ON THE TRANSFER FUNCTION OF THE PIECEWISE-CYLINDRICAL VOCAL TRACT MODEL

Tamara SMYTH (trsmyth@ucsd.edu) [1] and **Devansh ZURALE**[1]

[1]**University of California San Diego**, La Jolla, USA

ABSTRACT

In this work, a matrix formulation of a piecewise one-dimensional waveguide model of the vocal tract (having varying cross-sectional area along its length) is used to derive model transfer functions suitable for both cylindrical or conical sections, with outputs tapped at the position of the glottis and the lips. The transfer function tapped at the lips is then considered in more detail for cylindrical waveguide sections and presented in its more useful form as a ratio of polynomial functions in the discrete frequency variable z, with coefficients vectors calculated for two cases: one where model boundaries are scalar losses and the other where losses are dependent on frequency. Through a transfer function with coefficients that are dependent on parameters of cross-sectional area and boundary conditions, the model may not only be controlled in real time, but the relationship to other vocal tract representations, in particular linear prediction coding (LPC) of speech, can be more easily shown, laying the foundation for inverse problems such as parameter estimation and source-filter separation. Finally, a comparison between model transfer function coefficients and those estimated by LPC (which assumes an all-pole filter) is discussed, suggesting that lower-order (and less computationally costly) LPC estimators might benefit from acoustically-informed boundary losses and the resulting introduction of zeros into the transfer function.

1. INTRODUCTION

The work herein borrows strategies for waveguide modeling of wind instrument bores and bells which, like the vocal tract, have shapes that are frequently not cylindrical or conical and thus have no known analytic solution. Though round-trip propagation delay in purely cylindrical and/or conical tubes may be modeled as a single one-dimensional waveguide (bi-directional delayline) element, the vocal tract has a varying cross-sectional area along its length and is better modeled using a piecewise approach (see Figure 1). Here, the vocal tract model is presented from the perspective of musical instrument modeling the desire to have parameters that can be both estimated and controlled in real time. To that end, the theory of piecewise waveguide modeling is reviewed [1–3] and a matrix formulation is presented that leads to parametric transfer functions modeling the vocal tract, one with output tapped at the position of the glottis and the other at the lips, initially with no assumption on whether waveguide sections are cylindrical of conical.

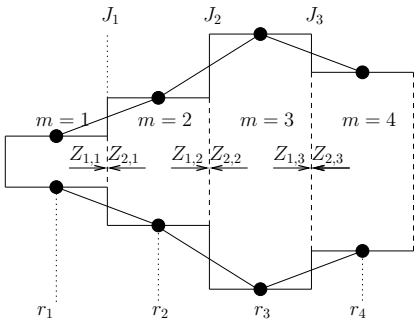

Figure 1. A sequence of $M = 4$ conical/cylindrical sections with radii r_m (and corresponding cross-sectional area) interleaved with $N = M - 1$ scattering junctions.

Though representations of the vocal tract have taken on different forms in the literature several can be made fundamentally equivalent—this is true of LPC [4], Kelly-Lockbaum [5] and piecewise models where sections are made uniformly cylindrical. These techniques model formants in the produced sound as a result of characteristic changes in the vocal tract shape and, under certain basic conditions/configurations, similarly result in an all-pole filter. The estimation of filter feedback coefficients using LPC is a frequently used technique and methods have been proposed to enhance its all-pole approximation by estimating, often iteratively, more accurate losses in the system [6]. Here, the contribution of vocal tract boundaries (lip reflection/transmission) and whether they are modeled as scalar or frequency-dependent losses, is shown to cause a divergence in model similarities. The suggestion is, therefore, that acoustically-informed boundaries as used in the piecewise cylindrical model, and the introduction of zeros into the transfer functions as a result of their inclusion, may enhance the all-pole LPC estimation without requiring higher and more computationally costly filter orders.

In the following, Section 2 reviews the theory of scattering junctions and first derives a scattering matrix for a single junction, then the "chain" scattering matrix for the complete vocal tract model. The matrix representations are then used to derive the model transfer functions in Section 3, applicable to both cylindrical and conical

waveguide sections. Section 4 considers the special case of cylindrical sections, showing a transfer function coefficient vector that is a function of the boundaries, first with an assumption of scalar losses that is most strongly related to LPC, then frequency-dependent losses, represented as a convolution in matrix form, that are more physically informed. Finally, in Section 5 a discussion is made on the relationship between waveguide model and LPC followed by a (preliminary) comparison of how both estimate the known glottal pulse from the output of a physical model.

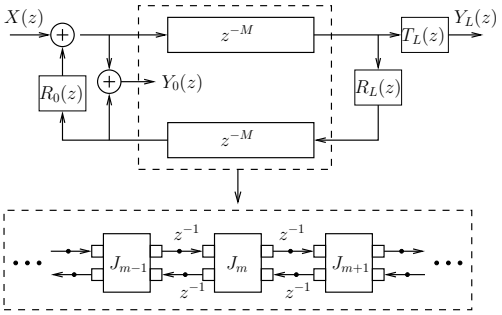

Figure 2. The vocal tract's varying cross-sectional area along its length may be implemented as a piecewise model—a cascade of two-port scattering junctions with interleaved unit-sample bi-directional delays (cylindrical/conical sections), with terminating boundary conditions $R_0(z)$ at the glottis and $R_L(z)$ at the lips and with outputs $Y_0(z)$ and $Y_L(z)$ in response to input $(X(z)$. Wall losses are omitted.

2. VOCAL TRACT SCATTERING MATRIX

As shown in Figures 1 and 2, the one-dimensional piecewise conical/cylindrical model of the vocal tract is comprised of M sections, each a bidirectional unit-sample delay, corresponding to acoustic propagation distance in one time sample, interleaved with $N = M - 1$ two-port $N_p = 2$ scattering junctions.

