Verifying a vocal tract model with a closed side-branch

Michel Tah-Tung Jackson,[a] Carol Espy-Wilson, and Suzanne E. Boyce[b]
Department of Electrical and Computer Engineering, Boston University, 8 St. Mary's Street, Boston, Massachusetts 02215

(Received 27 January 1999; revised 3 March 2001; accepted 13 March 2001)

In this article an implementation of a vocal tract model and its validation are described. The model uses a transmission line model to calculate pole and zero frequencies for a vocal tract with a closed side-branch such as a sublingual cavity. In the validation study calculated pole and zero frequencies from the model are compared with frequencies estimated using elementary acoustic formulas for a variety of vocal tract configurations. © *2001 Acoustical Society of America.*
[DOI: 10.1121/1.1370526]

PACS numbers: 43.70.Bk, 43.70.Fq [AL]

I. INTRODUCTION

Our purpose in this article is to describe a vocal tract model using a branching transmission line to model a sublingual cavity, and its validation. As research on the acoustics of speech production proceeds, models necessarily become more complex and detailed; the verification of these models also becomes more difficult. Thorough validation of the models is required before model-based inferences (e.g., ''analysis by synthesis'') can be accepted.

The construction of the model was motivated by a desire to test the hypothesis that a sublingual cavity plays a crucial role in determining the acoustics of various apical, retroflex, and rhotic speech sounds. This hypothesis has been advanced separately in the literature for /s/ (Perkell, Boyce, and Stevens, 1979), for retroflex and sublingual stops (Ladefoged and Bhaskararao, 1983), and retroflex American English /r/ (Stevens, 1999) and is addressed further in Espy-Wilson, Boyce, Jackson, Narayanan, and Alwan (2000). As in Espy-Wilson *et al.* (2000), we assume that the sublingual cavity behaves acoustically like a ''side-branch'' off the vocal tract, at least for relatively low frequencies.

II. DESCRIPTION OF THE MODEL

The branching transmission-line model is implemented as a MATLAB-callable routine called AFSB2XFB. It is based on the VTCALCS program described by Maeda (1982), as modified for use with MATLAB numerical analysis software by Dr. Ronan Scaife at Dublin City University. It is intended to calculate the acoustic behavior of a vocal tract that includes a side-branch such as a sublingual cavity. The AFSB2XFB routine, given section lengths and area functions of the ''main'' glottis-to-lips portion of the vocal tract and a side-branch, calculates the vocal tract transfer function. The transfer function may then be searched, using numerical peak-picking routines, to determine peak (pole) frequencies. The acoustic behavior of the ''main'' tract and the side-branch is approximated using the transmission-line analog sections specified in Maeda (1982) and Espy-Wilson *et al.* (2000), Fig. 6. Each section includes terms for viscous losses, heat-conduction losses, and vocal tract wall losses, as specified in Maeda (1982). For the purposes of this Letter, the glottal end of the vocal tract was modeled with a closed termination, and an R–L circuit model was used to approximate the effect of radiation at the lips.

III. VERIFICATION PROCEDURE

In order to verify the performance of the AFSB2XFB implementation of this model, the pole and zero frequencies of two groups of tube configurations were estimated using simple tube models. The first group of configurations tests the model when the side-branch is strongly coupled to the rest of the vocal tract because the vocal tract has no constriction. These configurations are similar to some configurations observed in r-colored vowels in American English [see, for example, the cineradiographic data in Lindau (1985), especially speaker P5, Fig. 11.5]. The second group of configurations tests the model when the vocal tract has a constriction that decouples the side-branch from the back cavity of the vocal tract. These configurations are similar to the configurations observed in some /r/'s in American English [for example, Lindau (1985), especially speaker P4, Fig. 11.5; and Espy-Wilson *et al.* (2000), Fig. 3]. Unlike VTCALCS, these estimates neglect radiation effects and the effects of vibrating vocal tract walls.

