ON THE ROLE OF LIP REFLECTION/TRANSMISSION IN THE RELATIONSHIP BETWEEN LPC AND WAVEGUIDE VOCAL TRACT MODELS

Tamara Smyth[1], *Devansh Zurale*

UC San Diego, Department of Music, La Jolla CA
[1]trsmyth, dzurale@ucsd.edu

ABSTRACT

In this work, the relationship and, under certain conditions equivalence, between LPC and a piecewise cylindrical waveguide (or Kelly-Lochbaum) model of the vocal tract is found to be largely tied to the lip reflection and transmission. Herein, the elements of a piecewise waveguide model are reviewed and final corresponding transfer functions presented for two cases, one in which the boundary losses are scalar, showing a more obvious relationship to the all-pole LPC estimation, and one in which the lip reflection is frequency dependent, which introduces a zero in the transfer function and a more obfuscated relationship to LPC. In previous work, it was found that the theoretical open-end reflection of a cylindrical tube is well approximated by a single pole filter. Here it is modeled as a first-order shelf filter, which has a useful property that simplifies its estimation from denominator polynomial coefficients. For both scalar and frequency-dependent cases, waveguide boundary losses are estimated directly from LPC coefficients showing a significant role of lip reflection/transmission in the relationship between waveguide and LPC models, and in particular, how the elements of one might be fit to elements of the other.

Index Terms— speech synthesis, waveguide model, LPC, vocal tract modeling, Kelly-Lochbaum

1. INTRODUCTION

It is widely accepted that the human voice can be modeled by source-filter techniques, that is, that pressure in the vocal tract, *the filter*, doesn't significantly influence the vibration of the vocal folds and the resulting train of glottal pulses into the vocal tract that produce *the source*, and that their coupling is unidirectional with very little (if any) feedback from filter to source.

Both waveguide synthesis [1] (including Kelly-Lochbaum [2] and the equivalent piecewise cylindrical model [3]) and linear predictive coding LPC [4] have been used to model the vocal tract filter, with a relationship between the two strongly implied [5]. One of the main differences between waveguide synthesis and LPC is in the modeling approach, the former simulating the physical system governing wave propagation in the vocal tract while the latter operating on the spectrum of the voice signal, analyzing and fitting an all-pole filter to its spectral envelope so as to identify (and even separate) components that are due to the source, the glottal pulse (or flow) and those that are due to the filter, the vocal tract. In the end, it is perhaps not surprising that the two methods, at least under certain situations, can be shown to be equivalent.

In this work, the transfer function corresponding to the piecewise cylindrical waveguide model is provided [6] and it's relationship to the all-pole filter is examined for two cases, one where the boundary losses are scalar and one more accurate representation, where the boundary, and specifically the lip reflection/transmission is frequency dependent. The first case produces an all-pole filter having a more obvious relationship to that estimated by LPC, while the second case can be modified (by deconvolving a simple first-order high-pass filter) to produce an all-pole from which the LPC coefficients may be estimated. In both cases, an accurate all-pole coefficient vector can be used to estimate the boundary losses.

2. PIECEWISE CYLINDRICAL WAVEGUIDE MODEL

Though round-trip propagation delay in purely cylindrical and/or conical tubes may be modeled as a single waveguide (see Figure 1, top), the vocal tract has a varying cross sectional area along its length (see Figure 2) that is better modeled using a piecewise approach [7, 8], consisting of a sequence of M sections, each a bidirectional unit-sample delay, interleaved with $N = M - 1$ two-port scattering junctions accounting for the reflection and transmission that occurs with a change in cross-section (see Figure 1, bottom). Boundary losses are modeled as a reflection at the glottis R_0 (assumed herein to be scalar) and a reflection and transmission at the lip, $R_L(z)$ and $T_L(z)$, respectively. The input $X(z)$ corresponds to the glottal flow, and the output $Y_L(z)$ to the speech signal.