Scattering, the reflection and transmission of a wave that occurs when there is a change in the wave's characteristic impedance, may be modeled using a multi-port scattering junction where the number of ports N_p is twice the dimensionality of the wave propagation and the wave impedance on each port is determined by the medium (or geometry) it serves to connect. For waves propagating along the length of a *diverging* conical section terminated at port n a distance l_n from the cone apex, the wave impedance is a complex function of frequency given by

$$Z_n(l,\omega) = \frac{\rho c}{S_n} \cdot \frac{j\omega}{j\omega + c/l_n}, \tag{1}$$

where S_n is the cross-sectional at the port, ρ is the medium density and c is the propagation velocity. If the wave is propagating in the opposite direction toward the cone apex, effectively seeing a *converging* conical section, its impedance is given by the complex conjugate,

$$Z_n^*(l,\omega) = \frac{\rho c}{S_n} \cdot \frac{j\omega}{j\omega - c/l_n}. \tag{2}$$

For plane waves traveling in cylindrical sections, the distance l_n to the cone apex is infinite and the characteristic impedance reduces to a real value:

$$Z_n = \rho c / S_n. \tag{3}$$

Each of the junction's ports has a physical pressure p_n and volume velocity U_n that is the sum of wave components propagating in "i" and out "o" of port:

$$p_n = p_n^i + p_n^o, \quad \text{and} \quad U_n = U_n^i + U_n^o. \tag{4}$$

and which are related by the characteristic impedance:

$$U_n^i = \frac{p_n^i}{Z_n}, \quad U_n^o = -\frac{p_n^o}{Z_n^*}, \tag{5}$$

(the negative output volume velocity accounts for the fact that it is a directional quantity and moves in the direction in which it generates pressure [7]). Because the junction is shared by all mediums it connects, *the law for conservation of mass and momentum* dictate that the pressure at the junction be continuous and equal to the pressure on each port:

$$p_J = p_n = p_n^i + p_n^o, \tag{6}$$

and the sum of volume velocity on each port is equal zero,

$$\sum_{n=1}^{N_p} U_n = \sum_{n=1}^{N_p} (U_n^i + U_n^o) = 0. \tag{7}$$

For the two-port ($N_p = 2$) junction used in the piecewise vocal tract model, it follows from (6) that

$$p_1^i + p_1^o = p_2^i + p_2^o, \tag{8}$$

and from (7), with the substitution given by (5), that

$$\frac{p_1^i}{Z_1} - \frac{p_1^o}{Z_1^*} = -\left(\frac{p_2^i}{Z_2} - \frac{p_2^o}{Z_2^*}\right). \tag{9}$$

Equations (8) and (9) may be conveniently expressed in matrix form,

$$\mathbf{C} \begin{bmatrix} p_1^i \\ p_1^o \end{bmatrix} = \mathbf{D} \begin{bmatrix} p_2^i \\ p_2^o \end{bmatrix}, \tag{10}$$

where $\mathbf{C}$ and $\mathbf{D}$ are 2×2 matrices given by

$$\mathbf{C} = \begin{bmatrix} 1 & 1 \\ \frac{1}{Z_1} & -\frac{1}{Z_1^*} \end{bmatrix} \quad \text{and} \quad \mathbf{D} = \begin{bmatrix} 1 & 1 \\ -\frac{1}{Z_2} & \frac{1}{Z_2^*} \end{bmatrix}, \tag{11}$$

and rearranged to yield the expression relating left and right port, $n = 1$ and 2, respectively, input and output pressure wave components:

$$\begin{bmatrix} p_1^i \\ p_1^o \end{bmatrix} = \mathbf{C}^{-1}\mathbf{D} \begin{bmatrix} p_2^i \\ p_2^o \end{bmatrix}, \tag{12}$$

where

$$\mathbf{C}^{-1}\mathbf{D} = \frac{1}{\frac{1}{Z_1^*}+\frac{1}{Z_1}}\begin{bmatrix}\frac{1}{Z_1^*} & 1\\ \frac{1}{Z_1} & -1\end{bmatrix}\begin{bmatrix}1 & 1\\ -\frac{1}{Z_2} & \frac{1}{Z_2^*}\end{bmatrix} = \begin{bmatrix}\frac{Z_1(Z_2-Z_1^*)}{Z_2(Z_1+Z_1^*)} & \frac{Z_1(Z_2^*+Z_1^*)}{Z_2^*(Z_1+Z_1^*)}\\ \frac{Z_1^*(Z_2+Z_1)}{Z_2(Z_1+Z_1^*)} & \frac{Z_1^*(Z_2^*-Z_1)}{Z_2^*(Z_1+Z_1^*)}\end{bmatrix}. \quad (13)$$

As may be seen in Figure 3, the *left* port's input and output wave components are equal to the right and left traveling pressure waves, denoted by $+$ and $-$ superscripts respectively, in section m,

$$\begin{bmatrix}p_1^i\\ p_1^o\end{bmatrix} = \begin{bmatrix}p_m^+\\ p_m^-\end{bmatrix} = \mathbf{p}_m, \quad (14)$$

but the inverse relationship exists between wave components on the junction's *right* port and traveling waves in neighbouring section $m+1$,

$$\begin{bmatrix}p_2^i\\ p_2^o\end{bmatrix} = \begin{bmatrix}p_{m+1}^- z^{-1}\\ p_{m+1}^+ z\end{bmatrix} = \begin{bmatrix}0 & z^{-1}\\ z & 0\end{bmatrix}\mathbf{p}_{m+1}, \quad (15)$$

with vector element ordering made consistent with (14) by multiplying with an antidiagonal matrix that also accounts for the unit-sample delay/advance in one section.

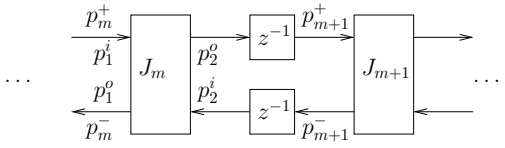

Figure 3. Relationship between the m^{th} junction's left and right (subscript 1 and 2, respectively) port input and output (superscript i and o, respectively) pressure wave components to the right ($+$ superscript) and left ($-$ superscript) traveling pressure waves in adjacent sections m and $m+1$.

With the change to traveling wave variables made by substituting (14) and (15) into (12), the relationship between right and left traveling pressure waves in adjacent sections m and $m+1$ may be given by

$$\mathbf{p}_m = \mathbf{A}_m\mathbf{p}_{m+1}, \quad (16)$$

where the scattering matrix for a single two-port junction

$$\mathbf{A}_m = \left(\mathbf{C}^{-1}\mathbf{D}\right)_m\begin{bmatrix}0 & z^{-1}\\ z & 0\end{bmatrix}, \quad (17)$$

is defined as a product of (13) for the m^{th} junction.