Then, the model was used to calculate the frequency response of the same tube configurations, and peak and zero frequencies were determined from this frequency response. The model's *calculated* peak and zero frequencies are compared with the *estimated* pole and zero frequencies from simple tubes. The results from the model are also compared with the results from VTCALCS, since the model results from side-branches tending toward zero length and zero cross-sectional area should converge to the same results as VTCALCS.

IV. RESULTS FOR VARIOUS TUBE CONFIGURATIONS

A. Side-branch strongly coupled to main tract

This configuration is schematized in Fig. 1(a). A uniform tube with the ''glottal'' end closed and the ''labial'' end open has a side-branch (with a closed end) attached. The position, length, and cross-sectional area of the side-branch vary.

[a]Current address: 86 Spring Ave., Arlington, MA 02476. Electronic mail: ladmttj@ix.netcom.com
[b]Currently at Department of Communication Sciences and Disorders, Mail Location 379, University of Cincinnati, Cincinnati, OH 45221.

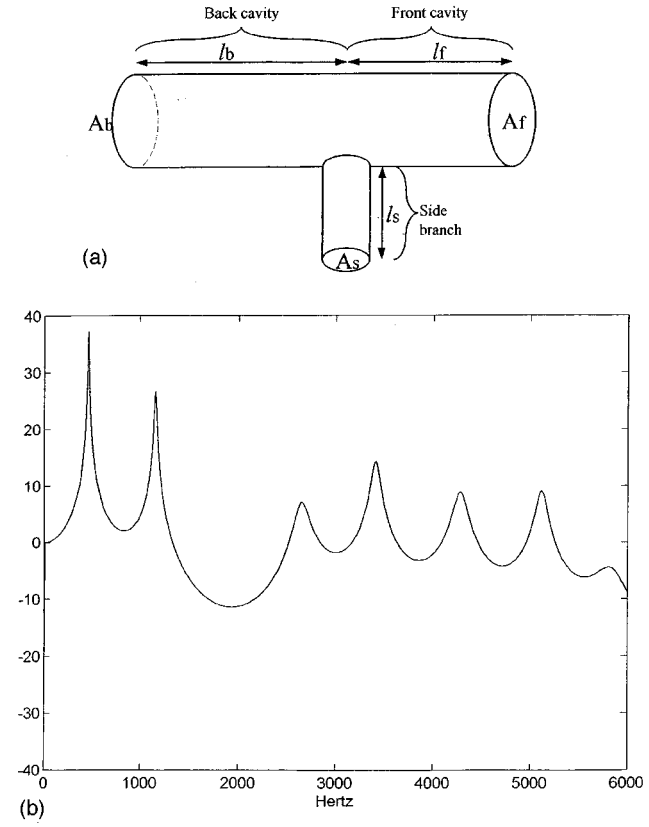

FIG. 1. (a) The tube configuration simulated in the ''strongly coupled'' models. (b) A sample plot of the calculated transfer function.

The acoustic impedance of ideal, hard-walled tubes in this configuration, as seen from the branching point, is given by Eq. (1), in which wall and other losses are neglected. The resonances of the glottis-to-lips transfer function are given by the zeros of Eq. (1),

$$Z(\omega)=A_b\tan(\omega l_b/c)+A_s\tan(\omega l_s/c)-A_f\cot(\omega l_f/c). \quad (1)$$

[Here, $Z(\omega)$ is the vocal tract impedance at the angular frequency ω, A_b, A_s, and A_f are the cross-sectional areas of the back tube, side-branch tube, and front tube, respectively; l_b, l_s, and l_f are the lengths of the same; and c is the speed of sound. In this Letter, c will be taken as 35 000 cm/s for the warm, humid air in the vocal tract.]