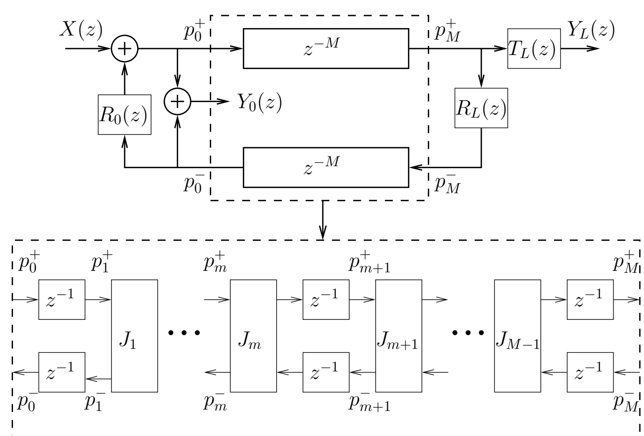

Figure 1: A waveguide model of an acoustic tube having terminating boundary conditions $R_0(z)$ (at the source/input $X(z)$) and $R_L(z)$, along with transmission $T_L(z)$ producing output $Y_L(z)$. The pure delay of M samples suitable for conical/cylindrical tubes is replaced by a sequence of unit-sample bidirectional delays interleaved with two-port scattering junctions that model reflection/transmission at changes in the vocal tract's cross-sectional area.

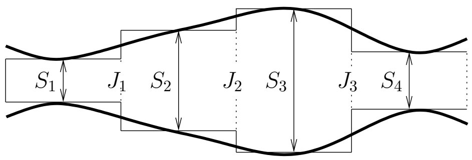

Figure 2: A vocal tract with varying cross-sectional area along its length is modeled here as a sequence of four cylindrical sections with cross-sectional areas S_1, S_2, S_3, and S_4, interleaved with three two-port scattering junctions J_1, J_2 and J_3.

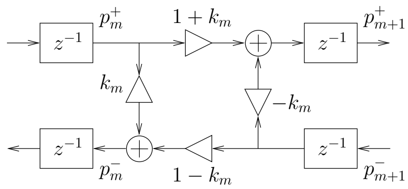

Figure 3: The Kelly-Lochbaum scattering junction, showing how right and left traveling wave components in adjacent sections m and $m+1$ are related by reflection coefficient k_m.

For cylindrical sections, the relationship between pressure wave components in adjacent sections m and $m+1$, separated by the two-port scattering junction J_m, may be represented in matrix form:

$$\mathbf{p}_m = \mathbf{A}_m \mathbf{p}_{m+1}, \tag{1}$$

where the column vector holding right and left traveling waves (denoted by $+$ and $-$ superscripts, respectively) in any section m is

$$\mathbf{p}_m = \begin{bmatrix} p_m^+ \\ p_m^- \end{bmatrix}, \tag{2}$$

and the scattering matrix for cross-sectional areas S_m and S_{m+1} is

$$\begin{aligned} \mathbf{A}_m &= \frac{1}{2S_m}\begin{bmatrix} S_m - S_{m+1} & S_m + S_{m+1} \\ S_m + S_{m+1} & S_m - S_{m+1} \end{bmatrix}\begin{bmatrix} 0 & z^{-1} \\ z & 0 \end{bmatrix} \\ &= \frac{z}{1+k_m}\begin{bmatrix} 1 & k_m z^{-2} \\ k_m & z^{-2} \end{bmatrix}, \end{aligned} \tag{3}$$

where

$$k_m = \frac{S_m - S_{m+1}}{S_m + S_{m+1}} \tag{4}$$

is the reflection coefficient used in the well-known Kelly-Lochbaum scattering junction (Figure 3) and linear predictive coding (LPC). Multiplication by scattering matrix (3) can be applied repeatedly,

$$\mathbf{p}_m = \mathbf{A}_m \mathbf{p}_{m+1} = \mathbf{A}_m \mathbf{A}_{m+1} \mathbf{p}_{m+2}, \tag{5}$$

to yield the expression relating traveling wave components in the first and last section,