The "chain" scattering matrix for the complete vocal tract model is obtained by first expanding (16),

$$\mathbf{p}_m = \mathbf{A}_m\overbrace{\mathbf{p}_{m+1}}^{\mathbf{A}_{m+1}\mathbf{p}_{m+2}} = \mathbf{A}_m\mathbf{A}_{m+1}\overbrace{\mathbf{p}_{m+2}}^{\mathbf{A}_{m+2}\mathbf{p}_{m+3}} \quad (18)$$

so that traveling pressure waves in the first section can be expressed as a product of those in the final section,

$$\mathbf{p}_1 = \mathbf{P}_{M-1}\mathbf{p}_M, \quad (19)$$

where, for a sequence of M sections and $N = M-1$ junctions, the model's final 2×2 chain scattering matrix is given by the repeated product

$$\mathbf{P}_{M-1} = \prod_{m=1}^{M-1}\mathbf{A}_m = \begin{bmatrix}P_{1,1} & P_{1,2}\\ P_{2,1} & P_{2,2}\end{bmatrix}. \quad (20)$$

3. MODEL TRANSFER FUNCTIONS

To adequately represent the vocal tract so that it may be coupled to a dynamic model of the vocal folds as in [8], it is necessary to obtain two (2) transfer functions representing the model: one with the output pressure tapped at the lips $Y_L(z)$ and the other with the output pressure tapped at the glottis $Y_0(z)$ (see Figure 4).

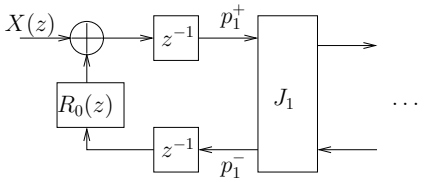

Figure 4. A signal flow diagram of the first waveguide section, showing how input $X(z)$ may be represented as a function of the glottis boundary $R_0(z)$ and traveling pressure waves p_1^+ and p_1^- as given by (21).

Both transfer functions are in response to input pressure $X(z)$ (corresponding to the product of the glottal flow and the characteristic impedance at the entry to the vocal tract) which, following Figure 4, can be defined in terms of the right and left traveling waves in the first section:

$$X(z) = p_1^+(z)z - R_0(z)p_1^-(z)z^{-1}, \quad (21)$$

which, by employing (19) and (20), may then be expressed in terms of traveling waves in the final section

$$X(z) = \left(P_{1,1}p_M^+(z) + P_{1,2}p_M^-(z)\right)z - R_0\left(P_{2,1}p_M^+ + P_{2,2}p_M^-\right)z^{-1}. \quad (22)$$

Finally, using the definition of the open-end lip reflection transfer function $R_L(z) = p_M^-(z)/p_M^+(z)$ and making the substitution $p_M^-(z) = R_L(z)p_M^+(z)$ in (22), the input may be expressed as a function of only the right traveling wave in the final section:

$$X(z) = p_M^+(z)\left(P_{1,1} + P_{1,2}R_L(z)\right)z - p_M^+(z)R_0(z)\left(P_{2,1} + P_{2,2}R_L(z)\right)z^{-1}. \quad (23)$$

The vocal tract transfer function $H_L(z) = Y_L(z)/X(z)$ tapped at the lips is defined as the ratio of output pressure $Y_L(z) = p_M^+(z)T_L(z)$ to input pressure $X(z)$ which, by substituting (23), yields

$$H_L(z) = \frac{T_L(z)z^{-1}}{P_{1,1}+P_{1,2}R_L(z) - R_0\left(P_{2,1}+P_{2,2}R_L(z)\right)z^{-2}}. \quad (24)$$

The transfer function $H_0(z) = Y_0(z)/X(z)$ is the ratio of the pressure at the glottis (vocal tract base)

$$\begin{aligned} Y_0(z) &= X(z) + p_1^-(z)(1 + R_0(z))z^{-1} \\ &= X(z) + \\ &\quad p_M^+(z)(P_{2,1} + P_{2,2}R_L(z))(1 + R_0(z))z^{-1}, \end{aligned}$$

to the system input $X(z)$ given by (23), yielding

$$H_0(z) = \frac{P_{1,1} + P_{1,2}R_L(z) + (P_{2,1} + P_{2,2}R_L(z))z^{-2}}{P_{1,1}+P_{1,2}R_L(z)-R_0\,(P_{2,1}+P_{2,2}R_L(z))\,z^{-2}}, \tag{25}$$

showing how boundary conditions $R_0(z)$ and $R_L(z)$ (further discussed in Section 4.2) and, for a cylindrical section, the assumed amplitude complementary transmission, which for pressure is given by

$$T_L(z) = 1 + R_L(z), \tag{26}$$

contribute to the vocal tract transfer functions.

Though the above transfer functions are sufficient for a frequency-domain representation/implementation of the model, it is preferable to represent it in its more useful form as a ratio of polynomials in the (discrete) frequency variable z, both to allow for time-domain implementation using the corresponding difference equation (obtained by taking the inverse z-transform) and also for comparison (or mapping) to all-pole filter coefficients estimated by LPC.

4. POLYNOMIAL TRANSFER FUNCTION FOR CYLINDRICAL SECTIONS

Though vocal tract sections may be modeled as being either cylindrical or conical (see Figure 1) and the above derivation makes no assumption of either, the section shape is dependent on the choice of expression for impedance (1)-(3) in the matrix given by (13). Conical sections in a time-domain synthesis would require fitting a digital filter to the complex impedances given by (1) and (2) as was done in [3] using the impulse-invariant method [9] and also in [10] using the bilinear transform. Using cylindrical sections, on the other hand, has considerable computational convenience for computing the transfer functions as a ratio of polynomials, as well as allowing better comparison with LPC and related Kelly-Lochbaum models.

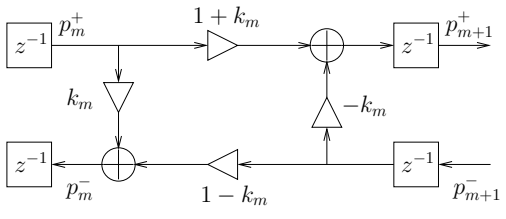

Figure 5. Kelly-Lochbaum scattering junction.