The side-branch produces zeros in the glottis-to-lip transfer function at the frequencies at which the acoustic impedance of the side-branch goes to zero, i.e., at the frequencies that would be the resonant frequencies of the isolated side-branch tube. This series of zero frequencies ζ_n is given by the well-known formula (2),

$$\zeta_n=(2n-1)c/4l_s. \quad (2)$$

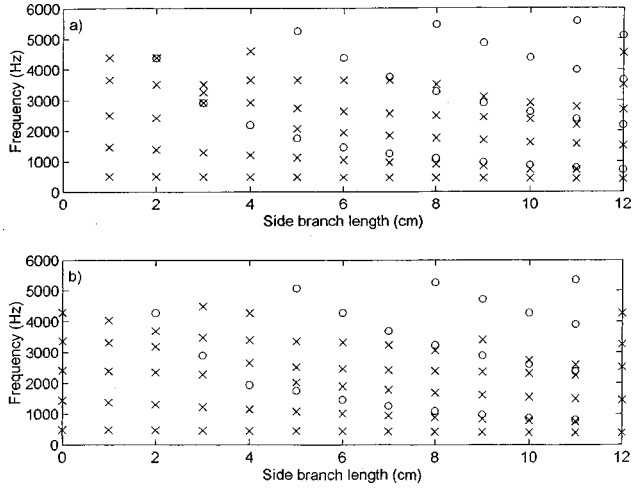

FIG. 2. (a) Estimated pole and zero frequencies of the tube configuration in Fig. 1 as l_s varies from 1 to 12 cm. In this and subsequent figures, ''×'' represents a pole frequency and ''○'' represents a zero frequency. (b) Calculated pole and zero frequencies.

For example, given a 17-cm vocal tract with a uniform cross-sectional area $A_b=A_s=A_f=4\ \mathrm{cm}^2$, with a back cavity length $l_b=12$, a side-branch length $l_s=4$, and a front cavity length $l_f=5$, the zeros of Eq. (1) can be numerically evaluated. The estimated resonant frequencies are ≈495, 1205, 2925, 3655, and 4605 Hz. An estimated zero frequency of 2190 Hz is found from Eq. (2). Figure 1(b) shows a plot of the glottis-to-lips transfer function calculated by the model for this configuration. It can be seen that there are peaks near 450, 1150, 2650, 3400, and 4300 Hz. In addition, there is a possible zero at 2000 Hz.

Table I summarizes the strongly coupled configurations tested. The results are presented as nomograms in Figs. 2–4, and summarized in Tables III and IV. The first two rows of Table I summarize configurations in which the side-branch length varied from 1 to 12 cm (with the minor variation of using 0.5-cm sections for the 1-cm side-branch and 1.0-cm sections for the rest). The third row of Table I summarizes configurations in which the cross-sectional area of the side-branch varied from 0.5 to 4 cm^2. The fourth row of Table I summarizes configurations in which the location of the side-

TABLE I. ''Strongly coupled'' model configurations tested. ''Fig.'' is the figure in which the results are presented. n_b, n_s, n_f: the number of tube sections (circuit elements) used to model the back, side, and front cavities, respectively. x_b, x_s, x_f: length in cm of tube sections used to model the back, side, and front cavities. Other symbols are as in the text.

	Back cavity				Side-branch				Front cavity			
Fig.	l_b	A_b	n_b	x_b	l_s	A_s	n_s	x_s	l_f	A_f	n_f	x_f
2	12.0	4.0	12	1.0	1.0	4.0	2	0.5	5.0	4.0	5	1.0
2	12.0	4.0	12	1.0	2.0→12.0	4.0	2→12	1.0	5.0	4.0	5	1.0
3	12.0	4.0	12	1.0	6.0	0.5→4.0	6	1.0	5.0	4.0	5	1.0
4	2.0→15.0	4.0	2→15	1.0	6.0	4.0	6	1.0	15.0→2.0	4.0	15→2	1.0

