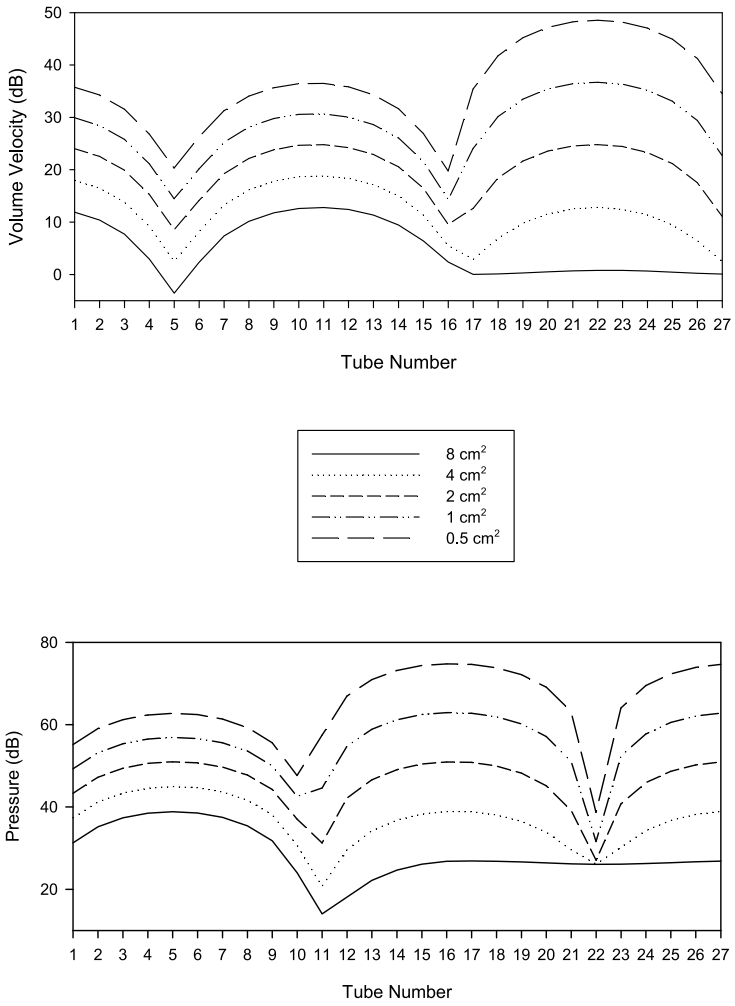

Fig. 4. The F3 spatial transfer functions of volume velocity and pressure for five blade-shifted areas using radiation loss only. The cross-sectional areas of the 6-tube blade-shifted pipe (tubes 11–16) vary from 0.5 to 8 cm^2 while the rest of the 17 cm pipe is constant at 2 cm^2.

If the 6-tube blade region is shifted back one quarter wavelength toward the glottis, the resulting quarter wavelength has a volume velocity maximum at the lip-oriented edge. The blade-shifted region extends from the F3 volume velocity maximum at three-fifths the vocal tract length *l* from the glottis to the F3 volume velocity minimum at two-fifths *l*. The blade-shifted region therefore comprises tubes No. 11–16 instead of tubes No. 5–10 that make up the original blade region. In Section 5.3, rule 1 for the volume velocity type states that $Q_{rad}^{KE} \propto 1/A_{A_{\lambda/4}}$. Hence the quality factor $\mathrm{Q3}_{\mathrm{rad}}^{\mathrm{KE}}$ should decrease –6 dB per doubling of the quarter wavelength area $A_{\lambda/4}$. The obtained log Q3 regression slope is –5.99 dB per doubling of blade-shifted area, an exact match. The F3 volume velocity and pressure levels along the vocal tract are displayed for the five blade-shifted areas in Fig. 4. When the blade-shifted area grows larger, the dB curves fall rather

than rise as they do for the blade areas in Fig. 3. For the blade-shifted region, the F3 spatial transfer functions exhibit a directly opposite pattern compared to the original blade region—much like the log Q3 regression slopes (–5.99 dB vs. +7.82 dB per doubling of $A_{\lambda/4}$). Hence at the same resonance mode, there is a stark difference in radiation bandwidth behavior between a quarter wavelength area $A_{\lambda/4}$ of the volume velocity type and a quarter wavelength area $A_{\lambda/4}$ of the pressure type. This is to be expected from the proportionalities $Q_{rad}^{KE} \propto 1/A_{A_{\lambda/4}}$ and $Q_{rad}^{PE} \propto A_{\lambda/4}/A_0^2$, where $A_{\lambda/4}$ is in the denominator of the first proportionality and in the numerator of the second. Remark finally that the power-energy equation for radiation bandwidth B_{rad} in Section 5.3 remains approximately valid when the vocal tract area A_{VT} is replaced by the quarter wavelength area $A_{\lambda/4}$ in each of the three tested cases: (i) the tongue body region, (ii) the blade region, (iii) the blade-shifted region. Hence these near equalities are confirmed:

$$Q_{rad}^{KE} \propto \frac{1}{A_{VT}} \approx \frac{1}{A_{\lambda/4}};\ Q_{rad}^{PE} \propto \frac{A_{VT}}{A_0^2} \approx \frac{A_{\lambda/4}}{A_0^2}.$$

5.4.4. Summary of results for the two-pipe models

Provided next is a summary of the findings developed so far for the radiation bandwidth behavior of the two-pipe FDVT models, in which the area of one articulator region ranges over 0.5, 1, 2, 4, 8 cm^2 while the area of the remaining regions defaults to 2 cm^2:

1a. The log Q2 regression slope is +4.13 dB per doubling of tongue body area.

2a. The log Q3 regression slope is +7.82 dB per doubling of blade area.

3a. The log Q4 regression slope is –7.48 dB per doubling of lip area.

The log Q values of the three regression slopes were calculated using radiation loss alone. When losses due to viscosity, heat conduction, and wall damping are additionally brought into the computation, the results are the following:

1b. The log Q2 regression slope is +2.10 dB per doubling of tongue body area.

2b. The log Q3 regression slope is +4.40 dB per doubling of blade area.

3b. The log Q4 regression slope is –4.28 dB per doubling of lip area.

For the pressure type of quarter wavelength, the full-loss log Q2 and log Q3 regression slopes are roughly ~ +3 dB per doubling of the tongue body and blade areas rather than +6 dB per doubling of $A_{\lambda/4}$ with radiation loss only ($Q_{rad}^{PE} \propto A_{\lambda/4}/A_0^2$). For the volume velocity type of quarter wavelength, the full-loss log Q4 regression slope is roughly ~ –3 dB per doubling of the lip area rather than –6 dB per doubling of $A_{\lambda/4}$ with radiation loss only ($Q_{rad}^{KE} \propto 1/A_{A_{\lambda/4}}$). When the losses due to viscosity, heat conduction, and wall damping are added, the quality factor Q becomes a distinctly less sensitive cue for articulator region area. Nevertheless, the rules of radiation bandwidth behavior, as

elaborated theoretically in Section 5.3, continue to dominate the overall pattern of the results.

Part III. Vowel acoustic-articulatory correlates

6. Vowel area functions

Table 1 above classifies the vowel area functions of the original 10-speaker data set analyzed in Pennington (2011). The vowel area functions are drawn from lists in two X-ray and five MRI studies. There are two adult female speakers (Japanese, English), seven adult male speakers (Russian, French, English, Japanese), and one 11-year-old male speaker (Japanese). In his MRI investigation Tiede (2013) assembled a 12-speaker data set with six American English and six Japanese subjects. There are five men and one woman in each language group. Area functions of the Japanese vowels [a e i o ɯ] and the English vowels [a æ ɛ e ɪ i o ʊ u] are tabulated for all speakers of the respective language. The cross-sectional area is estimated using two different methods, the first through pixel counting, the second through polygonal triangulation. As the area measures provided by the two methods are generally in close agreement (Tiede, p. 47), their mean is taken here to be the representative value.

The area functions of (i) the original 10-speaker data set and (ii) the Tiede 12-speaker data set are 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.

7. Articulatory parameters

Using the interpolated area functions, the FVDT program calculates 23 articulatory parameters. They are described in the following statements (see Fig. 1 for the tube numbers):

1) The lip length $L(lip)$ is the sum of tube lengths in the lip region from tube 1 to 4. The lip length $L(lip)$ is also converted to its base-2 logarithm: $\log_2 L(lip)$. Hence a length doubling leads to a unit increase in $\log_2 L(lip)$.

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 constriction, $I_{\min A}(body)$ the tongue body constriction, and $I_{\min A}(root)$ the tongue root constriction. The blade range is $6 \geq I_{\min A}(blade) \geq 1$ (from tube 5 to 10), the tongue body range is $9 \geq I_{\min A}(body) \geq 1$ (from tube 11 to 19), and the tongue root range is $8 \geq I_{\min A}(root) \geq 1$ (from tube 20 to 27).

3) The mean area $\overline{A}$ is the length-weighted mean of the tube areas in a given region. The summation indices *k* and *j* are tube numbers; the area and length of tube *i* are A_i and l_i:

$$\overline{A} = \frac{\sum_{i=k}^{j} l_i A_i}{\sum_{i=k}^{j} l_i}.$$

4) The summation indices are the following: the mean lip area $\overline{A}(lip)$ k = 1, j = 4; the mean blade area $\overline{A}(blade)$ k = 5, j = 10; the mean tongue body area $\overline{A}(body)$ k = 11, j = 19; the mean tongue root area $\overline{A}(root)$ k = 20, j = 27. When the tube length l_i remains constant, the mean area $\overline{A}$ is simply the sum of the tube areas in the region divided by the number of tubes. The mean area $\overline{A}$ is furthermore converted to its base-2 logarithm: $\log_2 \overline{A}(lip)$, $\log_2 \overline{A}(blade)$, $\log_2 \overline{A}(body)$, $\log_2 \overline{A}(root)$. Hence an area doubling leads to a unit increase in $\log_2 \overline{A}$.

5) The mean areas of the blade, tongue body, and tongue root regions are divided by the mean lip area yielding the blade aperture $\overline{A}(blade)/\overline{A}(lip)$, the tongue body aperture $\overline{A}(body)/\overline{A}(lip)$, and the tongue root aperture $\overline{A}(root)/\overline{A}(lip)$. The area ratios are transformed logarithmically: $\log_2 \overline{A}(blade)/\overline{A}(lip)$, $\log_2 \overline{A}(body)/\overline{A}(lip)$, $\log_2 \overline{A}(root)/\overline{A}(lip)$.

6) The mean areas of the 2-tube foreblade, the 2-tube hindblade, the 3-tube forebody, and the 3-tube hindbody are calculated. The summation indices are the following: the mean foreblade area $\overline{A}(foreblade)$ k = 5, j = 6; the mean hindblade area $\overline{A}(hindblade)$ k = 9, j = 10; the mean forebody area $\overline{A}(forebody)$ k = 11, j = 13; the mean hindbody area $\overline{A}(hindbody)$ k = 17, j = 19. The mean areas of the foreblade and forebody are respectively divided by the mean hindblade and hindbody areas, giving the blade and tongue body

positions: $\overline{A}(\mathit{foreblade}) / \overline{A}(\mathit{hindblade})$, $\overline{A}(\mathit{forebody}) / \overline{A}(\mathit{hindbody})$. The log ratios are $\log_2 \overline{A}(\mathit{foreblade}) / \overline{A}(\mathit{hindblade})$, $\log_2 \overline{A}(\mathit{forebody}) / \overline{A}(\mathit{hindbody})$.

8. Acoustic parameters

There are 9 acoustic parameters. The first four formant frequencies (F1–F4), the corresponding 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 octave scales of pitch and formant perception (Miller, 1989). As was mentioned in Section 2, the logarithmic quality factor log Q ($= 20 \log_{10} Q$) is a measure of the relative power or power gain at resonance. The reciprocal of the quality factor Q is equal to the formant bandwidth B normalized by the formant frequency F: $Q^{-1} = B / F$. Thus one might expect formant bandwidth itself to be a relevant acoustic parameter. However in Pennington (2011), it was discovered that the bandwidths B1–B4 gave unsatisfactory results relative to the quality factors log Q1–log Q4. In comparison with log Q, therefore, formant bandwidth B is not a useful acoustic parameter.