$$\mathbf{p}_1 = \prod_{m=1}^{N} \mathbf{A}_m \mathbf{p}_M = \frac{z^N}{\prod_{m=1}^{N}(1+k_m)} \mathbf{K}_N \mathbf{p}_M, \tag{6}$$

where the "chain" scattering matrix for $N = M-1$ junctions is

$$\mathbf{K}_N = \prod_{m=1}^{N} \begin{bmatrix} 1 & k_m z^{-2} \\ k_m & z^{-2} \end{bmatrix} = \begin{bmatrix} K_{1,1} & K_{1,2} \\ K_{2,1} & K_{2,2} \end{bmatrix}. \tag{7}$$

Polynomial entries in $\mathbf{K}_N$ are given by

$$\begin{aligned} K_{1,1} &= \sum_{m=0}^{N-1} c_{2m} z^{-2m}, \quad K_{1,2} = \sum_{m=1}^{N} d_{2(N-m)} z^{-2m}, \\ K_{2,1} &= \sum_{m=0}^{N-1} d_{2m} z^{-2m}, \quad K_{2,2} = \sum_{m=1}^{N} c_{2(N-m)} z^{-2m}, \end{aligned} \tag{8}$$

and have initial coefficients given by

$$c_0 = 1 \quad \text{and} \quad d_0 = k_1, \tag{9}$$

with remaining coefficients obtained recursively by

$$\begin{aligned} \mathbf{c}_N &= \begin{bmatrix} \mathbf{c}_{N-1} & 0 & 0 \end{bmatrix}^T + k_N \begin{bmatrix} 0 & \tilde{\mathbf{d}}_{N-1} & 0 \end{bmatrix}^T \\ \mathbf{d}_N &= \begin{bmatrix} \mathbf{d}_{N-1} & 0 & 0 \end{bmatrix}^T + k_N \begin{bmatrix} 0 & \tilde{\mathbf{c}}_{N-1} & 0 \end{bmatrix}^T, \end{aligned} \tag{10}$$

where $\tilde{\cdot}$ is used here to denote a vector in which the order of ts elements is reversed, and where the length-$2N$ coefficient (column) vectors have the form:

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

with zeros occupying every second element due to the round-trip delay of two samples in each section (an important defining structure of the piecewise-cylindrical waveguide model).

The transfer function corresponding to the waveguide model is given by the ratio of the output pressure at the mouth (the speech signal) $Y_L(z)$ to the input pressure at the glottis $X(z)$ which, following the signal flow in Figure 1, may be seen to be

$$H_L(z) = \frac{Y_L(z)}{X(z)} = \frac{T_L(z)p_M^+}{p_1^+ z - R_0 p_1^- z^{-1}} = \frac{T_L(z)p_M^+ z^{-1}}{p_1^+ - R_0 p_1^- z^{-2}}. \tag{11}$$

Employing the relationship between first and last section traveling waves given by (6) and further substituting $p_M^- = p_M^+ R_L(z)$ to allow for cancellation of all traveling wave components, yields the transfer function in its final *non-polynomial* form:

$$H_L(z) = \frac{T_L(z) z^{-(N+1)} \prod_{m=1}^{N}(1+k_m)}{K_{1,1} + K_{1,2}R_L(z) - R_0\left(K_{2,1} + K_{2,2}R_L(z)\right) z^{-2}} \tag{12}$$

Since sections here are assumed to be cylindrical, the transmission at the lips may be made amplitude complementary to the reflection which, for pressure, yields:

$$T_L(z) = 1 + R_L(z). \tag{13}$$

How (12) may be represented more conventionally as a ratio of polynomials in incremental integer powers of z^{-1} and how this relates to the all-pole filter that is estimated by LPC, is dependent on the nature of the boundaries and in particular, whether $R_L(z)$ is a scalar or a frequency-dependent loss.