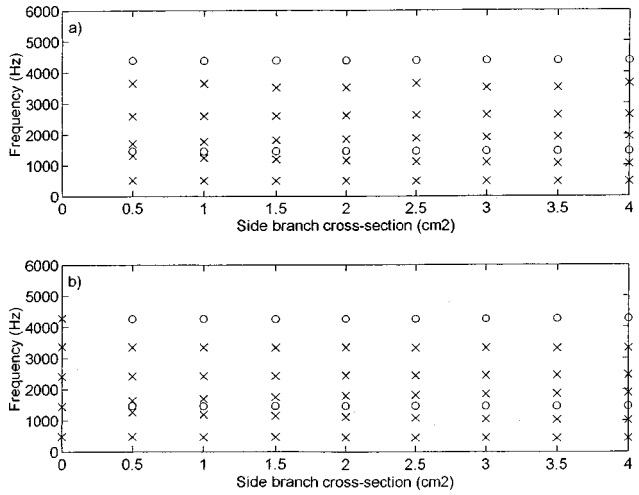

FIG. 3. (a) Estimated pole and zero frequencies of the tube configuration in Fig. 1 as A_s varies from 0.5 to 4 cm^2. (b) Calculated pole and zero frequencies.

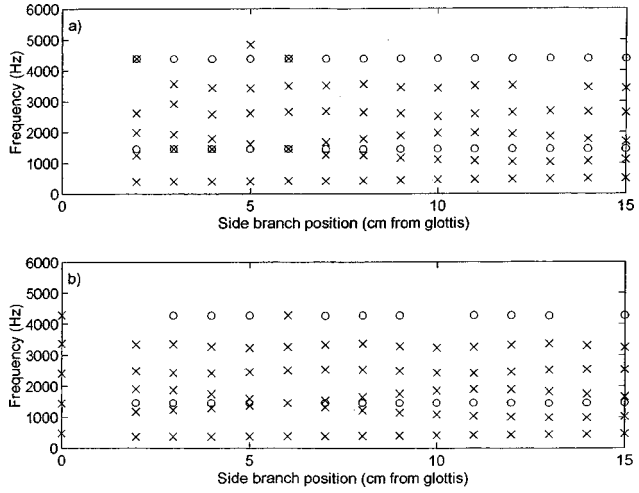

FIG. 4. (a) Estimated pole and zero frequencies of the tube configuration in Fig. 1 as l_b varies from 2 to 15 cm while l_b+l_f (the overall tract length) is held constant at 17 cm. (b) Calculated pole and zero frequencies.

branch (its distance from the glottal end of the model) varied from 2 to 15 cm.

Figure 2(a) shows a nomogram of the pole and zero frequencies beneath 5.5 kHz estimated from Eqs. (1) and (2) as the length of the side-branch varies. Figure 2(b) shows the (first five) pole and zero frequencies measured from the calculated transfer function, together with the limiting pole frequencies for a tube with no side-branch (i.e., the column of pole frequencies plotted on the y axis, at $l_s=0$ cm), as calculated by VTCALCS.

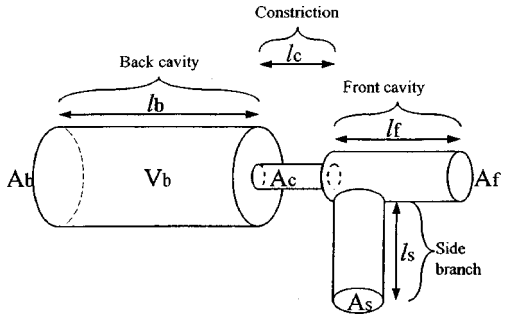

FIG. 5. The tube configuration simulated in the ''weakly coupled'' models.

Qualitative features of the nomogram, such as the abrupt decrease in the number of poles in the transfer function of the model vocal tract at $l_b=l_s=12$ cm, are seen in both Figs. 2(a) (estimated) and (b) (determined from model transfer functions.) In addition, the pole frequencies calculated by the model as the side-branch length tends to zero clearly converge to the pole frequencies for a tube with no side branch.

Figure 3 shows nomograms of the estimated and calculated (model) pole and zero frequencies as a function of the cross-sectional area of the side-branch together with the limiting pole frequencies for a tube with no side-branch. Figure 4 shows nomograms of the estimated and calculated pole and zero frequencies as a function of the position of the side-branch together with the limiting pole frequencies.

B. Side-branch weakly coupled to back cavity

This configuration is schematized in Fig. 5. In these configurations, a back cavity opens, through a constriction of length l_c and cross-sectional area A_c, into a front cavity with a side branch. Configurations in which l_s and A_s varied were tested. In addition, configurations in which l_s varied inversely with l_f were tested.