9. Acoustic-articulatory correlations

9.1. Acoustic-articulatory correlations: method

To estimate the strength of association between the acoustic and articulatory parameters, Pearson correlation matrices are calculated for the vowel system of each speaker. Then the correlation coefficients of the parameter pairs are averaged across the 10 speakers in the original data set and the 12 speakers in the Tiede data set. The means and standard deviations of the acoustic-articulatory correlations are presented in Appendix B. The top five absolute values of the mean correlation coefficient r are ranked from the largest (1) to the smallest (5) for every acoustic parameter in the two data sets. In light of Ratner's guidelines for big data sets (2017, p. 26), $|r| < 0.3$ will be considered a weak correlation, $0.3 \le |r| \le 0.7$ a moderate correlation, and $|r| > 0.7$ a strong correlation.

9.2. Acoustic-articulatory correlations: the F1, F2, and F3 formant frequencies

According to the perturbation analysis developed in Section 3.1.3, the general proportionality for formant frequency F_i can be stated as

$$F_i \propto \frac{1}{(\lambda_i / 4)} \left(\frac{A_{U\,max}}{A_{p\,max}} \right)^{a/b},$$

where $A_{U\,max}$ is the area at the volume velocity maximum and $A_{p\,max}$ the area at the pressure maximum. The fractional exponent a/b is the simplification or the experimental value of n/N, where n is the number of tubes spanning the 1/3 quarter wavelengths of formant frequency F_i and N equals 27, the total number of tubes. In the ratio $A_{U\,max}/A_{p\,max}$, the area $A_{U\,max}$ at the volume velocity maximum is in the numerator while the area $A_{p\,max}$ at the pressure maximum is in the denominator. As will be seen in the next section, the tongue root area $\overline{A}(root)$ $(=A_{p\,max})$ is the dominant term in the tongue root aperture $\overline{A}(root)/\overline{A}(lip)$ because the 8-tube tongue region is twice as long as the 4-tube lip region. Therefore the area $\overline{A}(root)$ at the pressure maximum is placed in the numerator of $\overline{A}(root)/\overline{A}(lip)$ whereas the lip area $\overline{A}(lip)$ becomes a normalizing term. To be consistent, the areas $\overline{A}(forebody)$ and $\overline{A}(foreblade)$ are likewise placed in the numerators of $\overline{A}(forebody)/\overline{A}(hindbody)$ and $\overline{A}(foreblade)/\overline{A}(hindblade)$ given that both are at the pressure maximum. Note that there is little consequence to this inversion of numerator and denominator since the absolute value of the correlation coefficient is the same in either case.

9.2.1. The F1 frequency

For the F1 resonance mode, $A_{p\,max}$ is the area of the pressure maximum at the glottis; $A_{U\,max}$ is the area of the volume velocity maximum at the lips. Thus in view of the general proportionality $F_i \propto (\lambda_i/4)^{-1}(A_{U\,max}/A_{p\,max})^{a/b}$, the F1 frequency should be raised when the mean tongue root area $\overline{A}(root)$ $(=A_{p\,max})$ becomes smaller, and vice versa. In a parallel manner, the F1 frequency should be lowered when the mean lip area $\overline{A}(lip)$ $(=A_{U\,max})$ becomes smaller, and vice versa. In Appendix B, the first- and second-ranked correlations for log F1 $(=\log_2 F1)$ are: (1^{st}) $\log_2 \overline{A}(root)/\overline{A}(lip)$ r = –0.9352 and (2^{nd}) $\log_2 \overline{A}(root)$ r = –0.9151 in the original data set, (1^{st}) $\overline{A}(root)$ r = –0.8267 and (2^{nd}) $\log_2 \overline{A}(root)/\overline{A}(lip)$ r = –0.8233 in the Tiede data set. The strong to very strong inverse correlations between the F1 frequency and the tongue root aperture $\overline{A}(root)/\overline{A}(lip)$ provide excellent evidence for a perturbation analysis governed by the general proportionality $F_i \propto (\lambda_i/4)^{-1}(A_{U\,max}/A_{p\,max})^{a/b}$. The 4-tube lip region (No. 1–4) is half the length of the 8-tube tongue root region (No. 20–27). As a result, the influence of the mean lip area $\overline{A}(lip)$ is comparatively diminished. This would explain why the correlations between F1 and the mean root area $\overline{A}(root)$ are very similar to those between F1 and the tongue root aperture $\overline{A}(root)/\overline{A}(lip)$. Consequently, $\overline{A}(root)$ is the dominant term in the ratio $\overline{A}(root)/\overline{A}(lip)$.

To determine how much log F1 decreases as the tongue root aperture $\overline{A}(root)/\overline{A}(lip)$ increases, log F1 regression slopes are calculated. The original data set shows a mean slope of –0.342 octave (s.d. 0.05) per doubling of $\overline{A}(root)/\overline{A}(lip)$. The Tiede data set has a mean slope of –0.306 octave (s.d. 0.08) per doubling of $\overline{A}(root)/\overline{A}(lip)$. The average slope for the two data sets is –0.324 octave per doubling of $\overline{A}(root)/\overline{A}(lip)$. In Section 3.1.3 the analytical F1 proportionality was found to be $F_1 \propto (\lambda_1/4)^{-1}(A_{U\,max}/A_{p\,max})^{1/3}$, where the theoretical fractional exponent a/b is 1/3. Hence when the 9-tube hindtract area $A_{p\,max}$ is doubled, the F1 frequency decreases by –1/3 (–0.333) octave ($=\log_2(2^{-1/3})$). Since the average regression slope is –0.324 octave per doubling of $\overline{A}(root)/\overline{A}(lip)$, the experimental fractional exponent a/b is about ~ 1/3 (0.333). Therefore the empirical F1 proportionality is $F_1 \propto (\lambda_1/4)^{-1}(A_{U\,max}/A_{p\,max})^{1/3}$.

9.2.2. The F2 frequency

With regard to the perturbation analysis, the two articulatory parameters most likely to be correlated with the F2 frequency are the following:

(a) The tongue body position $\overline{A}(forebody)/\overline{A}(hindbody)$ in which the mean forebody area is divided by the mean hindbody area.

(b) The tongue body constriction $I_{\min A}(body)$ where $I_{\min A}$ designates the integer index of the tube with the smallest constriction in the articulator region. The tongue body range is $9 \geq I_{\min A}(body) \geq 1$, where index 9 (tube 11) is under the hard palate and index 1 (tube 19) near the uvula.

For the F2 resonance mode, $A_{p\,max}$ is the area of the pressure maximum at the 3-tube forebody (No. 11–13); $A_{U\,max}$ is the area of the volume velocity maximum at the 3-tube hindbody (No. 17–19). Because the general proportionality for the formant frequency is $F_i \propto (\lambda_i/4)^{-1}(A_{U\,max}/A_{p\,max})^{a/b}$, the F2 frequency should be raised when the mean forebody area $\overline{A}(forebody)$ $(=A_{p\,max})$ becomes smaller, and vice versa. The F2 frequency should be lowered when the mean hindbody area $\overline{A}(hindbody)$ $(=A_{U\,max})$ becomes smaller, and vice versa.

When the tongue body constriction moves from the volume velocity maximum at the uvula index 1 (reduced $A_{U\,max}$) to the pressure maximum at the hard palate index 9 (reduced $A_{p\,max}$), the F2 frequency is also expected to rise from its lowest to its highest value. The tongue body constriction implements the traditional view that "in front vowels the 'front' of the tongue is raised in the direction of the hard palate, while in back vowels the 'back' of the tongue is raised in the direction of the soft palate (Jones, 1922, p. 17)." A contemporary account of front and back vowels in the Handbook of the International Phonetic Association (1999, pp. 10–11) gives nearly the same description (see Catford,

1981 for a historical synopsis of the front-back distinction and its acoustic effect, particularly on the F2 frequency).

In the original data set there are moderate inverse correlations between log F2 ($=\log_2 F2$) and the tongue body positions: $\overline{A}(forebody)/\overline{A}(hindbody)$ r = –0.6114 (s.d. 0.163); $\log_2 \overline{A}(forebody)/\overline{A}(hindbody)$ r = –0.6138 (s.d. 0.228). In the Tiede data set there are low-moderate inverse correlations between log F2 and tongue body positions: $\overline{A}(forebody)/\overline{A}(hindbody)$ r = –0.3612 (s.d. 0.255); $\log_2 \overline{A}(forebody)/\overline{A}(hindbody)$ r = –0.3627 (s.d. 0.327). In the original data set there is a moderate direct correlation between log F2 and the tongue body constriction: $I_{\min A}(body)$ r = 0.6171 (s.d. 0.280); in the Tiede data set there is also a moderate direct correlation between log F2 and the tongue body constriction: $I_{\min A}(body)$ r = 0.6208 (s.d. 0.237). These correlation results suggest that the tongue body position $\overline{A}(forebody)/\overline{A}(hindbody)$ and the tongue body constriction $I_{\min A}(body)$ can serve almost equally well as the articulatory parameter for F2 frequency (but see the discussion at the end of this section).

Inspection of the F2 formant variables in Appendix B reveals a strong inverse correlation between the F2 frequency and the mean blade area $\overline{A}(blade)$ in the original and Tiede data sets. Seeing that the F2 frequency is inversely correlated with the tongue body position $\overline{A}(forebody)/\overline{A}(hindbody)$, the mean blade area $\overline{A}(blade)$ manifestly behaves in the same way as the mean forebody area $\overline{A}(forebody)$ in the numerator of the tongue body position. In Fig. 1, there is a broad F2 pressure maximum (volume velocity minimum) from tube 7 to tube 10 which extends well into the blade region (tubes 5–10). Hence according to the perturbation analysis, a change in the blade area $\overline{A}(blade)$ should have a significant effect on the F2 frequency. In addition, the inverse correlation between F2 frequency and blade area does not appear to compromise the distinctive function of the blade area $\overline{A}(blade)$ in the numerator of the blade aperture $\overline{A}(blade)/\overline{A}(lip)$, the articulatory correlate of the Q3 quality factor (see Section 9.4.1 below).

To estimate how much log F2 decreases as the tongue body position $\overline{A}(forebody)/\overline{A}(hindbody)$ increases, log F2 regression slopes are calculated. The original data set has a mean slope of –0.177 octave (s.d. 0.09) per doubling of $\overline{A}(forebody)/\overline{A}(hindbody)$. The Tiede data set has a mean slope of –0.134 octave (s.d. 0.17) per doubling of $\overline{A}(forebody)/\overline{A}(hindbody)$. The average slope for the two data sets is then –0.156 octave per doubling of $\overline{A}(forebody)/\overline{A}(hindbody)$. In Section 3.1.3 the analytical F2 proportionality is $F_2 \propto (\lambda_2/4)^{-1}(A_{U\,max}/A_{p\,max})^{1/9}$, where the theoretical fractional exponent a/b is 1/9. Hence when the forebody area $A_{p\,max}$ is doubled or the hindbody area $A_{U\,max}$ is halved, the F2 frequency decreases by –1/9 (–0.111) octave ($=\log_2(2^{-1/9})$). As the average regression slope is –0.156 octave per doubling of

$\overline{A}(forebody)/\overline{A}(hindbody)$, the experimental fractional exponent a/b is approximately ~ 1/6 (0.167). Thus the empirical F2 proportionality is $F_2 \propto (\lambda_2/4)^{-1}(A_{U\,max}/A_{p\,max})^{1/6}$. In the preceding paragraph, (i) a broad F2 pressure maximum was shown to cover much of the blade region; (ii) an inverse correlation was found between the F2 frequency and the blade area $\overline{A}(blade)$. Accordingly, there are blade tubes which behave like forebody tubes. As a result, the forebody length at the F2 pressure maximum is effectively longer than the 3 tubes given by the theoretical fractional exponent 3/27 (1/9) for the 1/3 quarter wavelengths. This would explain the larger experimental fractional exponent of 1/6.