3. LIP REFLECTION/TRANSMISSION AND LPC

3.1. Scalar Reflection

Consider first the simplified case where R_L is a scalar value and not a function of frequency. For this (unphysical) interpretation, the vocal tract round-trip loss can be lumped into R_0 and the reflection at the mouth simply made inverting $R_L = -1$. Because this would yield an amplitude-complementary transmission ($T_L = 1 + R_L = 0$) in which there is no signal at the output, the transmission is neglected for this case and the numerator of (12) is simply a scalar value with a pure delay given by

$$B(z) = z^{-(N+1)} \prod_{m=1}^{N} (1 + k_m) \tag{14}$$

and the denominator, by (8), is a polynomial given by

$$\begin{aligned} A(z) &= \sum_{m=0}^{N-1} c_{2m} z^{-2m} + R_L \sum_{m=1}^{N} d_{2(N-m)} z^{-2m} \\ &\quad - R_0 \left(\sum_{m=0}^{N-1} d_{2m} z^{-2m} + R_L \sum_{m=1}^{N} c_{2(N-m)} z^{-2m} \right) z^{-2} \\ &= \sum_{m=0}^{2(N+1)} a_m z^{-m} \end{aligned}$$

having coefficients given the column vector of length $2N+3$:

$$\mathbf{A}_N = \begin{bmatrix} a_0 \\ 0 \\ a_2 \\ \vdots \\ a_{2(N+1)} \end{bmatrix} = \begin{bmatrix} \mathbf{c}_N & 0 & 0 & 0 \\ \cdot & \tilde{\mathbf{d}}_N & 0 & 0 \\ \cdot & \cdot & \mathbf{d}_N & 0 \\ \cdot & \cdot & \cdot & \tilde{\mathbf{c}}_N \\ 0 & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot \\ 0 & 0 & 0 & \cdot \end{bmatrix} \mathbf{R} \tag{15}$$

where $\mathbf{R}$ is the 4×1 column vector holding scalar losses

$$\mathbf{R} = \begin{bmatrix} 1 & R_L & -R_0 & -R_0 R_L \end{bmatrix}^T = \begin{bmatrix} 1 & -1 & -R_0 & R_0 \end{bmatrix}^T. \tag{16}$$

It is worth noting that, as with the coefficient vectors of which the matrix in (15) is comprised, every second element of $\mathbf{A}_N$ is zero and that, because the final element of $\tilde{\mathbf{c}}_N$ is one, the final element of $\mathbf{A}_N$ holds the scalar loss R_0.

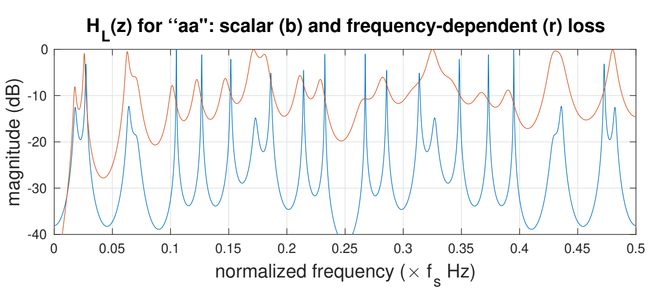

Figure 4: The magnitude response of $H_L(z)$ (blue) having scalar boundary loss ($R_0 = 0.8$ and $R_L = -1$) and $\hat{H}_L(z)$ (red) having frequency-dependent loss, where lip reflection/transmission $R_L(z)$ and $T_L(z)$ are modeled according to Figure 5. The former $H_L(z)$ shows a symmetry in the quarter-bandwidth.