A lumped vocal tract resonance exists in these configurations because the back cavity is separated from the front cavity of the vocal tract by a constriction. This resonant frequency is approximately the Helmholtz, or cavity resonance frequency F_{H} given by Eq. (3) (again, the radiation impedance and vocal tract losses are neglected),

$$F_{\mathrm{H}}=(c/2\pi)\sqrt{(A_c/(V_b l_c))}. \tag{3}$$

[In (4), A_c is the cross-sectional area of the constriction; $V_b=A_b*l_b$ is the back cavity volume, and l_c is the constriction length.]

The combination of the side-branch and front cavity gives rise to a series of quarter-wave resonant frequencies F_n given by Eq. (4); see Espy-Wilson *et al.* (2000), p. 345 ff,

TABLE II. ''Weakly coupled'' model configurations tested. n_c: number of tube sections (circuit elements) used to model the constriction. x_c: length in cm of tube sections used to model the constriction. Other symbols are as in the previous table.

	Back cavity				Constriction				Side-branch				Front cavity			
Fig.	l_b	A_b	n_b	x_b	l_c	A_c	n_c	x_c	l_s	A_s	n_s	x_s	l_f	A_f	n_f	x_f
6	12.0	4.0	12	1.0	1.0	0.2	1	1.0	0.75→6.0	2.0	2→16	3/8	4.0	2.0	4	1.0
7	12.0	4.0	24	0.5	1.0	0.2	2	0.5	2.5	0.25→2.0	5	0.5	4.0	2.0	8	0.5
8	12.0	4.0	24	0.5	1.0	0.2	2	0.5	1.0→6.0	2.0	2→12	0.5	7.0→2.0	2.0	14→4	0.5

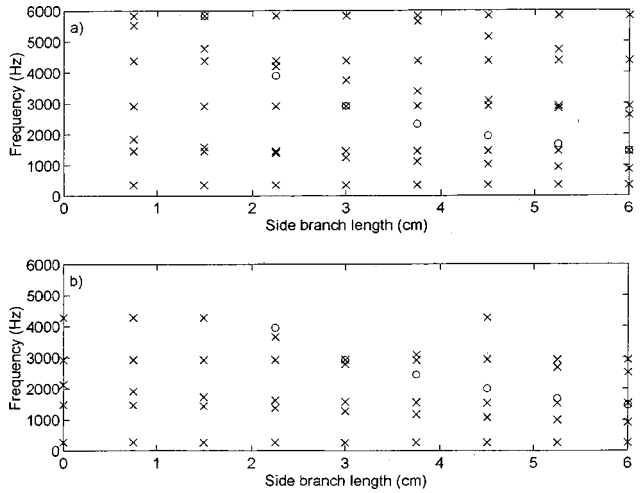

FIG. 6. (a) Estimated pole and zero frequencies of the tube configuration in Fig. 5 as l_s varies from 0.75 to 6 cm. (b) Calculated pole and zero frequencies.

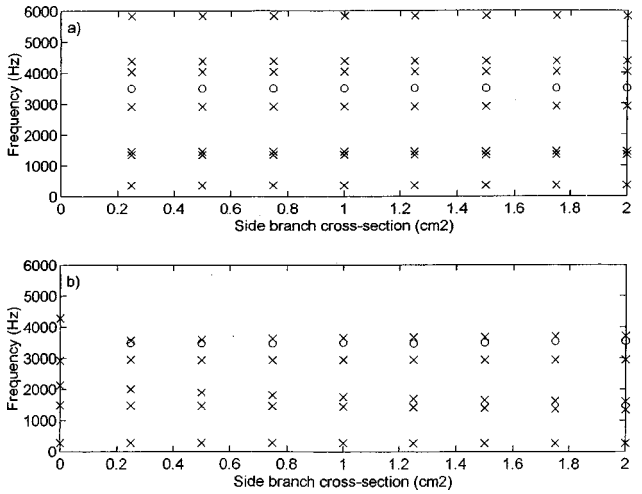

FIG. 7. (a) Estimated pole and zero frequencies of the tube configuration in Fig. 5 as A_s varies from 0.25 to 2 cm. (b) Calculated pole and zero frequencies.

$$F_n=(2n-1)c/4(l_s+l_f). \tag{4}$$

The back cavity contributes a half-wave resonance with the series of resonant frequencies given in (5):

$$F_n=nc/2l_s. \tag{5}$$

Finally, the side-branch causes zeros in the transfer function at the frequencies given in (2).