It was mentioned earlier that the vowel correlation results do not indicate if the tongue body position $\overline{A}(forebody)/\overline{A}(hindbody)$ or the tongue body constriction $I_{\min A}(body)$ is the better articulatory parameter for the F2 frequency. The tongue body constriction $I_{\min A}(body)$ specifies whether there is a constriction at the pressure maximum (reduced $A_{p\,max}$) for front vowels or a constriction at the volume velocity maximum (reduced $A_{U\,max}$) for back vowels. On the other hand, the tongue body position $\overline{A}(forebody)/\overline{A}(hindbody)$ assesses simultaneously both the forebody area ($=A_{p\,max}$) and the hindbody area ($=A_{U\,max}$). Because the tongue body constriction $I_{\min A}(body)$ detects only the smallest constriction along the tongue body region, the tongue body position $\overline{A}(forebody)/\overline{A}(hindbody)$ is the more sensitive articulatory parameter for the F2 frequency. Nevertheless, tongue body position does not exclude tongue body constriction. Given that $F_2 \propto (\lambda_2/4)^{-1}(A_{U\,max}/A_{p\,max})^{1/6}$, a front vowel constriction at the pressure maximum (reduced $A_{p\,max}$) or a back vowel constriction at the volume velocity maximum (reduced $A_{U\,max}$) may each be complemented by a dilation at the opposite end of the tongue body region.

9.2.3. The F3 frequency

The two articulatory parameters adopted for correlation with the F3 frequency are the following:

1. The blade position $\overline{A}(foreblade)/\overline{A}(hindblade)$ in which the mean foreblade area is divided by the mean hindblade area.
2. The blade constriction $I_{\min A}(blade)$ where $I_{\min A}$ designates the integer index of the tube with the smallest constriction in the articulator region. The tongue blade range is $6 \geq I_{\min A}(blade) \geq 1$, where index 6 (tube 5) is near the front teeth and index 1 (tube 10) under the postalveolar area.

For the F3 resonance mode, $A_{p\,max}$ is the area of the pressure maximum at the 2-tube foreblade (No. 5–6); $A_{U\,max}$ is the area of the volume velocity maximum at the 2-tube hindblade (No. 9–10). Because the general proportionality for the formant frequency is

$F_i \propto (\lambda_i / 4)^{-1}(A_{U\,max} / A_{p\,max})^{a/b}$, the F3 frequency should be raised when the mean foreblade area $\overline{A}(foreblade)$ $(= A_{p\,max})$ becomes smaller, and vice versa. The F3 frequency should be lowered when the mean hindblade area $\overline{A}(hindblade)$ $(= A_{U\,max})$ becomes smaller, and vice versa.

As the blade constriction moves from the volume velocity maximum at the postalveolar index 1 (reduced $A_{U\,max}$) to the pressure maximum at the front teeth index 6 (reduced $A_{p\,max}$), the F3 frequency is expected to rise from its lowest to its highest value. On the basis of known formant measurements, Fant (1960, p. 26) observed that "a very low F_3 signals retroflex modification" while "a high F_3 indicates a prepalatal or dental articulation."

In Appendix B, the first- and second-ranked correlations for log F3 $(= \log_2 F3)$ in the original data set are the blade positions: (1st) $\log_2 \overline{A}(foreblade) / \overline{A}(hindblade)$ r = –0.5744 and (2nd) $\overline{A}(foreblade) / \overline{A}(hindblade)$ r = –0.4875. The third- and fourth-ranked correlations for log F3 in the Tiede data set are also the blade positions: (3rd) $\log_2 \overline{A}(foreblade) / \overline{A}(hindblade)$ r = –0.5595 and (4th) $\overline{A}(foreblade) / \overline{A}(hindblade)$ r = –0.5518. In the original data set the fifth-ranked correlation for log F3 is the blade constriction (5th) $I_{\min A}(blade)$ r = 0.4003; in the Tiede data set there is a low-moderate direct correlation between log F3 and the blade constriction $I_{\min A}(blade)$ r = 0.3208 (s.d. 0.386). These correlation results show unambiguously that the blade position $\overline{A}(foreblade) / \overline{A}(hindblade)$ is a better articulatory parameter for the F3 frequency than the blade constriction $I_{\min A}(blade)$.

In order to determine how much log F3 decreases as the blade position $\overline{A}(foreblade) / \overline{A}(hindblade)$ increases, log F3 regression slopes are computed. The original data set has a mean slope of –0.087 octave (s.d. 0.05) per doubling of $\overline{A}(foreblade) / \overline{A}(hindblade)$. The Tiede data set has a mean slope of –0.112 octave (s.d. 0.10) per doubling of $\overline{A}(foreblade) / \overline{A}(hindblade)$. The average slope for the two data sets is –0.100 octave per doubling of $\overline{A}(foreblade) / \overline{A}(hindblade)$. In Section 3.1.3 the analytical F3 proportionality is $F_3 \propto (\lambda_3 / 4)^{-1}(A_{U\,max} / A_{p\,max})^{2/27}$, where the theoretical fractional exponent a / b is 2/27. Hence when the foreblade area $A_{p\,max}$ is doubled or the hindblade area $A_{U\,max}$ is halved, the F3 frequency decreases by –2/27 (–0.074) octave $(= \log_2(2^{-2/27}))$. As the average regression slope is –0.100 octave per doubling of $\overline{A}(foreblade) / \overline{A}(hindblade)$, the experimental fractional exponent a / b is about ~ 1/12 (0.083). Therefore the empirical F3 proportionality is $F_3 \propto (\lambda_3 / 4)^{-1}(A_{U\,max} / A_{p\,max})^{1/12}$.

The blade constriction $I_{\min A}(blade)$ specifies whether there is a constriction at the pressure maximum (reduced $A_{p\,max}$) for dental articulations or a constriction at the

volume velocity maximum (reduced $A_{U\,max}$) for postalveolar articulations. On the other hand, the blade position $\overline{A}(foreblade)/\overline{A}(hindblade)$ registers simultaneously both the foreblade area ($=A_{p\,max}$) and the hindblade area ($=A_{U\,max}$). Thus the blade position is a more sensitive articulatory parameter for the F3 frequency. Even so, blade position does not exclude blade constriction. Given that $F_3 \propto (\lambda_3/4)^{-1}(A_{U\,max}/A_{p\,max})^{1/12}$, a dental constriction at the pressure maximum (reduced $A_{p\,max}$) or a postalveolar constriction at the volume velocity maximum (reduced $A_{U\,max}$) may each be complemented by a dilation at the opposite end of the blade region.

9.3. Acoustic-articulatory correlations: further observations on the first formant

9.3.1. Vowel height

The Handbook of the International Phonetic Association (1999) describes the vowel [i] as close "since the tongue is near the roof of the mouth" (p. 10), whereas the vowel [ɑ] is described as open because "the space between the tongue and the roof of the mouth is as large as possible (p. 11)." These descriptions are quite similar to those put forward by Jones (1922, p. 17) where close vowels are "those in which the tongue is as high as possible" and open vowels are "those in which the tongue is as low as possible." However Wood (1975) noted that early X-ray studies (e.g. Russell, 1928) cast serious doubt on the equivalence between tongue height and IPA vowel height represented by the vertical axis of the IPA vowel quadrilateral. To pursue the question further, he examined tongue heights in X-ray tracings from 15 languages and discovered frequent reversals of vertical position among the traditional high [ɪ] and mid [e] vowels as well as the mid [ɔ] and low [a] vowels. Joos (1948, p. 50) stated that "it is generally agreed that the principal determinant of vowel identification in the articulatory field is the tongue position specified in two dimensions: (1) high-mid-low; (2) front-central-back." From spectrogram measurements he showed that the logarithmic F1 frequency increases as the IPA tongue position goes from high to low but concluded that (p. 54) "it appears that the classical or IPA tongue-position quadrilateral rather more closely resembles the acoustic vowel quadrilateral than it resembles the Carmody X-ray tongue-position quadrilateral." Delattre, Liberman, Cooper, and Gerstman (1952) synthesized two-formant vowels exemplified in the cardinal vowel quadrilateral. They found that a significant increase in F1 frequency caused a cardinal vowel to be identified as a more open one on the IPA chart. Subsequently, the hypothesis came to be accepted that F1 frequency increases when the tongue opens wider though some uncertainty persists (see Catford, 1981 for further elaboration).

To evaluate the tongue height hypothesis, the correlations between log F1 and the mean body area $\overline{A}(body)$ are provided. In the original data set, they are $\overline{A}(body)$

r = –0.1488 (s.d. 0.567); $\log_2 \overline{A}(body)$ r = –0.1541 (s.d. 0.532). In the Tiede data set, they are $\overline{A}(body)$ r = –0.5534 (s.d. 0.274); $\log_2 \overline{A}(body)$ r = –0.4891 (s.d. 0.270). The very weak to moderate inverse correlations reveal that Fl frequency tends to decrease as the tongue body region becomes more open. Yet, according to the tongue height hypothesis, the F1 frequency should *increase* when the tongue body region opens.

In Section 9.2.1, the F1 frequency was shown to have strong to very strong inverse correlations with the tongue root aperture $\overline{A}(root) / \overline{A}(lip)$, in which the mean tongue root area $\overline{A}(root)$ is the dominant term. To gain insight into a possible motor association between the tongue body and tongue root regions, the mean tongue body area $\overline{A}(body)$ is correlated with the mean tongue root area $\overline{A}(root)$. In the original data set, the Pearson coefficient is 0.2173 (s.d. 0.571); in the Tiede data, the Pearson coefficient is 0.5879 (s.d. 0.292). Hence there is a weak to moderate direct correlation between the areas of the tongue body and the tongue root, indicating that both regions tend to widen and narrow together. This accords with the preceding paragraph in which the F1 frequency decreases with the area of the tongue body region, albeit in a much less robust way than for the tongue root region.

Seeing that Fl frequency not only fails to increase but tends to decrease with the area of the tongue body region, it is of interest to consider the area of the lip region. For the F1 resonance mode, $A_{U\,max}$ is the area of the volume velocity maximum at the lips. Since $F_1 \propto (\lambda_1 / 4)^{-1} (A_{U\,max} / A_{p\,max})^{1/3}$, the F1 frequency should increase as the lip region $(= A_{U\,max})$ becomes more open, and vice versa. In the original data set (Appendix B), there is a fifth-ranked direct correlation between log F1 and the lip area $\log_2 \overline{A}(lip)$ r = 0.7226. In the Tiede data set, there are moderate direct correlations between log F1 and the mean lip area: $\overline{A}(lip)$ r = 0.6513 (s.d. 0.207); $\log_2 \overline{A}(lip)$ r = 0.6689 (s.d. 0.210). The direct correlations between log F1 and mean lip area $\overline{A}(lip)$ confirm that F1 frequency increases with the area of the lip region. Nevertheless, the high-moderate correlation values are smaller than the strong to very strong inverse correlations that F1 frequency has with the tongue root aperture $\overline{A}(root) / \overline{A}(lip)$ and the mean root area $\overline{A}(root)$. It was pointed out in Section 9.2.1 that the influence of the lip area is diminished relative to the tongue root area because the 4-tube lip region is half the length of the 8-tube tongue root region. To determine the direction and strength of the motor association between the lip and tongue root regions, the mean lip area $\overline{A}(lip)$ is correlated with the mean tongue root area $\overline{A}(root)$. In the original data set, the Pearson coefficient is –0.5073 (s.d. 0.241); in the Tiede data set, the Pearson coefficient is –0.5325 (s.d. 0.252). Hence a moderate inverse correlation exists between the areas of the lip and tongue root regions. This inverse motor association is anticipated in view of the trading relation given by the tongue root aperture $\overline{A}(root) / \overline{A}(lip)$ for the F1 frequency.

According to the perturbation analysis, area changes must be made near the pressure maximum or the volume velocity maximum to raise or lower the formant frequency (Sections 3.1.1–3.1.3). However the requirement is not satisfied for the tongue body region because of its intermediate location. The 9-tube tongue body region is 10 tubes away from the F1 volume velocity maximum at the lips and 8 tubes away from the F1 pressure maximum at the glottis (cf. Fig. 1). The somewhat closer proximity of the tongue body region to the glottis may explain why the F1 frequency decreases with the area of the tongue body region as it does with the tongue root region. Nevertheless, the inverse correlations between F1 frequency and the area of the tongue body region are still only very weak to moderate.

Chomsky and Halle (1968) specifically identified the tongue body as the main determinant of vowel height (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 tongue body features [high low] are retained in more recent feature proposals (Halle & Stevens, 1991; Stevens, 1998, p. 250; Halle, Vaux, & Wolfe, 2000; Gussenhoven & Jacobs, 2017, p. 78). Stevens (1998) takes the F1 frequency to be the acoustic cue for the tongue body features [high low], whereby F1 increases as the tongue body is displaced from a [+high] to a [+low] position (but see Chapter 6 for an extended discussion of F1 and the pharyngeal region). Thus the tongue height hypothesis continues to find acceptance.