For cylindrical sections, as mentioned in Section 2, the characteristic wave impedance (3) is not a function of frequency but rather a real value inversely proportional to the plane wave's surface area. For waves on left and right ports of the junction between sections m and $m + 1$, this area is the section's cross-sectional area S_m and S_{m+1}, respectively, and the scattering matrix (13) may be reduced to

$$\begin{aligned} \left(\mathbf{C^{-1}D}\right)_m &= \frac{1}{2S_m}\begin{bmatrix} S_m - S_{m+1} & S_m + S_{m+1} \\ S_m + S_{m+1} & S_m - S_{m+1} \end{bmatrix} \\ &= \frac{1}{1 + k_m}\begin{bmatrix} k_m & 1 \\ 1 & k_m \end{bmatrix}, \end{aligned} \tag{27}$$

where the reflection coefficient between sections

$$k_m = \frac{S_m - S_{m+1}}{S_m + S_{m+1}}, \tag{28}$$

is that used in LPC and forms the Kelly-Lochbaum scattering junction shown in Figure 5.

Substituting (27) into (17) yields the single junction scattering matrix between sections sections m and $m + 1$ for the piecewise *cylindrical* model

$$\mathbf{A}_m = \frac{1}{1 + k_m}\begin{bmatrix} k_m & 1 \\ 1 & k_m \end{bmatrix}\begin{bmatrix} 0 & z^{-1} \\ z & 0 \end{bmatrix}, \tag{29}$$

which, for $M = 2$ sections and $N = 1$ junction, yields a model scattering matrix (20) given by

$$\mathbf{P}_1 = \mathbf{A}_1 = \frac{z}{1 + k_1}\begin{bmatrix} 1 & k_1 z^{-2} \\ k_1 & z^{-2} \end{bmatrix}, \tag{30}$$

and for $M = 3$ sections and $N = 2$ junctions,

$$\begin{aligned} \mathbf{P}_2 = \quad & \mathbf{A}_1\mathbf{A}_2 = \mathbf{P}_1\mathbf{A}_2 \\ &= \frac{z}{1 + k_1}\begin{bmatrix} 1 & k_1 z^{-2} \\ k_1 & z^{-2} \end{bmatrix}\frac{z}{1 + k_2}\begin{bmatrix} 1 & k_2 z^{-2} \\ k_2 & z^{-2} \end{bmatrix} \\ &= \frac{z^2}{\prod_{m=1}^{2}(1 + k_m)}\begin{bmatrix} c_0 + c_2 z^{-2} & d_2 z^{-2} + d_0 z^{-4} \\ d_0 + d_2 z^{-2} & c_2 z^{-2} + c_0 z^{-4} \end{bmatrix}, \end{aligned}$$

where polynomial matrix elements have coefficients

$$c_0 = 1, \quad c_2 = k_1 k_2, \quad d_0 = k_1 \quad \text{and} \quad d_2 = k_2. \tag{31}$$

In general, for models having M sections and $N = M - 1$ junctions, the chain scattering matrix is given by

$$\mathbf{P}_N = \prod_{m=1}^{N}\mathbf{A}_m = \mathbf{P}_{N-1}\mathbf{A}_N = \frac{z^N}{\prod_{m=1}^{N}(1 + k_m)}\mathbf{K}_N, \tag{32}$$

where

$$\begin{aligned} \mathbf{K}_N &= \begin{bmatrix} K_{1,1} & K_{1,2} \\ K_{2,1} & K_{2,2} \end{bmatrix} = \prod_{m=1}^{N}\begin{bmatrix} 1 & k_m z^{-2} \\ k_m & z^{-2} \end{bmatrix} \\ &= \begin{bmatrix} \sum_{m=0}^{N-1} c_{2m} z^{-2m} & \sum_{m=1}^{N} d_{2(N-m)} z^{-2m} \\ \sum_{m=0}^{N-1} d_{2m} z^{-2m} & \sum_{m=1}^{N} c_{2(N-m)} z^{-2m} \end{bmatrix}. \end{aligned} \tag{33}$$

Polynomial entries in (33) have initial coefficients given by

$$c_0 = 1 \quad \text{and} \quad d_0 = k_1, \tag{34}$$

with remaining coefficients being recursively defined by

$$\begin{aligned}\mathbf{c}_N &= \begin{bmatrix}\mathbf{c}_{N-1} & 0 & 0\end{bmatrix}^\intercal + k_N \begin{bmatrix}0 & \tilde{\mathbf{d}}_{N-1} & 0\end{bmatrix}^\intercal \\ \mathbf{d}_N &= \begin{bmatrix}\mathbf{d}_{N-1} & 0 & 0\end{bmatrix}^\intercal + k_N \begin{bmatrix}0 & \tilde{\mathbf{c}}_{N-1} & 0\end{bmatrix}^\intercal,\end{aligned} \tag{35}$$

where $\tilde{\cdot}$ denotes the retrograde vector, one where the order of elements is reversed (e.g. by multiplying with the exchange, or backward identity, matrix of appropriate size), and where the length-$2N$ coefficient vectors have the form

$$\begin{aligned}\mathbf{c}_N &= \begin{bmatrix}c_0 & 0 & c_2 & 0 & \dots & c_{2(N-1)} & 0\end{bmatrix}^\intercal \\ \mathbf{d}_N &= \begin{bmatrix}d_0 & 0 & d_2 & 0 & \dots & d_{2(N-1)} & 0\end{bmatrix}^\intercal,\end{aligned} \tag{36}$$

with odd-ordered coefficients (even-numbered vector elements) being zero since polynomial entries in (33) have only even-ordered terms (corresponding to the one-sample propagation delay per section and between junctions). Consistent with (31), for a number of junctions $N > 1$, final coefficients are given by

$$c_{2(N-1)} = k_1 k_N \quad \text{and} \quad d_{2(N-1)} = k_N. \tag{37}$$

Coefficient vectors for $N = 1$ are obtained by (34):