Table II summarizes the weakly coupled configurations tested. The results are presented as nomograms in Figs. 6–8, and summarized in Tables III and IV. The first row of Table II summarizes configurations in which the side-branch length l_s varied from 0.75 to 6 cm. The second row summarizes configurations in which the side-branch cross-sectional area A_s varied from 0.25 to 2.0 cm^2. The third row summarizes configurations in which the side-branch length l_s varied from 1 to 6 cm while the front cavity length l_f varied inversely from 7 to 2 cm (thus keeping the total length l_s+l_f constant at 8 cm).

Figure 6 shows nomograms of the estimated and calculated (model) pole and zero frequencies as the length of the side-branch varies, together with the limiting pole frequencies for a tube with no side-branch, as calculated by VTCALCS. Figure 7 shows the estimated and calculated pole and zero frequencies as the cross-sectional area of the side-branch varies, together with the limiting pole frequencies. Figure 8 shows the estimated and calculated pole and zero frequencies as the lengths of the side-branch and front cavity vary inversely, together with the limiting pole frequencies.

Table III reports the rms differences between the estimated and model pole frequencies in Hz and in percent (of the average pole or zero frequency) for each series of configurations. Table IV reports the rms differences between the estimated and model zero frequencies.

V. DISCUSSION AND CONCLUSIONS

The patterns of pole and zero frequencies calculated by the model are in good qualitative agreement with the patterns of pole and zero frequencies estimated from simple tube models. Many details are seen in both the estimated and the calculated nomograms. For example, Figs. 2(a) and (b) both show an exact pole-zero cancellation at $l_s=3$ cm; in Fig. 2(b), F_5 at $l_s=3$ cm appears to be 1 kHz greater than F_5 at $l_s=2$ cm because a peak ''expected'' for F_4 at about 3 kHz has been cancelled. Finally, the calculated pole frequencies when the side-branch length or cross-sectional area are close to zero do in fact converge to the pole frequencies calculated by VTCALCS for a tube with no side-branch.

In many cases, the pole frequencies calculated by VTCALCS and this model implementation (AFSB2XFB) are lower than the estimated values. The model pole frequencies are from 5%–10% lower than the estimated ones, with the smallest discrepancies generally in F_2. The discrepancy is due to the fact that VTCALCS and the side-branch model take into account viscous, heat conduction, and wall losses and radiation loads that the simple tube models neglect. In addition, it is possible that (3), which estimates the lumped

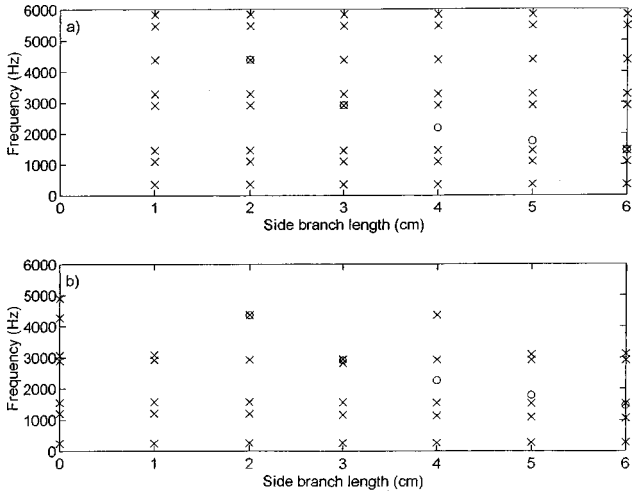

FIG. 8. (a) Estimated pole and zero frequencies of the tube configuration in Fig. 5 as l_s varies from 2 to 7 cm while l_s+l_f (the ''overall front cavity length'') is held constant at 8 cm. (b) Calculated pole and zero frequencies.