The correlation results demonstrate, however, that F1 frequency not only fails to increase with the area of the tongue body region, but tends to decrease instead. Furthermore, the intermediate location of the tongue body region provides a theoretical explanation for the observed behavior of the F1 frequency. To address these difficulties, the features [high low] may each be redefined as the vertical position of the inferior lip because the F1 frequency increases with the area of the lip region. Hence there is an equivalence between inferior lip height and IPA vowel height, in which F1 frequency increases with inferior lip opening—rather than tongue body opening as in the tongue height hypothesis. This interpretation has the benefit that other traditional terms for vowel height like *mid*, *raised*, and *lowered* remain meaningful.

A more accurate alternative to inferior lip height is replacing [high low] with the corresponding tongue root features [expanded contracted] since the tongue root area $\overline{A}(root)$ is the dominant term in the tongue root aperture $\overline{A}(root)/\overline{A}(lip)$. The primary feature [expanded] is adapted from Lindau (1979). The opposing primary feature [contracted] is similar to Perkell's term *constricted pharynx* (1971). The terms *root-*

advanced and *root-unadvanced* were first employed by Stewart (1967) to distinguish raised and unraised vowels in the cross-height vowel harmony of Akan. Later the same distinction was made using the secondary feature [ATR] (Advanced Tongue Root) which contrasts (raised) [+ATR] vowels and (non-raised) [–ATR] vowels (Halle & Stevens, 1969; see Trigo, 1991 for a survey). The opposing secondary feature [RTR] (Retracted Tongue Root) denotes the contrast between (lowered) [+RTR] vowels and (non-lowered) [–RTR] vowels (Vaux, 1999). The F1 frequency must be the acoustic cue for the secondary tongue root features [ATR RTR] because a [+ATR] vowel has a lower F1 than a [–ATR] vowel (Halle & Stevens, 1969). Therefore the F1 frequency serves as the acoustic cue for both the primary and secondary tongue root features [expanded contracted, ATR RTR]. Although the tongue root terms are notationally convenient, it should be emphasized that each tongue root feature characterizes the tongue root aperture $\overline{A}(root) / \overline{A}(lip)$, and not solely the dominant term $\overline{A}(root)$.

9.3.2. The quality factor Q1 and overall power PWR

Because the quality factor Q1 is equal to $F1/B1$, a linear relationship will occur between Q1 and the F1 frequency if the half-power bandwidth B1 remains constant. Then log Q1 $(= 20\log_{10}(F1/B1))$ increases by +6 dB for each octave or doubling of F1 frequency. Linear regression is applied to each data set with log Q1 being the dependent variable and log F1 $(= \log_2 F1)$ the independent variable. The original data set (N = 10) yields a mean slope of +7.61 dB per octave (s.d. 1.94), the Tiede data set (N = 12) a mean slope of +10.32 dB per octave (s.d. 0.91). Since losses due to wall damping diminish with increasing frequency, the B1 bandwidth becomes smaller as the F1 frequency increases (Flanagan, 1972, p. 69). As a result, the log Q1 slopes grow faster than +6 dB/octave. In Appendix B the original and Tiede correlations for log Q1 both display the highest ranking inverse associations with the tongue root aperture $\overline{A}(body) / \overline{A}(lip)$ and the mean root area $\overline{A}(root)$. The rankings thus resemble those found for log F1, which is not surprising under the assumption of a roughly linear relationship between Q1 and F1 frequency.

It was noted in Section 4 that the viscous and heat conduction bandwidths B_{visc} and B_{heat} are proportional to the square root of the formant frequency: $B \propto \sqrt{F}$. The kinetic energy radiation bandwidth B_{rad}^{KE} is $2F^2 A_{VT} / cl$; the potential energy radiation bandwidth B_{rad}^{PE} is $2F^2 A_0^2 / clA_{VT}$. Hence all the bandwidths grow larger as the formant frequency *F* increases. The greater damping causes a reduction in amplitude of the higher formants. If most of the vowel power is assumed to be concentrated in the first formant and therefore nearly equal to the relative power log Q1, then the overall power PWR should exhibit the same rough linearity with log F1 as log Q1 does. The mean regression slopes for PWR are +8.07 dB per octave (s.d. 2.31) for the original data set and +6.89 dB

per octave (s.d. 2.03) for the Tiede data set. Furthermore, the correlation rankings of PWR in Appendix B do not deviate appreciably from those of log Q1 for the original data set. The quasi-linear relationship between vowel power and logarithmic F1 frequency offers a straightforward explanation for the phenomenon of vowel specific intensity, in which (low) [+contracted] vowels with a high F1 frequency have greater intensity than (high) [+expanded] vowels with a low F1 frequency (for a review of vowel specific intensity, see Rossi, 1971).

9.3.3. The tongue root constriction $I_{\min A}(root)$

The author concluded in Pennington (2011) that F1 frequency increases as the location of the tongue root constriction moves upward in the pharynx. The inference was based on the moderate direct correlation (r = 0.5453, s.d. 0.213) between log F1 and the tongue root constriction $I_{\min A}(root)$ in the original data set. However there is only a weak direct correlation (r = 0.2272, s.d. 0.332) between log F1 and the tongue root constriction $I_{\min A}(root)$ in the Tiede data set. In contrast to the above result, Wood's X-ray observations (1979) indicated that the constriction locations for (low) [–expanded +contracted] vowels such as [ɑ a æ] with high F1s are in the lower pharynx, whereas the constriction locations for (mid) [–expanded –contracted] vowels like [o ɔ] with lower F1s are in the upper pharynx. To help clarify the mixed findings, the FDVT program calculates the F1 frequencies of 4 lossy two-pipe models 17 cm in length and 2 cm^2 in default area: (i) uniform pipe, (ii) whole pharynx constriction where the area of all 8 tongue-region tubes (tubes 20–27) is 1 cm^2, (iii) lower pharynx constriction where the area of the 4 tongue-region tubes nearest the glottis (tubes 24–27) is 1 cm^2, and (iv) upper pharynx constriction where the area of the 4 tongue-region tubes nearest the lips (tubes 20–23) is 1 cm^2. The F1 frequencies of the uniform pipe, the whole pharynx constriction, the lower pharynx constriction, and the upper pharynx constriction are (i) 536 Hz, (ii) 611 Hz, (iii) 573 Hz, (iv) 563 Hz. Whole pharynx constriction is evidently the most effective way to increase F1 frequency (536 to 661 Hz), followed by lower pharynx constriction (536 to 573 Hz), then upper pharynx constriction (536 to 563 Hz). Because the four lower pharynx tubes are closer to the F1 pressure maximum at the glottis than the four upper pharynx ones, the lower pharynx constriction increases the F1 frequency more efficiently, as expected from the perturbation analysis. Hence there is a theoretical basis for Wood's observations. On the other hand, the frequency difference between lower and upper pharynx constrictions is small (573 vs. 563 Hz), which may account for the variability of the correlations between log F1 and tongue root constriction $I_{\min A}(root)$.

9.4. Acoustic-articulatory correlations: the Q2, Q3, and Q4 quality factors

9.4.1. The Q2 and Q3 quality factors: pressure types of quarter wavelength

In Section 5.4.4, the full-loss two-pipe model displays a log Q2 regression slope of +2.10 dB per doubling of tongue body area when the lip area is held constant. The original and Tiede vowel data sets yield respectively mean log Q2 regression slopes of +2.53 dB (s.d. 2.77) and +2.53 dB (s.d. 0.98) per doubling of the tongue body aperture $\overline{A}(body) / \overline{A}(lip)$.

Also in Section 5.4.4, the full-loss two-pipe model shows a log Q3 regression slope of +4.40 dB per doubling of blade area when the lip area is held constant. The original and Tiede vowel data sets give respectively mean log Q3 regression slopes of +3.85 dB (s.d. 2.62) and +3.31 dB (s.d. 1.71) per doubling of the blade aperture $\overline{A}(blade) / \overline{A}(lip)$.

Hence for the lossy two-pipe models and vowel data sets alike, the log Q2 and log Q3 regression slopes reveal an increase on the order of ~ +3 dB for each doubling of quarter wavelength area or area ratio. Recall that with radiation loss alone, there is an increase of +6 dB per doubling of $A_{\lambda/4}$ in accordance with $Q_{rad}^{PE} \propto A_{\lambda/4} / A_0^2$ for the pressure type of quarter wavelength (Section 5.3). The tongue body aperture $\overline{A}(body) / \overline{A}(lip)$ and the blade aperture $\overline{A}(blade) / \overline{A}(lip)$ have about the same regression slopes as the two-pipe models with a constant lip area. Thus the vowel data sets suggest that the lip area $\overline{A}(lip)$ is relatively constant when the tongue body area $\overline{A}(body)$ or the blade area $\overline{A}(blade)$ is varied. It was mentioned at the end of Section 5.4.2 that the square of the terminating lip area A_0 grows substantially faster than the quarter wavelength area $A_{\lambda/4}$ in the quality factor $Q_{rad}^{PE} \propto A_{\lambda/4} / A_0^2$. Simply from a theoretical standpoint then, the area of the lip region should be fixed relative to the variable area $A_{\lambda/4}$ of the blade or tongue body region.

An inspection of the original and Tiede correlations in Appendix B finds that:

1. log Q2 has the highest ranking direct correlations with the tongue body aperture $\overline{A}(body) / \overline{A}(lip)$ and its base-2 logarithm $\log_2 \overline{A}(body) / \overline{A}(lip)$;
2. log Q3 has very high ranking direct correlations with the blade aperture $\overline{A}(blade) / \overline{A}(lip)$ and its base-2 logarithm $\log_2 \overline{A}(blade) / \overline{A}(lip)$.

The results from both data sets demonstrate that log Q2 and log Q3 are the optimal acoustic correlates of the tongue body aperture $\overline{A}(body) / \overline{A}(lip)$ and the blade apeature $\overline{A}(blade) / \overline{A}(lip)$.

Flanagan (1957) investigated the just-noticeable difference (JND) for the second formant (F2) amplitude of the synthetic vowel [æ]. He obtained a JND of about 3 dB from 224 judgments of four subjects. Since $10\log_{10} 2 = +3$ dB and $10\log_{10}(1/2) = -3$ dB,

the power of F2 must be either doubled or halved in order for a difference to be detected. As was shown above, the relative power log Q2 is roughly doubled (~ + 3 dB) for each doubling of the tongue body aperture $\overline{A}(body) / \overline{A}(lip)$. Therefore a doubling of tongue body aperture is the smallest motor change that produces an auditorily detectable difference in F2 amplitude.

Two reviews of studies on vowel bandwidth found that B1–B3 bandwidth values have (i) "little effect on synthetic vowel recognition" (Hawks & Miller, 1995, p. 1343); (ii) "only a small effect on vowel identity in perceptual studies" (Kent & Vorperian, 2018, p. 87). However in 1956, House and Stevens (p. 225) confirmed earlier reports suggesting that B1 bandwidth is the main cue for vowel nasalization. In addition, the corresponding quality factor Q1 serves to distinguish between the transient phases of buccal stops [d g b] and those of nasals [n ŋ m] (cf. Section 10.2 below). Using an auditory estimator PE3 of the quality factor Q3, Pennington (2017) analyzed the coronal consonants of American English, Toda, Ubykh, and Central Arrernte. Laminal coronals with a narrow blade aperture $\overline{A}(blade) / \overline{A}(lip)$ revealed smaller PE3 measures than apical coronals with a wider blade aperture (for more details, see final Section 11). Although B3 bandwidth seems to have little influence on the identity of vowels, its corresponding quality factor Q3 is the cue for the opposition between apical and laminal coronal consonants. Thus for coronal and nasal consonants, the contrastive function of formant bandwidth appears well-established.