The vector $\mathbf{A}_N$ corresponds to the LPC coefficients that are estimated when using a predictor such as the Levinson-Durbin recursion. In practice, an accurate LPC estimation of $A_N(z)$ from actual speech is fraught with difficulties. Nevertheless, the relationship to the waveguide model may be illustrated by omitting consideration of the glottal flow (and other contributions) and performing an order $2(N+1)$ estimation on the impulse response of $H_L(z)$, a contrived situation that would estimate $\mathbf{A}_N$ precisely [6]. Of course, like $\mathbf{A}_N$, the LPC coefficients would be interleaved with zeros and, as shown in Figure 4 for vowel "aa", produce a magnitude response that is symmetric about $f_s/4$ rather than half the sampling rate $f_s/2$. This redundancy suggests a more efficient estimation may be made (without loss of information) by halving the LPC order, with the required order corresponding to the number of sections in the waveguide model $N + 1 = M$ and/or the number of formant peaks in the spectrum. Notably however, this would also result in an assumed round-trip delay of only one sample per waveguide model section, or equivalently, the original round-trip delay of two samples but at twice the sampling rate (half the sampling period). This is significant because it implies that, given an LPC estimation, an equivalent waveguide model could be implemented at twice the sampling rate, corresponding to interleaving the LPC vector with zeros. As this would not produce an accurate physical representation (and would impede estimation of reflection coefficients and the vocal tract area function) and would not be practical for real-time use, the case of frequency-dependent loss is considered next.

3.2. Frequency-dependent reflection

Because the waveguide sections are cylindrical, the lip reflection may be described theoretically by Levine and Schwinger [9] and, though a better match is obtained using a second-order filter [10, 6], as shown in Figure 5 for lip opening corresponding to the vowel sound "aa", a good approximation (with advantages described herein) may be obtained using the first-order shelf filter:

$$R_L(z) = \frac{B_L(z)}{A_L(z)} = -\frac{(b_L)_0 + (b_L)_1 z^{-1}}{(a_L)_0 + (a_L)_1 z^{-1}}, \tag{17}$$

with coefficients, described in [11], given by

$$(b_L)_0 = \frac{\beta_0 + \rho\beta_1}{1 + \rho\alpha}, \quad (b_L)_1 = \frac{\beta_1 + \rho\beta_0}{1 + \rho\alpha}, \quad (a_L)_1 = \frac{\rho + \alpha}{1 + \rho\alpha}, \tag{18}$$

where

$$\beta_0 = \frac{(1 + g_\pi) + (1 - g_\pi)\alpha}{2}, \quad \beta_1 = \frac{(1 - g_\pi) + (1 + g_\pi)\alpha}{2} \tag{19}$$

and, if simplified for band-edge (Nyquist limit) gain $0 < g_\pi < 1$,

$$\alpha = \frac{(\sqrt{g_\pi} - 1)^2}{g_\pi - 1} = \frac{(\sqrt{g_\pi} - 1)^2}{(\sqrt{g_\pi} - 1)(\sqrt{g_\pi} + 1)} = \frac{\sqrt{g_\pi} - 1}{\sqrt{g_\pi} + 1}. \tag{20}$$

The parameter

$$\rho = \sin(f_t\pi - \pi/4)/\sin(f_t\pi + \pi/4), \tag{21}$$

is the coefficient warping the filter to the proper transition frequency $0 < f_t < 0.5$ (0.5 corresponding to the Nyquist limit). Noting that

$$\begin{aligned} 1 - (b_L)_0 &= ((1 - \beta_0) + \rho(\alpha - \beta_1))/(1 + \rho\alpha) \\ (a_L)_1 - (b_L)_1 &= (\rho(1 - \beta_0) + (\alpha - \beta_1))/(1 + \rho\alpha), \end{aligned}$$

and that applying (20) to (19) for $0 < g_\pi < 1$ yields

$$\beta_0 = \left((1 + g_\pi) - (\sqrt{g_\pi} - 1)^2\right)/2 = \sqrt{g_\pi} \tag{22}$$

and

$$\alpha - \beta_1 = -(1 - \sqrt{g_\pi}), \tag{23}$$

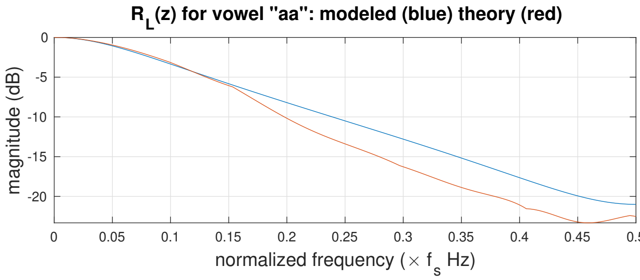

Figure 5: A shelf filter with band-edge gain $g_\pi = 0.09$ (-21 dB) and transition $f_t = 0.25$ produces a close match (at low frequencies) to the theoretical reflection of an open cylinder (with opening corresponding to that of vowel sound "aa") as given by Levine and Schwinger.