TABLE III. Root-mean-square differences between estimated and model pole frequencies in Hz, and in % of the mean pole frequency, for each series of configurations.

Varying parameter	F_1 [Hz (%)]	F_2 [Hz (%)]	F_3 [Hz (%)]	F_4 [Hz (%)]	F_5 [Hz (%)]
			''Strongly coupled''		
l_s	37.2(7.8)	51.5(4.7)	216.9(10.0)	207.8(7.1)	451.7(12.1)
A_s	37.3(7.5)	37.5(3.3)	52.3(2.8)	175.9(6.7)	253.8(7.1)
l_b, l_f	34.4(7.6)	108.5(8.7)	287.3(14.6)	321.1(11.6)	568.1(15.5)
			''Weakly coupled''		
l_s	84.3(23.4)	37.9(3.2)	113.0(7.4)	79.1(2.8)	559.5(15.4)
A_s	83.3(23.2)	95.2(7.1)	328.0(22.5)	33.7(1.2)	386.9(9.6)
l_s, l_f	88.6(24.6)	76.7(7.0)	104.3(7.2)	38.6(1.3)	654.7(20.0)

vocal tract resonant frequency from the back cavity volume and the dimensions of the constriction, unduly neglects the load due to the front cavity and side-branch.

In summary, these results demonstrate that the pole and zero frequencies produced by the side-branch model, as implemented by the AFSB2XFB routine, are plausible for /r/-like consonants produced by forming an acoustic side-branch in the vocal tract.

TABLE IV. Root-mean-square differences between estimated and model zero frequencies in Hz, and in % the mean zero frequency, for each series of configurations.

Varying parameter	ζ_1 [Hz (%)]	ζ_2 [Hz (%)]	ζ_3 [Hz (%)]	ζ_4 [Hz (%)]
		''Strongly coupled''		
l_s	85.4(4.8)	83.3(2.4)	140.8(3.0)	220.0(4)
A_s	4.1(0.3)	105.7(2.4)		
l_b, l_f	5.6(0.4)	103.7(2.4)		
		''Weakly coupled''		
l_s	85.1(3.0)			
A_s	21.4(0.6)			
l_s, l_f	41.3(1.6)			

ACKNOWLEDGMENTS

This work was supported primarily by Grant No. NIH IR03-C2576-01 to Dr. Suzanne E. Boyce. Further support was provided by National Science Foundation Grant No. IRI-9310518 and NIH Grant No. 1 K02 DC00149 to Carol Espy-Wilson. The general strategy for verification was suggested by Dr. Ken Stevens, of MIT. The authors would also like to thank Dr. Ronan Scaife, of Dublin City University, for the original MATLAB-callable code implementing VTCALCS.

Espy-Wilson, C. Y., Boyce, S. E., Jackson, M. T.-T., Narayanan, S., and Alwan, A. (**2000**). ''Acoustic modeling of American English /r/,'' J. Acoust. Soc. Am. **108**, 343–356.

Ladefoged, P., and Bhaskararao, P. (**1983**). ''Non-quantal aspects of consonant production: A study of retroflex consonants,'' J. Phonetics **11**, 291–302.

Lindau, M. (**1985**). ''The story of /r/,'' in *Phonetic Linguistics: Essays in Honor of Peter Ladefoged*, edited by V. A. Fromkin (Academic, Orlando), pp. 157–168.

Maeda, S. (**1982**). ''A digital simulation method of the vocal-tract system,'' Speech Commun. **1**, 199–229.

Perkell, J. S., Boyce, S. E., and Stevens, K. N. (**1979**). ''Articulatory and acoustic correlates of the [s-š] distinction,'' J. Acoust. Soc. Am. Suppl. 1 **65**, S24.

Stevens, K. N. (**1999**). *Acoustic Phonetics* (MIT Press, Cambridge, MA).