9.4.2. The Q4 quality factor: volume velocity type of quarter wavelength

In Section 5.4.4, the full-loss two-pipe model exhibits a log Q4 regression slope of –4.28 dB per doubling of lip area when the area of the rest of the pipe is held constant. The original and Tiede vowel data sets yield respectively mean log Q4 regression slopes of –5.67 dB (s.d. 3.76) and +1.39 dB (s.d. 1.82) per doubling of the lip area $\overline{A}(lip)$. The positive log Q4 slope of +1.39 dB obtained for the Tiede data set is clearly an outlier. With radiation loss alone, there should be a negative log Q4 slope of –6 dB per doubling of lip area $A_{\lambda/4}$ in accordance with $Q_{rad}^{KE} \propto 1 / A_{A_{\lambda/4}}$ for the volume velocity type of quarter wavelength (Section 5.3).

For the original data set in Appendix B, the first- and second-ranked log Q4 correlations are: (1^{st}) $\overline{A}(lip)$ r = –0.6222 and (2^{nd}) $\log_2 \overline{A}(lip)$ r = –0.6194. These highest ranking inverse correlations between log Q4 and the lip area $\overline{A}(lip)$ conform to the anticipated result. On the other hand, in the Tiede data set there are very weak direct correlations between log Q4 and lip area: $\overline{A}(lip)$ r = 0.1938 (s.d. 0.258); $\log_2 \overline{A}(lip)$ r = 0.2095 (s.d. 0.235).

It is evident that the Tiede data set fails to meet the expectations of both a negative log Q4 regression slope of ~ –3 dB per doubling of lip area $\overline{A}(lip)$ and a high

ranking inverse correlation between log Q4 and lip area $\overline{A}(lip)$. Teeth are not easily visualized with conventional MRI because of their very low water content. Consequently, tooth contours are segmented manually and then superimposed on the vowel scans. Tiede (2013, p. 31) mentioned that the incisor contours were of marginal accuracy. Therefore errors could have occurred in the determination of the area of the 4-tube lip region.

9.5. Acoustic-articulatory correlations: the F4 frequency

In Section 3.1.1 the first four formant frequencies of a lossless open-closed pipe of length *l* were shown to be

$$F_i = \frac{2i-1}{4} \times \frac{c}{l},$$

where $1 \le i \le 4$ is the formant number. Each formant frequency F_i is then inversely proportional to the vocal tract length:

$$F_i \propto \frac{1}{l}.$$

The inverse proportionality between F_i and vocal tract length *l* is the same as the inverse proportionality between F_i and any other fraction of the vocal tract length, including the quarter wavelength $\lambda_i / 4$:

$$F_i \propto \frac{1}{l} \propto \frac{1}{l/(2i-1)} = \frac{1}{\lambda_i / 4}.$$

This reflects the fact that the relation of proportionality is invariant under the linear operations of multiplication or division. In the proportionality above, for example, the vocal tract length *l* is divided by the integer constant $2i - 1$. A correlation variable may similarly be multiplied or divided by a constant without changing the value of the coefficient. Thus the correlation between formant frequency F_i and its quarter wavelength $\lambda_i / 4$ should be numerically the same as the correlation between F_i and vocal tract length *l*.

The lip length $L(lip)$ is the sum of the tube lengths in the lip region (tubes 1–4). As the tube lengths are equal (Section 6), $L(lip)$ is 4/27 the length of the 27-tube vocal tract. Therefore at the F4 resonance mode, the lip length $L(lip)$ is very nearly one F4 quarter wavelength long or one-seventh (4/28) the length of the vocal tract. In Appendix B, the original data set reveals that the first- and third-ranked correlations for log F4 ($= \log_2 F4$) are: (1^{st}) $\log_2 L(lip)$ r = –0.3911 and (3^{rd}) $L(lip)$ r = –0.3884. The Tiede data set shows that the first- and second-ranked correlations for log F4 are: (1^{st}) $L(lip)$ r = –0.6991 and (2^{nd}) $\log_2 L(lip)$ r = –0.6976. As predicted, these log F4 correlations for lip length $L(lip)$ are identical to those computed for vocal tract length *l*.

Due to the general proportionality $F_i \propto (\lambda_i / 4)^{-1}$, the F4 frequency is expected to be inversely proportional to the F4 quarter wavelength $F_4 \propto (\lambda_4 / 4)^{-1}$. Hence when $\lambda_4 / 4$ is doubled, F_4 should decrease by a factor of 2^{-1} or a negative change of –1 octave ($= \log_2(2^{-1})$). To estimate how much log F4 decreases as the lip length $L(lip)$ increases, log F4 regression slopes are calculated. The original data set yields a mean slope of –0.731 octave (s.d. 1.03) per doubling of $L(lip)$. The Tiede data set has a mean slope of –0.992 octave (s.d. 0.61) per doubling of $L(lip)$. Thus the Tiede data set confirms the anticipated log F4 decrease of –1 octave for each doubling of $L(lip)$. This outcome accords with the much stronger correlations between log F4 and $L(lip)$ for the Tiede data set than for the original data set. In consequence, log F4 appears to be the optimal acoustic cue for the lip length $L(lip)$.

Lip protrusion is not responsible for all changes in vocal tract length. Raising and lowering the larynx can also shorten and lengthen the vocal tract (Ewan & Krones, 1974). Wood (1979, p. 33) remarked that the larynx tends to be lower for protruded vowels like [u o y ø] than for drawn vowels like [ɨ ɤ i ɛ]. Hoole and Kroos (1998) discovered substantial inter-speaker variability in larynx height. Two of their three male German speakers showed a maximum difference of 0.7–1.0 cm between protruded and drawn vowels, yet the other displayed only a 0.2 cm difference. Riordan (1977, Fig. 5) found an equally small difference in larynx height between protruded and drawn vowels for two of four male French speakers. Because the F4 frequency is identically correlated with both lip length $L(lip)$ and total vocal tract length, larynx height must have an acoustic effect. Nevertheless, lip protrusion will be considered the primary determinant of vocal tract length for the following reasons:

(i) Differences in larynx height are rather variable for the German and French speakers. Note that the two languages contrast the protruded and drawn vowel qualities [y ø œ] and [i e ɛ].

(ii) For a female French speaker the largest difference in lip protrusion is observed to be about 2 cm (Abry & Boë, 1986, p. 101). However only one of four male French speakers reaches a larynx height difference of 1 cm (Riordan, 1977).

(iii) The lips constitute the most visible active articulator (see Rosenblum, 2008 for a brief review of the lips in visual speech perception).

9.6. Acoustic-articulatory correlations: summary of the results

The F1–F3 formant frequencies and their articulatory correlates are as follows:

1. log F1 and tongue root aperture $\overline{A}(root) / \overline{A}(lip)$ are inversely correlated (see Section 9.3.2 where log Q1 and PWR also pattern like log F1 with respect to $\overline{A}(root) / \overline{A}(lip)$).

2. log F2 and tongue body position $\overline{A}(forebody) / \overline{A}(hindbody)$ are inversely correlated.
3. log F3 and blade position $\overline{A}(foreblade) / \overline{A}(hindblade)$ are inversely correlated.

As seen throughout Sections 9.2.1–9.2.3, these inverse correlations result from the general proportionality for the formant frequency $F_i \propto (\lambda_i / 4)^{-1} (A_{U\,max} / A_{p\,max})^{a/b}$, where $A_{U\,max}$ is the area at the volume velocity maximum, $A_{p\,max}$ the area at the pressure maximum, and a / b the fractional exponent. For the F1 frequency, the theoretical fractional exponent is 1/3 (0.333). Since the average regression slope of the vowel data sets is –0.324 octave per doubling of $\overline{A}(root) / \overline{A}(lip)$, the experimental fractional exponent a / b is ~ 1/3 (0.333). For the F2 frequency, the theoretical fractional exponent is 1/9 (0.111). Since the average regression slope is –0.156 octave per doubling of $\overline{A}(forebody) / \overline{A}(hindbody)$, the experimental fractional exponent a / b is ~ 1/6 (0.167). For the F3 frequency, the theoretical fractional exponent is 2/27 (0.074). Since the average regression slope is –0.100 octave per doubling of $\overline{A}(foreblade) / \overline{A}(hindblade)$, the experimental fractional exponent a / b is ~ 1/12 (0.083). Hence considerable similarity exists between the theoretical and experimental fractional exponents. This is strong evidence that (a) the general proportionality $F_i \propto (\lambda_i / 4)^{-1} (A_{U\,max} / A_{p\,max})^{a/b}$ applies to the area functions of spoken vowels and (b) the perturbation analysis remains valid when losses due to radiation, viscosity, heat conduction, and wall damping are implemented.

4. log F4 and lip length $L(lip)$ are inversely correlated.

This inverse correlation is due to the inverse proportionality $F_4 \propto (\lambda_4 / 4)^{-1}$, where the lip length $L(lip)$ spans the F4 quarter wavelength $\lambda_4 / 4$ (see the preceding Section 9.5).

The Q2–Q3 quality factors and their articulatory correlates are as follows:

1. log Q2 and tongue body aperture $\overline{A}(body) / \overline{A}(lip)$ are directly correlated.
2. log Q3 and blade aperture $\overline{A}(blade) / \overline{A}(lip)$ are directly correlated.

These direct correlations result from the proportionality $Q_{rad}^{PE} \propto A_{\lambda/4} / A_0^2$. The area $A_{\lambda/4}$ is associated with a pressure type of quarter wavelength. The terminating lip area A_0 appears to be relatively constant in the vowel data sets (Section 9.4.1).

3. log Q4 and lip area $\overline{A}(lip)$ are inversely correlated.

This inverse correlation is due to the inverse proportionality $Q_{rad}^{KE} \propto 1 / A_{A_{\lambda/4}}$, where the area $A_{\lambda/4}$ is associated with a volume velocity type of quarter wavelength (Section 9.4.2).

The acoustic parameters in the foregoing summary are logarithmic. The formant frequency F (Hz) is converted to its octave value: log F $= \log_2 F$ (Section 8). The dimensionless quality factor Q is converted to a measure of relative power: log Q $= 20 \log_{10} Q$ (Section 2). To transform a proportionality logarithmically, one must take the logarithm of both sides of the expression. In view of the proportionalities between the

acoustic and articulatory parameters described above, the articulatory parameters should in principle be logarithmic because the acoustic parameters are logarithmic. Nevertheless, there are a number of instances in Appendix B where the absolute value of the mean correlation coefficient *r* is smaller for the logarithmic articulatory variable than for the non-logarithmic one. To illustrate this with log Q2 correlations, both the original and the Tiede data sets show that $\overline{A}(body) / \overline{A}(lip)$ is ranked first, whereas $\log_2 \overline{A}(body) / \overline{A}(lip)$ is ranked second. Hence logarithmic transformation of the articulatory parameters does not appear to be a strict necessity.

Part IV. Consonants, overview table, and future direction

10. Consonants

Sections 10.1 and 10.2 below summarize the results obtained by the FDVT program for the area functions of buccal obstruents and nasals. The complete data and analysis are presented in Pennington (2017, Section 2.7).

10.1. Consonants: buccal obstruents

The area functions of [t], [k], and [p] (Story, Titze, & Hoffman, 1996) are fit to the 27 tubes of the FVDT model. The areas of the tube corresponding to the stop closure are set to 0.05, 0.1 and 0.2 cm^2, thereby simulating a stop transient or fricative (cf. the 0.05–0.2 cm^2 fricative range of Stevens, 1998, p. 33). The areas of the [t] closure are additionally set to 0.0125, 0.025 and 0.4 cm^2. To evaluate the effects of spectral zeros produced by supraglottal excitation (Fant, 1960, p. 42; Flanagan, 1972, pp. 72–73), each area function is calculated with a glottal and a supraglottal source. Like the vowels earlier, the volume velocity source at the glottis (tube 27) is assigned the value of 1 for all frequencies (12–6502 Hz). The series pressure sources of [t] (tube 5), [k] (tube 12) and [p] (tube 1) are also assigned the value of 1 for all frequencies (regarding the series connection of the supraglottal pressure source, see Fant, 1960, p. 36; Flanagan, 1972, p. 54; Stevens, 1998, p. 102).

The results for the F1–F4 formant frequencies reveal that both the glottal and the supraglottal sources excite the odd resonance modes of the vocal tract. Even when the stop closure of [t] is set to the smallest area of 0.0125 cm^2, the odd resonance frequencies are maintained. For the glottal source, they are $F1_{gl}$ 235 Hz, $F2_{gl}$ 1674 Hz, $F3_{gl}$ 2656 Hz, $F4_{gl}$ 3281 Hz. For the supraglottal source, they are $F1_{sgl}$ 247 Hz, $F2_{sgl}$ 1681 Hz, $F3_{sgl}$ 2663 Hz, $F4_{sgl}$ 3282 Hz. Formant frequency differences between glottal and supraglottal sources are generally slight like those of [t] with the 0.0125 cm^2 closure. Quality factor differences between the two source locations are also minor, with two classes of

exceptions. For [k] and [p], the supraglottal log Q4 values are considerably less than the glottal ones. For the smaller [t] closures of 0.0125 and 0.025 cm^2, the supraglottal log Q2 and log Q3 values are also less than their glottal counterparts.