$$\mathbf{c}_1 = \begin{bmatrix}c_0 \\ 0\end{bmatrix} = \begin{bmatrix}1 \\ 0\end{bmatrix} \quad \text{and} \quad \mathbf{d}_1 = \begin{bmatrix}d_0 \\ 0\end{bmatrix} = \begin{bmatrix}k_1 \\ 0\end{bmatrix}, \tag{38}$$

for $N = 2$, by (35) or directly from (37):

$$\mathbf{c}_2 = \begin{bmatrix}c_0 \\ 0 \\ c_2 \\ 0\end{bmatrix} = \begin{bmatrix}1 \\ 0 \\ k_1 k_2 \\ 0\end{bmatrix}, \quad \mathbf{d}_2 = \begin{bmatrix}d_0 \\ 0 \\ d_2 \\ 0\end{bmatrix} = \begin{bmatrix}k_1 \\ 0 \\ k_2 \\ 0\end{bmatrix}, \tag{39}$$

and for $N = 3$, by (35) (expanded for illustration):

$$\begin{aligned}\mathbf{c}_3 &= \begin{bmatrix}\mathbf{c}_2 & 0 & 0\end{bmatrix}^\intercal + k_3 \begin{bmatrix}0 & \tilde{\mathbf{d}}_2 & 0\end{bmatrix}^\intercal \\ &= \begin{bmatrix}1 \\ 0 \\ k_1 k_2 \\ 0 \\ 0 \\ 0\end{bmatrix} + k_3 \begin{bmatrix}0 \\ 0 \\ k_2 \\ 0 \\ k_1 \\ 0\end{bmatrix} = \begin{bmatrix}1 \\ 0 \\ k_1 k_2 + k_2 k_3 \\ 0 \\ k_1 k_3 \\ 0\end{bmatrix} = \begin{bmatrix}c_0 \\ 0 \\ c_2 \\ 0 \\ c_4 \\ 0\end{bmatrix} \\ \mathbf{d}_3 &= \begin{bmatrix}\mathbf{d}_2 & 0 & 0\end{bmatrix}^\intercal + k_3 \begin{bmatrix}0 & \tilde{\mathbf{c}}_2 & 0\end{bmatrix}^\intercal \\ &= \begin{bmatrix}k_1 \\ 0 \\ k_2 \\ 0 \\ 0 \\ 0\end{bmatrix} + k_3 \begin{bmatrix}0 \\ 0 \\ k_1 k_2 \\ 0 \\ 1 \\ 0\end{bmatrix} = \begin{bmatrix}k_1 \\ 0 \\ k_2 + k_1 k_2 k_3 \\ 0 \\ k_3 \\ 0\end{bmatrix} = \begin{bmatrix}d_0 \\ 0 \\ d_2 \\ 0 \\ d_4 \\ 0\end{bmatrix}\end{aligned} \tag{40}$$

and so on for models having a greater number of junctions.

With the scattering matrix $\mathbf{P}_N$ defined in (32) for cylindrical sections, substitution may be made into (24) to yield transfer functions in their more useful form, as a ratio of polynomial functions in z. The final expression for numerator and denominator polynomials are, however, dependent on the boundary conditions $R_0(z)$ and $R_L(z)$, whether they are scalar or frequency dependent and, in the latter case, the filter order.

4.1 $H_L(z)$ for Scalar Boundaries

In the simplified case (yet important because of its close relationship to LPC coefficients) of scalar boundaries, any losses may be lumped in R_0 and the reflection at the lips simply made lossless but inverting $R_L = -1$ (and thus no longer a function of z). Since, by (26), this would yield a transmission at the lips given by $T_L = 1 + R_L = 0$ and a complete attenuation of the signal, the transmission is omitted for this case. After a substitution of (32), the transfer function (24) therefore becomes

$$\begin{aligned}H_L(z) &= \frac{z^{-1}}{P_{1,1} + P_{1,2}R_L - R_0(P_{2,1} + P_{2,2}R_L)z^{-2}} \\ &= \frac{z^{-(N+1)} \prod_{m=1}^{N}(1 + k_m)}{K_{1,1} + K_{1,2}R_L - R_0(K_{2,1} + K_{2,2}R_L)z^{-2}} \\ &= \frac{B(z)}{A(z)},\end{aligned} \tag{41}$$

with the numerator being a pure delay with a scalar value,

$$B(z) = z^{-(N+1)} \prod_{m=1}^{N}(1 + k_m), \tag{42}$$

showing $H_L(z)$ has no zeros, and a denominator given by

$$\begin{aligned}A(z) &= K_{1,1} + K_{1,2}R_L - R_0\left(K_{2,1} + K_{2,2}R_L\right) z^{-2} \\ &= a_0 z^{-0} + a_1 z^{-1} + \cdots + a_{2(N+1)} z^{-2(N+1)},\end{aligned} \tag{43}$$

with polynomial coefficients given by the (column) vector

$$\mathbf{A}_N = \mathbf{C}_N \mathbf{R}, \tag{44}$$

where $\mathbf{C}_N$ is $(2N+3)\times 4$ matrix with columns constructed from coefficient vectors $\mathbf{c}_N$ and $\mathbf{d}_N$ given in (35),

$$\begin{aligned}\mathbf{C}_N &= \begin{bmatrix}\mathbf{c}_N & 0 & 0 & 0 \\ 0 & \tilde{\mathbf{d}}_N & 0 & 0 \\ 0 & 0 & \mathbf{d}_N & 0 \\ 0 & 0 & 0 & \tilde{\mathbf{c}}_N\end{bmatrix} \\ &= \begin{bmatrix}c_0 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ 0 & 0 & \mathbf{0} & \mathbf{0} \\ c_2 & d_{2(N-1)} & d_0 & \mathbf{0} \\ 0 & 0 & 0 & 0 \\ c_4 & d_{2(N-2)} & d_2 & c_{2(N-1)} \\ 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ c_{2(N-1)} & d_2 & d_{2(N-2)} & c_4 \\ 0 & 0 & 0 & 0 \\ \mathbf{0} & d_0 & d_{2(N-1)} & c_2 \\ \mathbf{0} & \mathbf{0} & 0 & 0 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & c_0\end{bmatrix}\end{aligned} \tag{45}$$

with each column extending the length-$2N$ vector by 3 zeros (in bold for better visibility) to accomodate a downward shift by one element from one column to the next, and where $\mathbf{R}$ is a 4×1 column vector