it may be seen that coefficients have the relationship

$$(a_L)_0 - (b_L)_0 = -\left((a_L)_1 - (b_L)_1\right). \tag{24}$$

Incorporating amplitude-complementary transmission (13),

$$T_L(z) = 1 + R_L(z) = (A_L(z) + B_L(z))/A_L(z), \tag{25}$$

the model transfer function (12) becomes

$$\hat{H}_L(z) = \frac{(A_L(z) + B_L(z))/A_L(z) z^{-(N+1)} \prod_{m=1}^{N}(1+k_m)}{K_{1,1} + K_{1,2}\frac{B_L(z)}{A_L(z)} - R_0\left(K_{2,1} + K_{2,2}\frac{B_L(z)}{A_L(z)}\right) z^{-2}}. \tag{26}$$

Multiplying numerator and denominator of (26) by $A_L(z)$ yields a numerator polynomial that is different from (14) and given by

$$\hat{B}(z) = (A_L(z) + B_L(z)) z^{-(N+1)} \prod_{m=1}^{N}(1+k_m) \tag{27}$$

which, when incorporating (24), may be reduced to

$$\hat{B}(z) = g(1 - z^{-1}) z^{-(N+1)}, \tag{28}$$

a product of a pure delay and a first-order high-pass filter with gain

$$g = (1 - (b_L)_0) \prod_{m=1}^{N}(1 + k_M), \tag{29}$$

showing $\hat{H}_L(z)$ is made to have zeros when the lip reflection is frequency dependent. The denominator polynomial of $\hat{H}_L(z)$ becomes

$$\begin{aligned}\hat{A}(z) = {} & K_{1,1}A_L(z) + K_{1,2}B_L(z) \\ & - R_0\left(K_{2,1}A_L(z) + K_{2,2}B_L(z)\right) z^{-2},\end{aligned} \tag{30}$$

having coefficient vector $\hat{\mathbf{A}}_N = [\hat{a}_0 \quad \hat{a}_1 \quad \ldots \quad \hat{a}_{2N+3}]^T$ given by

$$\hat{\mathbf{A}}_N = \begin{bmatrix} \mathbf{c}_N & 0 & 0 & 0 \\ \cdot & \tilde{\mathbf{d}}_N & 0 & 0 \\ \cdot & \cdot & \mathbf{d}_N & 0 \\ \cdot & \cdot & \cdot & \tilde{\mathbf{c}}_N \\ 0 & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot \\ 0 & 0 & 0 & . \\ 0 & 0 & 0 & 0 \end{bmatrix} \hat{\mathbf{R}}_0 + \begin{bmatrix} 0 & 0 & 0 & 0 \\ \mathbf{c}_N & 0 & 0 & 0 \\ \cdot & \tilde{\mathbf{d}}_N & 0 & 0 \\ \cdot & \cdot & \mathbf{d}_N & 0 \\ \cdot & \cdot & \cdot & \tilde{\mathbf{c}}_N \\ 0 & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot \\ 0 & 0 & 0 & . \end{bmatrix} \hat{\mathbf{R}}_1 \tag{31}$$

with the column vector holding reflection filter coefficients given by

$$\hat{\mathbf{R}}_n = \left[(a_L)_n \quad -(b_L)_n \quad -R_0(a_L)_n \quad R_0(b_L)_n\right]^T. \tag{32}$$