Odd-numbered resonances are observed for the F2–F4 frequencies of Swedish, American English, and Polish fricatives (Jassem, 1965, Table 1). The odd resonance modes show that the open-closed boundary conditions are satisfied not only for the synthetic stop transients and fricatives, but also for naturally occurring fricatives. Consequently, the acoustic-articulatory rules established on the basis of the vowels and two-pipe models are equally applicable to the buccal obstruents.

10.2. Consonants: nasals

To analyze the transient phase of nasals, a side branch is placed in parallel with tube 20 of the FDVT model, the most anterior tube of the tongue root region. The nasal pipe, 19 vocal tract (VT) tubes long , consists of three tubes: the nasal port (variable area, 4 VT tubes long), the nasal cavity (3 cm^2, 11 VT tubes long), and the nostrils (1.5 cm^2, 4 VT tubes long). The shape factors of the port, cavity, and nostril tubes are fixed at 1, 4, and 2, respectively (Dang, Honda, & Suzuki, 1994). Without extra damping. the calculated nasal resonance in the range of 550–1100 Hz would be too pronounced because nasal peaks exceeding the first formant are not often observed in nasalized vowel spectra (Chen, 1997). To keep the level of the nasal peak generally below that of the first formant, a resistance of 30 cgs acoustic ohms is added in series to the nasal cavity tube. The transient phase of nasals is modeled using (a) the area functions of [t k p] as in the previous section, (b) a buccal closure set to 0.05 cm^2 like a stop transient, and (c) a glottal source. Two nasal port areas are selected: (i) Stevens' estimate of 0.2 cm^2 for nasal consonants (1998, p. 487) and (ii) a smaller one of 0.05 cm^2.

The overall results indicate that log Q1 decreases and F1 frequency increases as the nasal port area $A(nasal)$ grows larger, whereas the Q2–Q4 quality factors and the F2–F4 formant frequencies remain almost unchanged. The results are in agreement with those of House and Stevens (1956) as well as later nasalization studies (cf. the literature survey by Pruthi, 2007, Chapter 2). For [n], the log Q1 difference between the nasal port areas of 0.05 and 0.2 cm^2 is 16.25–8.74 = 7.51 dB; for [ŋ], the log Q1 difference between the two areas is 14.14–6.53 = 7.61 dB; for [m], the log Q1 difference between the two areas is 16.39–9.27 = 7.12 dB. The average log Q1 decline is then –7.41 dB per quadrupling of the nasal port area $A(nasal)$, or equivalently, ~ –3 dB per doubling of $A(nasal)$.

Because $Q1 = F1 / B1$, a linear relation is expected between the Q1 quality factor and the F1 frequency on the condition that B1 stays constant. Vowel regression analyses in Section 9.3.2 confirm the quasi-linear relation between Q1 and F1, whereby log Q1 increases somewhat more than +6 dB for each doubling of F1 frequency. A large F1 shift

will evidently give rise to a large Q1 shift. Hence the Q1 quality factor can be used as an acoustic cue for the nasal port area $A(nasal)$ only at a given F1 frequency. Then log Q1 will decrease ~ –3 dB per doubling of $A(nasal)$. The Q1 quality factor illustrates the notion of relational invariance in which the acoustic cue for a distinctive feature "is normalized with respect to individual and contextual factors (Fant, 1986, p. 486)."

11. Toward distinctive features of the vocal tract

The acoustic-articulatory associations found in this vocal tract study are displayed in Table 3 below. The table also gives the section where each acoustic-articulatory correspondence is discussed. Remark that the relative power log Q is roughly doubled (~ + 3 dB) for every doubling of the blade and tongue body apertures; it is roughly halved (~ – 3 dB) for every doubling of the lip and nasal port areas. Thus an overview of the findings reveals important regularities.

To demonstrate how the acoustic-articulatory correspondences can provide a foundation for distinctive features, the acoustic-articulatory correlates of the blade will be briefly considered: (a) log F3 and blade position $\overline{A}(foreblade) / \overline{A}(hindblade)$; (b) log Q3 and blade aperture $\overline{A}(blade) / \overline{A}(lip)$. The following paragraph summarizes some of the observations and conclusions in Pennington (2017, Section 3).

Formant measurements indicate that F3 frequencies range over 6 one-sixth octaves. The primary features of blade position are [anterior posterior]. The higher 3 one-sixth octaves are classified as [+anterior –posterior] dentalveolars, the lower 3 one-sixth octaves as [–anterior +posterior] postalveolars. When speech is processed by a bank of auditory filters, the spectra are too smoothed for the half-power frequencies to be of use in assessing bandwidth. Therefore an auditorily-based estimator of the relative power log Q was developed: the peak energy factor PE. The measured acoustic cue for blade aperture $\overline{A}(blade) / \overline{A}(lip)$ is then the peak energy factor PE3 instead of log Q3. Blade aperture is captured by the features [elevated depressed]. Laminals are [+elevated –depressed]. Apicals are [–elevated –depressed]. Similar to what would be expected for log Q3, the PE3 value is small for [+elevated –depressed] laminals with a narrow blade aperture whereas the PE3 value is larger for [–elevated –depressed] apicals with a wider blade aperture. The coronal fricatives of American English, Toda, and Ubykh were examined. The PE3 measures as well as palatographic data consistently show the laminality of [s̪ s] and the apicality of [ʃ ʂ]. Furthermore, the [s ʃ] sounds are [+anterior –posterior] dentalveolars. In the American English recordings, for example, there is no statistically significant difference in F3 frequency between laminal [s] and apical [ʃ], indicating that both sounds have very similar blade positions.

Table 3
Overview of the acoustic-articulatory associations. The acoustic change produced by a doubling of the articulatory parameter is expressed in octaves log F ($= \log_2 F$) or as relative power log Q ($= 20\log_{10} Q$).

Acoustic Parameter	Name of Articulatory Parameter	Articulatory Parameter	Acoustic Change per doubling of Articulatory Parameter	Section
log F1	Tongue Root Aperture	$\overline{A}(root) / \overline{A}(lip)$	~ –1/3 octave	9.2.1
log Q1	Nasal Port Area	$A(nasal)$	~ –3 dB	10.2
log F2	Tongue Body Position	$\overline{A}(forebody) / \overline{A}(hindbody)$	~ –1/6 octave	9.2.2
log Q2	Tongue Body Aperture	$\overline{A}(body) / \overline{A}(lip)$	~ +3 dB	9.4.1
log F3	Blade Position	$\overline{A}(foreblade) / \overline{A}(hindblade)$	~ –1/12 octave	9.2.3
log Q3	Blade Aperture	$\overline{A}(blade) / \overline{A}(lip)$	~ +3 dB	9.4.1
log F4	Lip Length	$L(lip)$	~ –1 octave	9.5
log Q4	Lip Area	$\overline{A}(lip)$	~ –3 dB	9.4.2

In the earlier research, the evidence for distinctive features was almost entirely empirical because it was based on the acoustic-articulatory correlations of the original 10-speaker data set. By contrast, each correlation pair in the summary of Section 9.6 is followed by a rule stating the theoretical proportionality between the acoustic and articulatory parameters. The proportionality is $F_i \propto (\lambda_i / 4)^{-1} (A_{U\,max} / A_{p\,max})^{a/b}$ for the F1–F3 formant frequencies and $F_4 \propto (\lambda_4 / 4)^{-1}$ for the F4 formant frequency. The proportionality is $Q_{rad}^{PE} \propto A_{\lambda/4} / A_0^2$ for the Q2–Q3 quality factors and $Q_{rad}^{KE} \propto 1 / A_{A_{\lambda/4}}$ for the Q4 quality factor. As the correlation results are now grounded theoretically, the acoustic-articulatory correspondences in Table 3 provide a firm foundation for distinctive features of the vocal tract.

Acknowledgements

I wish to thank the researchers who painstakingly measured the vocal tract area functions used in this study. Without their work the paper could not have been written. I would also like to thank Alessia Cherici and three anonymous reviewers for their careful readings, corrections, and helpful suggestions. I am also indebted to Kenneth de Jong for bringing a number of points to my attention during the course of the study.

Appendix A

Table A1
Acoustic parameters calculated by the FDVT and VTAR models for a 4 cm^2 uniform pipe.

Formant	*f* Hz	*Amp* dB	*B* Hz	B_{sum} Hz	B_{rad} Hz	B_{visc} Hz	B_{heat} Hz	B_{wall} 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^2uniform pipe								
F1	493	30.98	17					
F2	1396	24.73	34					
F3	2316	17.94	71					
F4	3240	13.21	115					

Table A2
Acoustic parameters calculated by the FDVT and VTAR models for a 1 cm^2 uniform pipe.

Formant	*f* Hz	*Amp* dB	*B* Hz	B_{sum} Hz	B_{rad} Hz	B_{visc} Hz	B_{heat} Hz	B_{wall} 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					

Appendix B

Table B1
Acoustic-articulatory correlations for the F1 formant variables and overall power PWR. The formant frequency log F1 and quality factor log Q1 are equal to $\log_2 F1$ and $20\log_{10} Q1$. The units of PWR are dB.

Data set	Rank	Acoustic Variable	Articulatory Variable	Correlation Mean *r*	s.d.
Original	1	log F1	$\log_2 \overline{A}(root) / \overline{A}(lip)$	–0.93515	.062
	2	log F1	$\log_2 \overline{A}(root)$	–0.91510	.071
	3	log F1	$\overline{A}(root)$	–0.90918	.069
	4	log F1	$\overline{A}(root) / \overline{A}(lip)$	–0.89934	.105
	5	log F1	$\log_2 \overline{A}(lip)$	0.72255	.224
Tiede	1	log F1	$\overline{A}(root)$	–0.82669	.146
	2	log F1	$\log_2 \overline{A}(root) / \overline{A}(lip)$	–0.82334	.108
	3	log F1	$\log_2 \overline{A}(root)$	–0.81884	.141
	4	log F1	$\overline{A}(root) / \overline{A}(lip)$	–0.81054	.107
	5	log F1	$\overline{A}(body) / \overline{A}(lip)$	–0.74512	.131
Original	1	log Q1	$\overline{A}(root) / \overline{A}(lip)$	–0.87012	.141
	2	log Q1	$\overline{A}(root)$	–0.85944	.100
	3	log Q1	$\log_2 \overline{A}(root) / \overline{A}(lip)$	–0.84201	.121
	4	log Q1	$\log_2 \overline{A}(root)$	–0.84109	.110
	5	log Q1	$\log_2 \overline{A}(blade)$	0.68678	.180
Tiede	1	log Q1	$\overline{A}(root) / \overline{A}(lip)$	–0.77020	.170
	2	log Q1	$\log_2 \overline{A}(root) / \overline{A}(lip)$	–0.72285	.183
	3	log Q1	$\overline{A}(body) / \overline{A}(lip)$	–0.71621	.181
	4	log Q1	$\overline{A}(root)$	–0.70190	.247
	5	log Q1	$\log_2 \overline{A}(blade)$	0.69239	.174
Original	1	PWR	$\log_2 \overline{A}(root) / \overline{A}(lip)$	–0.86199	.082
	2	PWR	$\overline{A}(root) / \overline{A}(lip)$	–0.82734	.116
	3	PWR	$\log_2 \overline{A}(root)$	–0.78968	.111
	4	PWR	$\overline{A}(root)$	–0.76902	.131
	5	PWR	$\log_2 \overline{A}(lip)$	0.71211	.197
Tiede	1	PWR	$L(lip)$	–0.70541	.215
	2	PWR	$\log_2 L(lip)$	–0.70254	.214
	3	PWR	$\overline{A}(root)$	–0.66383	.203
	4	PWR	$\log_2 \overline{A}(root)$	–0.65434	.194
	5	PWR	$\log_2 \overline{A}(root) / \overline{A}(lip)$	–0.50073	.285

Table B2
Acoustic-articulatory correlations for the F2 formant variables. The formant frequency log F2 and quality factor log Q2 are equal to $\log_2 F2$ and $20\log_{10} Q2$.