$$\mathbf{R} = \begin{bmatrix}1 & R_L & -R_0 & -R_0 R_L\end{bmatrix}^\intercal = \begin{bmatrix}1 & -1 & -R_0 & R_0\end{bmatrix}^\intercal \tag{46}$$

holding scalar boundaries R_0 and $R_L = -1$. It may be observed from (44)-(46) that the interleaved zeros characterizing vectors $\mathbf{c}_N$ and $\mathbf{d}_N$ carries over to the coefficient vector $\mathbf{A}_N$. Further, since by (34) $c_0 = 1$, the first element of $\mathbf{A}_N$ is $a_0 = 1$ and the last element is $a_{2(N+1)} = R_0$ so that the structure of the length $2N + 3$ coefficient vector is

$$\mathbf{A}_N = \begin{bmatrix} 1 & 0 & a_2 & \dots & 0 & a_{2N} & 0 & R_0 \end{bmatrix}^\mathsf{T}. \quad (47)$$

Equation (43) shows that for M cylindrical sections and $N = M - 1$ junctions using scalar boundaries R_0 and $R_L = -1$, the transfer function $H_L(z)$ is of order $2(N+1)$ and, save a scalar with pure delay in the numerator (42), an all-pole filter consistent with the assumption made in LPC of speech [4]. In fact, for cylindrical sections with simplified scalar boundaries, the coefficient vector $\mathbf{A}_N$ corresponds (save rounding error) to the autoregressive predictor coefficients estimated directly from the impulse response of $H_L(z)$ by LPC when the order is $2(N + 1)$ (and the unknown glottal flow and *true* boundary losses do not contribute to the observed signal and complicate the prediction—see Appendix 1).

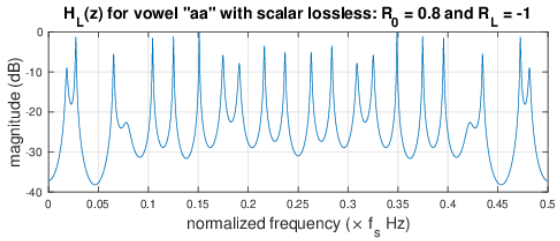

Figure 6. A scalar boundary loss, $R_0 = 0.8$ and $R_L = -1$ produces a symmetry in the half-bandwidth of the frequency response magnitude.

Factoring $\mathbf{A}_N$ for the scalar loss case show poles that are symmetric about the unit circle and a corresponding symmetry in the quarter-bandwidth (sampling rate divided by four) of the frequency response magnitude, as seen in Figure 6 for area functions of the vowel sound "aa" [11].

4.2 $H_L(z)$ for Frequency-Dependent Boundaries

In more practical applications of voice modeling, though the reflection at the glottis can often be approximated by a scalar, the reflection at the mouth is better modeled by accounting for frequency-dependent loss. Borrowing from work in which waveguide elements are estimated from measurement [12] and in which the open-end reflection of a cylindrical tube was shown to be very close to theoretical expectation [13], it was found here that a cascade of two first-order shelf filters, producing a second-order-section (SOS) and having the form

$$R_L(z) = \frac{B_L(z)}{A_L(z)} = -\frac{(b_L)_0 + (b_L)_1 z^{-1} + (b_L)_2 z^{-2}}{1 + (a_L)_1 z^{-1} + (a_L)_2 z^{-2}}, \quad (48)$$

with coefficient vectors

$$\begin{aligned} \mathbf{B}_L &= \begin{bmatrix} (b_L)_0 & (b_L)_1 & (b_L)_2 \end{bmatrix} \\ \mathbf{A}_L &= \begin{bmatrix} (a_L)_0 & (a_L)_1 & (a_L)_2 \end{bmatrix}, \end{aligned} \quad (49)$$

and with a transition frequency of $\omega_t = c/r_M$, as shown in Figure 7, produced a very good fit. With the assumption that the transmission is the amplitude complement of the lip reflection (26), it may be given here by

$$T_L(z) = 1 + R_L(z) = \frac{A_L(z) + B_L(z)}{A_L(z)}. \quad (50)$$

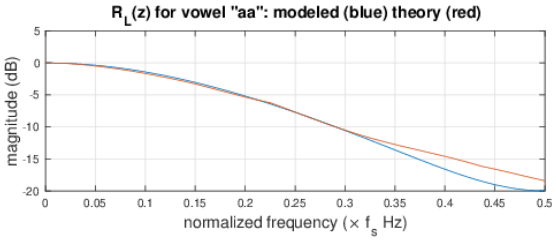

Figure 7. A cascade of two first-order shelf filters, each with band-edge gain of -10 dB, produces a second-order filter having the form given in (48) and with transition $f_t = c/(2\pi r_M)$ Hz (blue). This modeled response produces a good fit to the theoretical response of an open-end cylindrical tube having radius r_M, the radius of the final cylindrical section when the vocal tract model is configured to the vowel sound "aa" (red).

Substituting (48) and (50) into (41) yields the vocal tract transfer function tapped at the lips with frequency-dependent boundaries (denoted by $\hat{\cdot}$):

$$\begin{aligned} \hat{H}_L(z) &= \frac{T_L(z) z^{-(N+1)} \prod_{m=1}^{N} (1 + k_m)}{K_{1,1} + K_{1,2} \dfrac{B_L(z)}{A_L(z)} - R_0 \left(K_{2,1} + K_{2,2} \dfrac{B_L(z)}{A_L(z)} \right) z^{-2}} \\ &= \frac{\hat{B}(z)}{\hat{A}(z)}, \end{aligned} \quad (51)$$

where the numerator as a polynomial in z is given by

$$\begin{aligned} \hat{B}(z) &= (A_L(z) + B_L(z)) z^{-(N+1)} \prod_{m=1}^{N} (1 + k_m) \\ &= \left(b_0 + b_1 z^{-1} + b_2 z^{-2} \right) z^{-(N+1)} \prod_{m=1}^{N} (1 + k_m), \end{aligned}$$

having coefficients obtained by summing vectors in (49),

$$\begin{bmatrix} b_0 & b_1 & b_2 \end{bmatrix} = \mathbf{A}_L + \mathbf{B}_L, \quad (52)$$

