Direct computational method of including piriform fossae and nasal cavity in a time-domain acoustic model of the vocal tract

Parham Mokhtari *
ATR Cognitive Information Science Laboratories, and NiCT Universal Media Research Center, Kyoto JAPAN.

This research was supported by the National Institute of Information and Communications Technology (NICT), and by SCOPE-R of Japan.

3a. Electrical Circuit Analog of the Vocal Tract

3b. Equivalent impedances \( b_n \), \( H_n \), and \( H_{\text{INT}} \)

6. Single-Matrix Formulation Extended to 3 side-branches

The system matrix is **sparse, diagonally-dominant** and symmetric. Its size is \((M+1) \times (N+1) \times (WW+1) \times (WR+1) \rightarrow 94 \times 94\)

7. Some Notes on Implementation

- The system of linear equations was solved at every sampling instant with the LAPACK (Linear Algebra PACKage) routine \( \text{"dgesv"} \)… a general algorithm for Gaussian elimination using double-precision.
- To attain reasonable accuracy up to about 6 kHz, the simulation was oversampled by 4 times, at \( f_{\text{sim}} = 48 \) kHz (output signals were then downsampled to \( f_s = 12 \) kHz).
- The non-optimized C program runs (on a 1.8 GHz laptop PC) on the order of 10 times real-time. About half of the processing time is used by the \( \text{dgesv} \) routine. Faster speeds may be attained by:
  - optimizing the C code
  - using a more efficient (specialized) linear-algebra routine
  - reducing the simulation sampling frequency \( f_{\text{sim}} \)
  - reducing the number of VT sections (including side-branches)

Please see Kitamura et al. (poster 1pSC5) for details on a speech synthesis system based on the present work.
1. Background & Motivations

Flanagan et al. (1968,...,1975) and Maeda (1982) described a time-domain simulation of 1-D acoustic wave propagation in the vocal tract. For numerical simulation, the wave equations relating acoustic pressure and volume–velocity were discretized in space (finite number of VT sections) and in time (digital sampling). This led to a set of linear finite-difference equations that must be solved at every sampling instant. However, existing formulations of these time-domain methods apparently do not accommodate any more than one VT side-branch (usually configured as the nasal tract). On the other hand, there is increasing evidence (e.g., Honda et al., 2004; Takemoto et al., 2006) of the acoustic importance of the detailed morphology of the hypopharynx, which includes the piriform fossae – small bilateral cavities that are best modelled as side-branches to the main VT (e.g., Dang & Honda, 1997).


Four steps at every simulation sampling instant...

Step 1: Update the global area and VT area functions, then compute all circuit elements (as shown in the previous panel).

Step 2: From the circuit elements, compute the following variables:
- Equivalent impedances $Z_n$ and $R_n$:

$$
Z_n = 2/(\omega n) \quad \text{and} \quad R_n = 1/(2\omega n) \quad \text{where} \quad \omega = 2\pi f
$$

- Equivalent pressure (force) terms $F_n$:

$$
F_n = (\text{Term 1}) + (\text{Term 2}) \quad \text{where} \quad (\text{Term 1}) = (1-\text{Term 2}) \quad \text{and} \quad \text{Term 2} = \text{Term 3} \quad \text{Term 4} \quad \text{Term 5}
$$

Step 3: Separately solve the following 3 sets of linear equations, to get the right hand-side vectors:

$$
\begin{align*}
A & = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} \\
B & = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} \\
C & = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{bmatrix}
\end{align*}
$$

Step 4: Compute the remaining variables (see text below):

$$
\begin{align*}
P &= a_{11}c_1 + a_{12}c_2 + a_{13}c_3 + a_{14}c_4 \\
Q &= a_{21}c_1 + a_{22}c_2 + a_{23}c_3 + a_{24}c_4 \\
R &= a_{31}c_1 + a_{32}c_2 + a_{33}c_3 + a_{34}c_4 \\
S &= a_{41}c_1 + a_{42}c_2 + a_{43}c_3 + a_{44}c_4
\end{align*}
$$

8. Acoustic Validation

- $A_p$: step-down function simulating a sudden glottal closure ($f_{\text{on}} = 48$ kHz).
- Area Functions: from MRI data of an adult male Japanese.
  - VT: 44 equal-length sections (about 0.4cm each)
  - NT: 38 equal-length sections (0.3cm each) [open glabellar area = 0.01cm²]
  - PFL & PFR: 4 equal-length sections (left: 0.52cm, right: 0.47cm)

9. Conclusions

- We extended Maeda’s (1982) time-domain method of vocal tract acoustic simulation. The new formulation simulates a VT with 3 side-branches: the nasal tract, and the left and right piriform fossae. This method can be easily extended further to allow any number of side-branches, such as the sub-lingual cavity, inter-dental spaces, and nasal sinuses.
- The resulting system matrix is much larger than in Maeda’s method: its size is proportional to the total number of sections in the VT and all side-branches. However, in simulations of both sustained and dynamic utterances including vowels and consonants, we have not encountered any problems of instability. Stability is probably helped by the fact that the system matrix is always symmetric, sparse and diagonally-dominant.
- In future work, we would like to: allow for frequency-dependent losses, improve the source ($A_p$) modelling, and increase the overall naturalness of the synthesizer to include personal characteristics.

5. New Formulation with a Single Matrix

(only 1 side-branch)

Expand Maeda’s 3 matrix equations:

(a) $F_n = H_n E_n + h_n F_{n+1}$
(b) $F_{n+1} = h_n E_n + h_{n+1} F_{n+2}$
(c) $F_{n+2} = h_{n+1} E_n + h_{n+2} F_{n+1}$
(d) $E_n = h_n F_n + h_{n+1} E_{n+1}$
(e) $E_{n+1} = h_{n+1} F_n + h_{n+2} E_{n+2}$
(f) $E_{n+2} = h_{n+2} F_n + h_{n+3} E_{n+3}$

Rewrite the new set of equations in single-matrix form:

$$
\begin{align*}
F_n &= \begin{bmatrix} F_{n-1} \\ F_n \\ F_{n+1} \\ F_{n+2} \end{bmatrix} \\
E_n &= \begin{bmatrix} E_{n-1} \\ E_n \\ E_{n+1} \\ E_{n+2} \end{bmatrix}
\end{align*}
$$

The 4 off-diagonal elements and other terms (circled) account for the VT-NT junction.