Data set	Rank	Acoustic Variable	Articulatory Variable	Correlation Mean *r*	s.d.
Original	1	log F2	$\log_2 \overline{A}(blade) / \overline{A}(lip)$	–0.90725	.073
	2	log F2	$\overline{A}(blade) / \overline{A}(lip)$	–0.84932	.074
	3	log F2	$\log_2 \overline{A}(blade)$	–0.82299	.096
	4	log F2	$\overline{A}(blade)$	–0.77827	.132
	5	log F2	$I_{\min A}(body)$	0.61705	.280
Tiede	1	log F2	$\log_2 \overline{A}(blade)$	–0.85812	.059
	2	log F2	$\overline{A}(blade)$	–0.81771	.111
	3	log F2	$\log_2 \overline{A}(body) / \overline{A}(lip)$	0.79211	.126
	4	log F2	$\log_2 \overline{A}(body)$	0.78784	.241
	5	log F2	$\overline{A}(body)$	0.75681	.281
Original	1	log Q2	$\overline{A}(body) / \overline{A}(lip)$	0.64380	.323
	2	log Q2	$\log_2 \overline{A}(body) / \overline{A}(lip)$	0.60638	.294
	3	log Q2	$\log_2 \overline{A}(lip)$	–0.57947	.235
	4	log Q2	$\overline{A}(lip)$	–0.53095	.268
	5	log Q2	$\overline{A}(root) / \overline{A}(lip)$	0.46169	.391
Tiede	1	log Q2	$\overline{A}(body) / \overline{A}(lip)$	0.83429	.144
	2	log Q2	$\log_2 \overline{A}(body) / \overline{A}(lip)$	0.78290	.110
	3	log Q2	$\log_2 \overline{A}(blade)$	–0.77933	.157
	4	log Q2	$\overline{A}(body)$	0.77345	.178
	5	log Q2	$\overline{A}(root) / \overline{A}(lip)$	0.72109	.257

Table B3
Acoustic-articulatory correlations for the F3 formant variables. The formant frequency log F3 and quality factor log Q3 are equal to $\log_2 F3$ and $20\log_{10} Q3$.

Data set	Rank	Acoustic Variable	Articulatory Variable	Correlation Mean *r*	s.d.
Original	1	log F3	$\log_2 \overline{A}(foreblade) / \overline{A}(hindblade)$	–0.57438	.263
	2	log F3	$\overline{A}(foreblade) / \overline{A}(hindblade)$	–0.48748	.302
	3	log F3	$\log_2 L(lip)$	–0.43955	.439
	4	log F3	$L(lip)$	–0.43898	.438
	5	log F3	$I_{\min A}(blade)$	0.40025	.313
Tiede	1	log F3	$L(lip)$	–0.56703	.240
	2	log F3	$\log_2 L(lip)$	–0.56409	.240
	3	log F3	$\log_2 \overline{A}(foreblade) / \overline{A}(hindblade)$	–0.55945	.309
	4	log F3	$\overline{A}(foreblade) / \overline{A}(hindblade)$	–0.55183	.285
	5	log F3	$\overline{A}(body)$	0.45881	.261
Original	1	log Q3	$\log_2 \overline{A}(blade) / \overline{A}(lip)$	0.61130	.231
	2	log Q3	$L(lip)$	0.58685	.505
	3	log Q3	$\log_2 L(lip)$	0.58669	.504
	4	log Q3	$\overline{A}(blade) / \overline{A}(lip)$	0.58657	.154
	5	log Q3	$\overline{A}(lip)$	–0.50486	.335
Tiede	1	log Q3	$\log_2 \overline{A}(blade) / \overline{A}(lip)$	0.70050	.296
	2	log Q3	$\overline{A}(blade) / \overline{A}(lip)$	0.63044	.286
	3	log Q3	$L(lip)$	0.58571	.208
	4	log Q3	$\log_2 L(lip)$	0.58303	.209
	5	log Q3	$\overline{A}(body)$	–0.51074	.372

Table B4
Acoustic-articulatory correlations for the F4 formant variables. The formant frequency log F4 and quality factor log Q4 are equal to $\log_2 F4$ and $20\log_{10} Q4$.

Data set	Rank	Acoustic Variable	Articulatory Variable	Correlation Mean *r*	s.d.
Original	1	log F4	$\log_2 L(lip)$	–0.39111	.291
	2	log F4	$\log_2 \overline{A}(body) / \overline{A}(lip)$	0.38915	.371
	3	log F4	$L(lip)$	–0.38838	.293
	4	log F4	$\overline{A}(body) / \overline{A}(lip)$	0.38557	.378
	5	log F4	$\overline{A}(lip)$	–0.34886	.391
Tiede	1	log F4	$L(lip)$	–0.69908	.279
	2	log F4	$\log_2 L(lip)$	–0.69761	.279
	3	log F4	$\overline{A}(body)$	0.51881	.291
	4	log F4	$\log_2 \overline{A}(body)$	0.49897	.302
	5	log F4	$\overline{A}(body) / \overline{A}(lip)$	0.48471	.219
Original	1	log Q4	$\overline{A}(lip)$	–0.62215	.283
	2	log Q4	$\log_2 \overline{A}(lip)$	–0.61936	.252
	3	log Q4	$\log_2 \overline{A}(root) / \overline{A}(lip)$	0.58372	.259
	4	log Q4	$\overline{A}(root) / \overline{A}(lip)$	0.55222	.206
	5	log Q4	$L(lip)$	0.53260	.483
Tiede	1	log Q4	$\log_2 L(lip)$	0.71564	.317
	2	log Q4	$L(lip)$	0.71411	.316
	3	log Q4	$I_{\min A}(blade)$	–0.44877	.237
	4	log Q4	$\overline{A}(blade) / \overline{A}(lip)$	0.39170	.420
	5	log Q4	$\log_2 \overline{A}(body)$	–0.35359	.369

Appendix C

The summary of Section 3.1.3 shows that when $A_{p\,max}$ is doubled or $A_{U\,max}$ is halved, the negative octave changes computed by FDVT models (F1: –0.318, F2: –0.114, F3: –0.076) are almost identical to the values determined theoretically by the perturbation analysis (F1: –0.333, F2: –0.111, F3: –0.074). On the other hand, when $A_{p\,max}$ is halved or $A_{U\,max}$ is doubled, the positive octave changes computed by the FDVT models (F1: +0.233, F2: +0.060, F3: +0.031) are less than the values found by the perturbation analysis (F1: +0.333, F2: +0.111, F3: +0.074), particularly for the F2 and F3 formants.

To account for the divergence of the positive octave changes, it is necessary to consider the shortening and lengthening of the quarter wavelengths as a function of the corresponding F1–F3 frequencies. The acoustic partition of the vocal tract into four articulator regions presupposes that the F1–F3 quarter wavelengths are equal to those of the open-closed uniform pipe. However the assumption is only approximately met since the quarter wavelength $\lambda_i / 4 = c / 4F_i$, where F_i is the formant frequency, *i* the formant number, and *c* the sound velocity (Section 3.1.1). Thus as the formant frequency F_i increases, its quarter wavelength $\lambda_i / 4$ is shortened, and vice versa.

In Table C1 the F1–F3 frequencies of the uniform pipe are presented: 486 Hz, 1458 Hz, 2430 Hz. They are taken from Table 2 and were computed using lossless FDVT models. Their quarter wavelengths are 18 cm, 6 cm, 3.6 cm according to the equation $\lambda_i / 4 = c / 4F_i$, where $c = 35000$ cm/s. These quarter wavelength calculated from the F1–F3 frequencies are numerically exact because (a) the uniform pipe is 18 cm long and (b) 1 quarter wavelength fits into the pipe at F_1, 3 quarter wavelengths at F_2, and 5 quarter wavelengths at F_3 (Section 3.1.1). The raised and lowered F1–F3 frequencies are likewise taken from Table 2. A raised FDVT frequency is the geometric mean of two formant frequencies: one computed when $A_{p\,max}$ is halved, the other when $A_{U\,max}$ is doubled. A lowered FDVT frequency is the geometric mean of two formant frequencies: one computed when $A_{p\,max}$ is doubled, the other when $A_{U\,max}$ is halved. The perturbation analysis of Section 3.1.3 indicates that the formant frequency F_i is proportional to $(\lambda_i / 4)^{-1}(A_{U\,max} / A_{p\,max})^{n/27}$, where *n* is respectively 9, 3, 2 for F_1, F_2, F_3. A raised theoretical frequency is the formant frequency of a uniform pipe multiplied by $(2^{+1})^{n/27}$, where either $A_{p\,max}$ is halved or $A_{U\,max}$ is doubled. A lowered theoretical frequency is the formant frequency of a uniform pipe multiplied by $(2^{-1})^{n/27}$, where either $A_{p\,max}$ is doubled or $A_{U\,max}$ is halved. The quarter wavelengths of all the raised and lowered frequencies are calculated as well as the associated quarter wavelength ratios. The quarter wavelength ratio is defined as a shortened (raised F_i) or lengthened (lowered F_i) quarter wavelength divided by the corresponding quarter wavelength of the 18 cm uniform pipe.

Table C1
Raised and lowered F1–F3 frequencies together with their respective quarter wavelengths and quarter wavelength ratios. The quarter wavelength $\lambda_i / 4 = c / 4F_i$, where F_i is the formant frequency and *i* the formant number. The sound velocity *c* is 35000 cm/s. The formant frequencies (uniform, raised, lowered), calculated by lossless FDVT models are taken from Table 2. The quarter wavelength ratio is a shortened (raised F_i) or lengthened (lowered F_i) quarter wavelength divided by the corresponding quarter wavelength of the 18 cm uniform pipe.

	F1 (Hz)	F1 quarter wavelength (cm)	F1 quarter wavelength ratio
uniform FDVT	486	18.00	
raised theoretical	$486 \times (2^{+1})^{9/27} = 612.32$	14.29	14.29/18.00 = 0.794
raised FDVT	571	15.32	15.32/18.00 = 0.851
lowered theoretical	$486 \times (2^{-1})^{9/27} = 385.74$	22.68	22.68/18.00 = 1.260
lowered FDVT	390	22.44	22.44/18.00 = 1.246

	F2 (Hz)	F2 quarter wavelength (cm)	F2 quarter wavelength ratio
uniform FDVT	1458	6.00	
raised theoretical	$1458 \times (2^{+1})^{3/27} = 1574.73$	5.56	5.56/6.00 = 0.926
raised FDVT	1520	5.76	5.76/6.00 = 0.959
lowered theoretical	$1458 \times (2^{-1})^{3/27} = 1349.93$	6.48	6.48/6.00 = 1.080
lowered FDVT	1347	6.50	6.50/6.00 = 1.083

	F3 (Hz)	F3 quarter wavelength (cm)	F3 quarter wavelength ratio
uniform FDVT	2430	3.60	
raised theoretical	$2430 \times (2^{+1})^{2/27} = 2558.03$	3.42	3.42/3.60 = 0.950
raised FDVT	2482	3.53	3.53/3.60 = 0.979
lowered theoretical	$2430 \times (2^{-1})^{2/27} = 2308.38$	3.79	3.79/3.60 = 1.053
lowered FDVT	2306	3.79	3.79/3.60 = 1.054

As anticipated from the negative octave changes, an examination of Table C1 reveals that the lowered FDVT frequencies and the lowered theoretical frequencies are nearly the same for formants F1–F3. Consequently, their quarter wavelengths and quarter wavelength ratios are also very similar.