and showing an introduction of zeros into the all-pole transfer function $H_L(z)$ given in (41) for scalar boundaries. The denominator of (51) as a polynomial in z is given by

$$\begin{aligned} \hat{A}(z) &= K_{1,1} A_L(z) + K_{1,2} B_L(z) - \\ &\qquad R_0 \left(K_{2,1} A_L(z) + K_{2,2} B_L(z) \right) z^{-2} \\ &= \hat{a}_0 z^{-0} + \hat{a}_1 z^{-1} + \cdots + \hat{a}_{2(N+2)} z^{-2(N+2)}, \end{aligned} \quad (53)$$

showing polynomial multiplication terms that require (acyclic) convolution of coefficients to produce coefficient vector $\hat{\mathbf{A}}_N$ which, in matrix form, is given by

$$\begin{aligned}\hat{\mathbf{A}}_N &= \begin{bmatrix}\hat{a}_0 & \hat{a}_1 & \hat{a}_2 & \dots & \hat{a}_{2(N+2)}\end{bmatrix} \\ &= \begin{bmatrix}\mathbf{C}_N \\ 0\,0\,0\,0 \\ 0\,0\,0\,0\end{bmatrix}\hat{\mathbf{R}}_0 + \begin{bmatrix}0\,0\,0\,0 \\ \mathbf{C}_N \\ 0\,0\,0\,0\end{bmatrix}\hat{\mathbf{R}}_1 + \begin{bmatrix}0\,0\,0\,0 \\ 0\,0\,0\,0 \\ \mathbf{C}_N\end{bmatrix}\hat{\mathbf{R}}_2,\end{aligned} \tag{54}$$

where $\mathbf{C}_N$ from (45) is extended by two rows of zeros to accommodate the convolution length $(2N+3)+2$ (the sum of $\mathbf{C}_N$ column length and $R_L(z)$ coefficient vector length minus one) and the downward shift of one row for each subsequent term in the sum. Equation (54) also requires a modification of the scalar boundary vector in (46) so it holds coefficients of $R_L(z)$ for corresponding n^{th}-order terms as indicated by the subscript n:

$$\hat{\mathbf{R}}_n = \begin{bmatrix}(a_L)_n & (b_L)_n & -(a_L)_n R_0 & -(b_L)_n R_0\end{bmatrix}^\intercal. \tag{55}$$

Finally, in addition to being two elements longer than $\mathbf{A}_N$, it is, perhaps, worthwhile to note that the coefficient vector $\hat{\mathbf{A}}_N$ no longer has the structure of interleaved zeros for odd-ordered terms, and is a vector more typical of an LPC estimation from an actual recorded speech signal.

5. DISCUSSION AND CONCLUSIONS

As mentioned in Section 4 and Appendix 1, the coefficient vector $\mathbf{A}_N$ corresponds to the linear prediction coefficients estimated from the impulse response of $H_L(z)$ if the LPC order is $2(N+1)$. If, on the other hand, the LPC estimation is on the impulse response of $\hat{H}_L(z)$ and, correspondingly, the order is increased to $2(N+2)$, the order of $\hat{\mathbf{A}}_N$, the estimated coefficients will not accurately correspond to $\hat{\mathbf{A}}_N$ since the assumption of an all-pole filter no longer holds. Though an increased order creates a better fit, this also introduces computational cost and, more significantly, impedes the inverse problem by placing the burden of representing losses on an increased number of reflection coefficients k_m, thus reducing their correlation to vocal tract length and cross-sectional area (parameters frequently estimated from LPC coefficients).

Another way of comparing the LPC estimation to the piecewise cylindrical waveguide model is by testing its (in)accuracy in the inverse problem of source-filter separation and glottal flow estimation. Though a rigorous treatment is beyond the scope of this work, a preliminary attempt at separating a model [8] generated volume flow (source) from the vocal tract model $H_L(z)$ presented herein (filter), can provide some insight into the role of the boundaries. Consistent with expectation, as shown in Figure 9, the inverse filter constructed with LPC estimated coefficients produces a signal (middle) that is closer to the signal produced by the inverse of $\hat{H}_L$ *without* the transmission filter $T_L(z)$ (bottom), a signal known as the flow *derivative*, frequently estimated and fit to the well-known parametric LF source model [14, 15]). When this flow derivative is passed through an inverse lip radiation

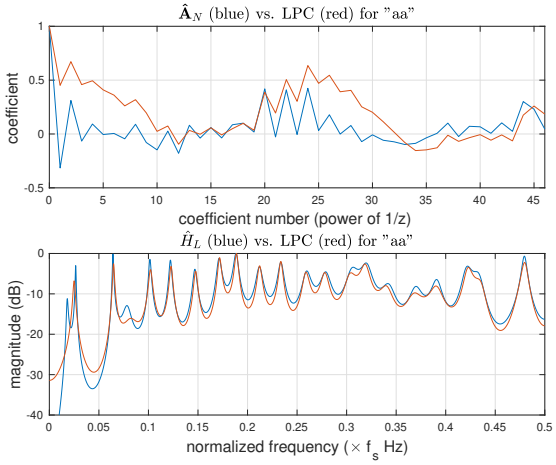

Figure 8. Coefficient vector $\hat{\mathbf{A}}_z$ and same-order LPC estimation from the impulse response of $\hat{H}_L(z)$ (top) and corresponding frequency response magnitudes (bottom).

(derivative) filter $L(z) = 1 - dz^{-1}$, where d is close to one ([6, 16]), the resulting signal (Figure 10, middle) is effectively integrated and shows a closer fit to the original volume flow (Figure 10, top), strongly suggesting that an accurate (acoustically-informed) lip reflection with amplitude-complementary transmission, may improve the problem of source estimation.