Just as the scalar loss R_0 may be estimated as the last element of $\mathbf{A}_N$, so may the reflection filter coefficients be estimated from $\hat{\mathbf{A}}_N$, albeit with a little more effort. Since the first coefficient of $\mathbf{c}_N$ and $\tilde{\mathbf{d}}_N$ is one and zero, respectively, it can be seen from (31) and (32) that coefficient $(a_L)_1$ is simply the second element of $\hat{\mathbf{A}}_N$ so that, with lip reflection filter coefficients in (17) defined by

$$\mathbf{b}_L = \left[-(b_L)_0 \quad -(b_L)_1\right]^T \quad \text{and} \quad \mathbf{a}_L = \left[1 \quad (a_L)_1\right]^T, \tag{33}$$

denominator coefficients may be estimated as first and second elements of $\hat{\mathbf{A}}_N$,

$$\hat{\mathbf{a}}_L = \left[1 \quad \hat{a}_1\right]^T. \tag{34}$$

Further, it may be shown algebraically that, given by (24), the numerator coefficients may be estimated from the first and last four elements of $\hat{\mathbf{A}}_N$: $\hat{a}_0, \hat{a}_1, \hat{a}_2, \hat{a}_3$ and $\hat{a}_{2N}, \hat{a}_{2N+1}, \hat{a}_{2N+2}, \hat{a}_{2N+3}$, respectively, to yield

$$\hat{\mathbf{b}}_L = \left[E - \sqrt{E^2 + D\hat{a}_1} \quad E + \sqrt{E^2 + D\hat{a}_1}\ \right]^T \tag{35}$$

where

$$E = -\frac{1 + \hat{a}_1}{2} \quad \text{and} \quad D = -\frac{\frac{\hat{a}_3}{\hat{a}_1} - \hat{a}_2}{\frac{\hat{a}_{2N}}{\hat{a}_{2N+2}} - \frac{\hat{a}_{2N+1}}{\hat{a}_{2N+3}}}. \tag{36}$$

Finally, the scalar loss at the glottis is be given by

$$R_0 = -\frac{\hat{a}_{2N+3}}{(\hat{\mathbf{b}}_L)_2} = -\frac{\hat{a}_{2N+2}}{(\hat{\mathbf{b}}_L)_1}, \tag{37}$$

showing that boundary losses may be estimated from $\hat{A}(z)$, independent of the vocal tract shape. Though there remains the problem of actually having an accurate $\hat{A}(z)$ (since LPC would estimate a much higher order filter as it tries to fit an all-pole filter to one having zeros), it is suggested that first deconvolving the simple high-pass filter in (28) from the speech signal before the LPC estimation is a step toward that goal.

4. CONCLUSIONS

In this work, a matrix formulation of a piecewise cylindrical waveguide model of the vocal tract is revisited to produce a transfer function that is dependent on both the varying cross-sectional area along the vocal tract length and boundary losses due to reflection at the glottis and the lip. Boundary losses are considered for two cases, one where they are scalar, producing an all-pole transfer function with a strong relationship to LPC (discussed herein) and the other where the lip reflection is made frequency-dependent, a more physical interpretation that has the effect of introducing zeros to the transfer function and creating a relationship to LPC that is less clear (since LPC is known to estimate an all-pole filter). Nevertheless, it is shown that if the lip reflection loss is assumed to have the properties of a first-order shelf filter, the numerator of the vocal tract can be reduced to a product of a pure delay and simple high-pass FIR filter with a gain which can be more easily deconvolved from the transfer function to yield the denominator polynomial (and coefficients) from which lip reflection filter coefficients and (scalar) losses at the glottis may be estimated.

5. REFERENCES

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

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

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

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

[5] 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.

[6] T. Smyth and D. Zurale, "On the transfer function of the piecewise cylindrical vocal tract model," in *Proceedings of the* 18^{th} *Sound and Music Computing Conference*, online, June 2021, virtual conference.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] J. Abel, T. Smyth, and J. O. Smith, "A simple, accurate wall loss filter for acoustic tubes," in *DAFX 2003 Proceedings*. London, UK: International Conference on Digital Audio Effects, September 2003, pp. 53–57.