By definition, raised frequencies exhibit quarter wavelength ratios less than 1. The raised FDVT frequencies therefore shorten the quarter wavelengths relative to the corresponding quarter wavelengths of the uniform pipe. Nevertheless, the acoustic partition of the vocal tract into constant articulator regions remains based on the quarter wavelengths of the uniform pipe. Hence for a raised FDVT frequency, a change in area $A_{U\,max}$ or $A_{p\,max}$ extends over a longer region than the one delimited by the shortened quarter wavelength. As the longer region has longer inductance and capacitance lengths l_L and l_C, the acoustic mass ($M = Ll_L$) increases and the acoustic stiffness ($s = 1/Cl_C$) decreases, thereby depressing the raised FDVT frequency (cf. the Helmholtz resonator

$f_0 = (2\pi)^{-1}\sqrt{s/M}$ in Section 3.1.1). Because the raised FDVT frequencies are depressed, they display larger quarter wavelengths ratios than the raised theoretical frequencies in Table C1. Remark however that for the F2 and F3 formants, quarter wavelength differences between the raised FDVT frequencies and the raised theoretical frequencies are small (F2: $5.76 - 5.56 = 0.20$ cm; F3: $3.53 - 3.42 = 0.11$ cm). It was noted earlier that the vocal tract contains three F2 quarter wavelengths and five F3 quarter wavelengths. The F2 and F3 quarter wavelength differences can then be multiplied by 3 and 5 to obtain the vocal tract length differences, which are also comparatively small ($0.20 \times 3 = 0.60$ cm; $0.11 \times 5 = 0.55$ cm). If the vocal tract is shortened by the length difference of 5 mm ($\approx 0.55 - 0.60$ cm), the inductance and capacitance lengths will be shortened by the same distance that they were lengthened. Consequently, the decrease in acoustic mass and the increase in acoustic stiffness should elevate the raised FDVT frequency to the same value as the raised theoretical frequency (see Section 9.5 for vocal tract shortening and lengthening).

To test the prediction, the raised F2 and F3 frequencies of the 18 cm vocal tract presented in Table C1 are contrasted with those of a 17.5 cm vocal tract computed by lossless FDVT models. Table C1 shows that the raised theoretical F2 frequency is 1575 Hz, but the raised FDVT F2 frequency is 1520 Hz; the raised theoretical F3 frequency is 2558 Hz, but the raised FDVT F3 frequency is only 2482 Hz. Thus for the 18 cm vocal tract, the calculated FDVT formant frequencies are well below the theoretical ones. For the 17.5 cm vocal tract, on the other hand, the raised FDVT F2 frequency is 1564 Hz and therefore close to the theoretical value of 1575 Hz; the raised FDVT F3 frequency is 2554 Hz and even closer to the theoretical value of 2558 Hz. As in Table 2, the raised FDVT F2 frequency 1564 Hz is the geometric mean of the 1 cm^2 forebody frequency 1543 Hz and the 4 cm^2 hindbody frequency 1586 Hz; the raised FDVT F3 frequency 2554 Hz is the geometric mean of the 1 cm^2 foreblade frequency 2583 Hz and the 4 cm^2 hindblade frequency 2525 Hz.

The results demonstrate that a minor shortening of the vocal tract can easily offset the decline of the raised F2–F3 formant frequencies caused by longer inductance and capacitance lengths. In this case, the positive octave changes of both the FDVT frequency and the theoretical frequency will be nearly the same—as is already observed for the negative octave changes.

References

Aarts, R. M., & Janssen, A. J. E. M. (2003). Approximation of the Struve function H_1 occurring in impedance calculations. *The Journal of the Acoustical Society of America*, *113*, 2635–2637.

Abry, C., & Boë, L.-J. (1986). "Laws" for lips. *Speech Communication*, *5*, 97–104.

Badin, P. (1989). Acoustics of voiceless fricatives: Production theory and data. *Speech Transmission Laboratory Quarterly Progress and Status Report*, *30*, 33–55. Stockholm: Royal Institute of Technology.

Badin, P., Perrier, P., Boë, L.-J., & Abry, C. (1990). Vocalic nomograms: Acoustic and articulatory considerations upon formant convergences. *The Journal of the Acoustical Society of America*, *87*, 1290–1300.

Baer, T., Gore, J. C., Gracco, L. C., & Nye, P. W. (1991). Analysis of vocal tract shape and dimensions using magnetic resonance imaging: Vowels. *The Journal of the Acoustical Society of America*, *90*, 799–828.

Boë, L.-.J., & Perrier, P. (1990). Comments on "Distinctive regions and modes: A new theory of speech production" by M. Mrayati, R. Carré and B. Guérin. *Speech Communication*, *9*, 217–230.

Boylestad, R. L. (2016). *Introductory circuit analysis*. Essex: Pearson Education.

Carré, R., Divenyi, P., & Mrayati, M. (2017). *Speech: A dynamic process*. Berlin: Walter de Gruyter.

Catford, J. C. (1981). Observations on the recent history of vowel classification. In R. E. Asher, & E. J. A. Henderson (Eds.), *Towards a history of phonetics* (pp. 19–32). Edinburgh: University Press.

Chen, M. Y. 1997. Acoustic correlates of English and French nasalized vowels. *The Journal of the Acoustical Society of America*, *102*, 2360–2370.

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

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

Clements, G. N., & Hume, E. V. (1995). The internal organization of speech sounds. In J. A. Goldsmith (Ed.), *The handbook of phonological theory* (pp. 245–306). Cambridge: Blackwell.

Dang, J., Honda, K., & Suzuki, H. (1994). Morphological and acoustical analysis of the nasal and the paranasal cavities. *The Journal of the Acoustical Society of America*, *96*, 2088–2100.

Delattre, P., Liberman, A. M., Cooper, F. S., & Gerstman, L. J. (1952). 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.

Dunn, H. K. (1950). The calculation of vowel resonances, and an electrical vocal tract. *The Journal of the Acoustical Society of America*, *22*, 740–753.

Ewan, W. G., & Krones, R. (1974). Measuring larynx movement using the thyroumbrometer. *Journal of Phonetics*, *2*, 327–335.

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

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

Fant, G. (1975). Vocal-tract area and length perturbations. *Speech Transmission Laboratory Quarterly Progress and Status Report*, *16*, 1–14. Stockholm: Royal Institute of Technology.

Fant, G. (1986). Features: Fiction and facts. In J. S. Perkell, & D. H. Klatt (Eds.), *Invariance and variability in speech processes* (pp. 480–492). Hillsdale, New Jersey: Lawrence Erlbaum.

Fant, G., & Pauli, S. (1975). Spatial characteristics of vocal tract resonance modes. In *Proceedings of the Speech Communication Seminar*, Stockholm, August 1–3, 1974 (pp. 121–132).

Fitch, W. T., & Giedd, J. (1999). Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *The Journal of the Acoustical Society of America*, *106*, 1511–1522.

Flanagan, J. L. (1957). Difference limen for formant amplitude. *Journal of Speech and Hearing Disorders*, *22*, 205–212.

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

Flemming, E. S. (2002). *Auditory representations in phonology*. New York: Routledge.

Goldstein, U. G. (1980). *An articulatory model for the vocal tracts of growing children* (Ph.D. dissertation), MIT.

Green, E. I. (1955). The story of Q. *American Scientist*, *43*, 584–594.

Gussenhoven, C., & Jacobs, H. (2017). *Understanding Phonology*. New York: Routledge.

Halle, M., & Stevens, K. N. (1969). On the feature "advanced tongue root." *MIT Research Laboratory of Electronics Quarterly Progress Report*, *94*, 209–215.

Halle, M., & Stevens, K. N. (1991). Knowledge of language and the sounds of speech. In J. Sundberg, L. Nord, & R. Carlson (Eds.), *Music, language, speech and brain* (pp. 1–19). London: Macmillan Press.

Halle, M., Vaux, B., & Wolfe, A. (2000). On feature spreading and the representation of place of articulation. *Linguistic Inquiry*, 31, 387–444.

Handbook of the International Phonetic Association (1999). New York: Cambridge University Press.

Hawks, J. W., & Miller, J. D. (1995). A formant bandwidth estimation procedure for vowel synthesis. *The Journal of the Acoustical Society of America*, *97*, 1343–1344.

Hoole, P., & Kroos, C. (1998). Control of larynx height in vowel production. *Fifth International Conference on Spoken Language Processing (ICSLP 98)*, Sydney (pp. 531–534).

House, A. S., & Stevens, K. N. (1956). Analog studies of the nasalization of vowels. *Journal of Speech and Hearing Disorders*, *21*, 218–232.

Ishizaka, K., French, J. C., & Flanagan, J. L. (1975). Direct determination of vocal tract wall impedance. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, *23*, 370–373.

Jakobson, R., Fant, G., & Halle, M. (1952). *Preliminaries to speech analysis*. *The distinctive features and their correlates*. Cambridge: MIT Press.

Jassem, W. (1965). The formants of fricative consonants. *Language and Speech*, *8*, 1–16.

Joos, M. (1948). *Acoustic phonetics*. Language Monograph No. 23. Linguistic Society of America.

Jones, D. (1922). *An outline of English phonetics*. New York: G. E. Stechert & Co.

Keating, P. (1991). Coronal places of articulation. In C. Paradis, & J.-F. Prunet (Eds.), *The special status of coronals: Internal and external evidence* (pp. 29–48). San Diego: Academic Press.

Kent, R. D., & Vorperian, H. K. (2018). Static measurements of vowel formant frequencies and bandwidths: A review. *Journal of Communication Disorders*, *74*, 74–97.

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

Lindau, M. (1979). The feature expanded. *Journal of Phonetics*, *7*, 163–176.

McGowan, R. S. (2018). *Acoustics of speech production*. Cambridge: CReSS Books.

Miller, J. D. (1989). Auditory-perceptual interpretation of the vowel. *The Journal of the Acoustical Society of America*, *85*, 2114–2134.

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

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

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

Pennington, M. (2011). Toward phonetically grounded distinctive features. Part I: Acoustic-articulatory correlations in a four-region model of the vocal tract. *IULC Working Papers*, *11*, 1–27.

Pennington, M. (2017). Acoustic-articulatory correlations in a four-region model of the vocal tract: Experimental evidence for blade features. *IULC Working Papers*, *17*, 1–42.

Perkell, J. S. (1971). Physiology of speech production: A preliminary study of two suggested revisions of the features specifying vowels. *MIT Research Laboratory of Electronics Quarterly Progress Report, 102*, 123–139.

Pruthi, T. (2007). *Analysis, vocal-tract modeling and automatic detection of vowel nasalization* (Ph.D. dissertation), University of Maryland, College Park.

Ratner, B. (2017). *Statistical and machine-learning data mining. Techniques for better predictive modeling and analysis of big data.* Boca Raton, Florida: CRC Press Taylor & Francis Group.

Riordan, C. J. (1977). Control of vocal-tract length in speech. *The Journal of the Acoustical Society of America*, *62*, 998–1002.

Rosenblum, L. D. (2008). Speech perception as a multimodal phenomenon. *Current Directions in Psychological Science*, *17*, 405–409.

Rossi, M. (1971). L'intensité spécifique des voyelles. *Phonetica*, *24*, 129–161.

Rossing, T. D., & Fletcher, N. H. (2004). *Principles of Vibration and Sound*. New York: Springer Verlag.

Russell, G. O. (1928). *The vowel. Its physiological mechanism as shown by X-ray*. Columbus: Ohio State University Press.

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

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

Story, B. H., Titze, I. R., & Hoffman, E. A. (1996). Vocal tract area functions from magnetic resonance imaging. *The Journal of the Acoustical Society of America*, *100*, 537–554.

Story, B. H., Titze, I. R., & Hoffman, E. A. (1998). Vocal tract area functions for an adult female speaker based on volumetric imaging. *The Journal of the Acoustical Society of America*, *104*, 471–487.

Takemoto, H., Honda, K., Masaki, S., Shimada, Y., & Fujimoto, I. (2006). Measurement of temporal changes in vocal tract area function from 3D cine-MRI data. *The Journal of the Acoustical Society of America*, *119*, 1037–1049.

Tiede, M. K. (2013). *An MRI-based morphological approach to vocal tract area function estimation* (Ph.D. dissertation), Yale University.

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

Turner, R. E., Walters, T. C., Monaghan, J. J. M., & Patterson, R. D. (2009). A statistical, formant-pattern model for segregating vowel type and vocal-tract length in developmental formant data. *The Journal of the Acoustical Society of America*, *125*, 2374–2386.

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

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

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

Yang, C.-S., & Kasuya, H. (1994). Accurate measurement of vocal tract shapes from magnetic resonance images of child, female, and male subjects. *Third International Conference on Spoken Language Processing (ICSLP 94)*, Yokohama (pp. 623–626).

Zhang, Z., & Espy-Wilson, C. Y. (2004). A vocal-tract model of American English /l/. *The Journal of the Acoustical Society of America*, *115*, 1274–1280.

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