6. APPENDIX 1

For the transfer function

$$H_L(z) = \frac{Y_L(z)}{X(z)} = \frac{z^{-(N+1)}\prod_{m=1}^{N}(1+k_m)}{1+\sum_{i=1}^{2(N+1)} a_i z^{-i}}, \tag{56}$$

the difference equation is given by the inverse z-transform to yield output $y(n)$ at time sample n:

$$y(n) = \prod_{m=1}^{N}(1+k_m)x(n-(N+1)) - \sum_{i=1}^{2(N+1)} a_i y(n-i). \tag{57}$$

The impulse response $h(n)$ is the output in response to an input that is the unit step function $x(n) = u(n)$ which, by definition, has a non-zero value only when $n = N+1$, yielding

$$h(n) = \begin{cases} 0, & \text{for } n < N+1 \\ \prod_{m=1}^{N}(1+k_m), & \text{for } n = N+1 \\ -\sum_{i=1}^{2(N+1)} a_i h(n-i), & \text{for } n > N+1. \end{cases} \tag{58}$$

In linear prediction, future values of a discrete-time signal are estimated as a linear function of previous samples, a model that may be represented by an expression that is very similar to the final case of (58) where the impulse response $h(n)$ is defined for $n > N+1$ (an actual model

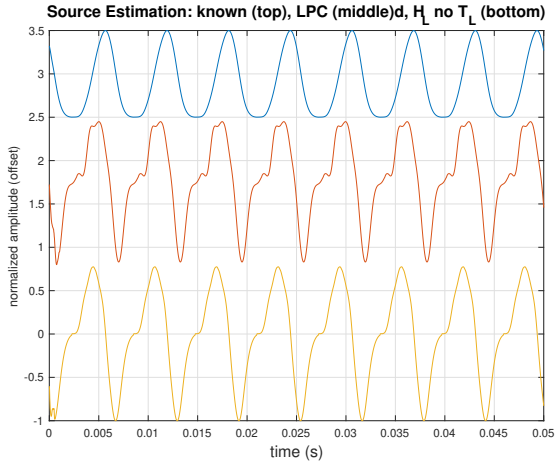

Figure 9. Known model glottal flow (top); signal from inverse filter with LPC estimated coefficients (middle); signal from $1/\hat{H}_L(z)$ *without* transmission $T_L(z)$ (bottom).

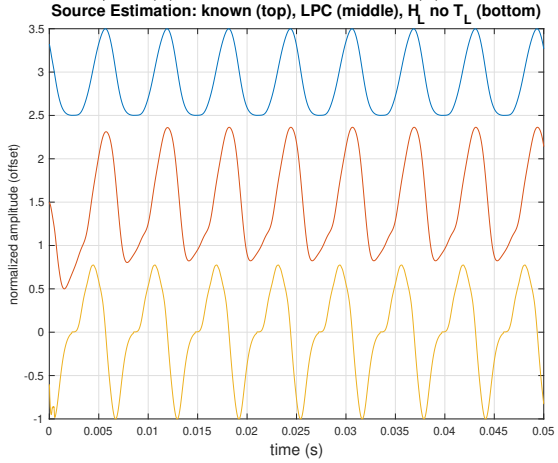

Figure 10. Known model glottal flow (top); signal from inverse filter with LPC estimated coefficients *with* the inverse of a lip radiation filter $1/L(z)$ (middle); signal from $1/\hat{H}_L(z)$ *without* transmission $T_L(z)$ (bottom).

would also account for prediction error). In practice, for a sufficiently long $h(n)$, one with enough samples that the infinite impulse response is allowed to decay very close to zero, the estimation of coefficients corresponding to $\mathbf{A}_N$ may be made with negligible error.

7. REFERENCES

[1] D. P. Berners, "Acoustics and signal processing techniques for physical modeling of brass instruments," Ph.D. dissertation, Stanford University, Stanford, California, July 1999.

[2] T. Smyth and F. S. Scott, "Trombone synthesis by model and measurement," *EURASIP Journal on Advances in Signal Processing*, vol. 2011, no. Article ID 151436, p. 13 pages, 2011, doi:10.1155/2011/151436.

[3] V. Välimäki and M. Karjalainen, "Improving the Kelly-Lochbaum vocal tract model using conical tube sections and fractional delay filtering techniques," in *Proceedings of the 3^{rd} International Conference on Spoken Language Processing (ICSLP)*, Yokohama, Japan, January 1994.

[4] J. D. Markel and A. H. Gray, *Linear Prediction of Speech*. Springer-Verlag, 1976.

[5] J. Jr and C. Lochbaum, "Speech synthesis," *Proceedings of the fourth International Congress on Acoustics*, vol. G42, pp. 1–4, 1962.

[6] O. Perrotin and I. McLoughlin, "A spectral glottal flow model for source-filter separation of speech," in *Proceedings of 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brighton, UK, May 2019.

[7] J. O. Smith, *Digital Waveguide Modeling of Musical Instruments*. ccrma.stanford.edu/~jos/waveguide/, 2003, last viewed 12/4/08.

[8] T. Smyth and A. Fathi, "Voice synthesis using the generalized pressure controlled-valve," in *Proceedings of ICMC 2008*, Belfast, Ireland, August 2008, pp. 57–60.

[9] J. O. Smith, "Physical audio signal processing for virtual musical instruments and audio effects," http://ccrma.stanford.edu/~jos/pasp/, December 2008, last viewed 8/24/2009.

[10] H. W. Strube, "Are conical segments useful for vocal tract simulation?" *Journal of the Acoustical Society of America*, vol. 114, no. 6, pp. 3028–3031, December 2003.

[11] B. H. Story and I. R. Titze, "Vocal tract area functions from magnetic resonance imaging," *Journal of the Acoustical Society of America*, vol. 100, no. 1, pp. 537–554, July 1996.

[12] T. Smyth and J. Abel, "Estimating waveguide model elements from acoustic tube measurements," *Acta Acustica united with Acustica*, vol. 95, no. 6, pp. 1093–1103, 2009.

[13] H. Levine and J. Schwinger, "On the radiation of sound from an unflanged circular pipe," *Phys. Rev*, vol. 73, no. 4, pp. 383–406, 1948.

[14] G. Fant, J. Liljencrants, and Q. Lin, "A four-parameter model of glottal flow," *Royal Institute of Technologies - Dept. for Speech, Music and Hearing, Quarterly Progress and Status Report 4*, 1985.

[15] H.-L. Lu and J. O. Smith, "Joint estimation of vocal tract filter and glottal source waveform via convex optimization," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA'99)*, New Paltz, NY, October 1999, pp. 79–92.

[16] I. V. McLoughlin, *Speech and Audio Processing: a MATLAB-based approach*. Cambridge University Press, 2016.