5pSC6

4th Joint Meeting of the Acoustical Society of America & the Acoustical Society of Japan (ASA/J)
Honolulu, Hawaii, 28 Nov – 2 Dec, 2006

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

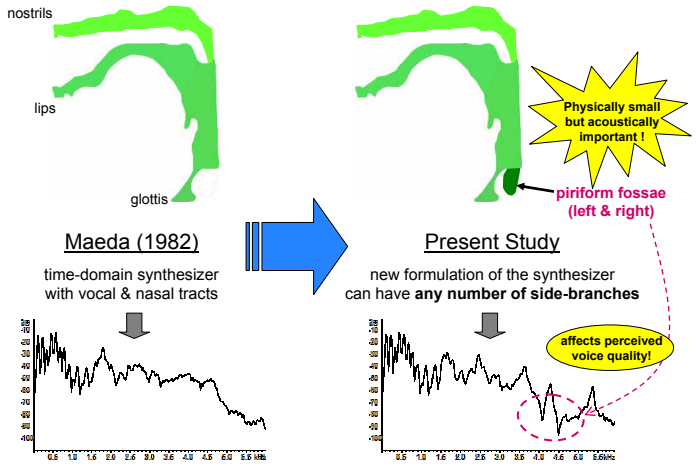
Parham Mokhtari *

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

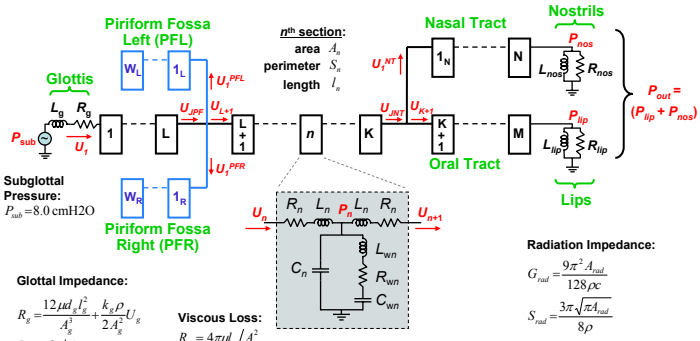
* parham@atr.jp

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

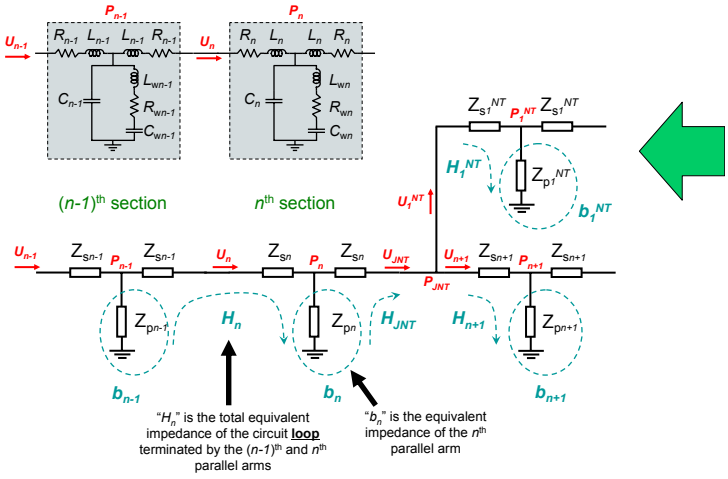
Basic Idea: to allow the flexibility of adding VT sub-branches in a time-domain articulatory synthesizer



3a. Electrical Circuit Analog of the Vocal Tract

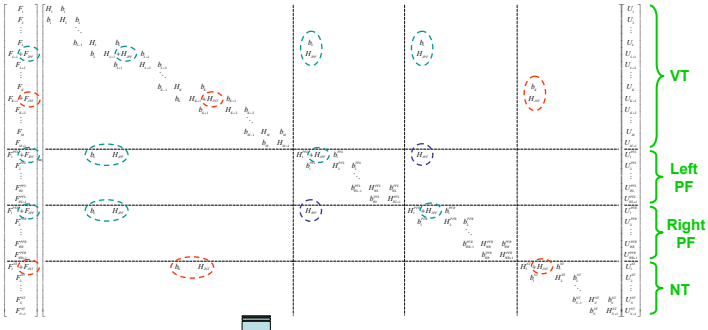


3b. Equivalent impedances b_n, H_n, and H_{JNT}



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

The same mathematical steps applied to the VT with 3 side-branches (NT & PFL & PFR) lead to the following single-matrix equation:



The system matrix is **sparse**, **diagonally-dominant** and **symmetric**.
Its size is (M+1) + (N+1) + (WL+1) + (WR+1) ➔ **94 x 94**

The extra terms (circled) account for:
------ VT-PF junction
------ VT-NT junction

7. Some Notes on Implementation

- The system of linear equations was solved at every sampling instant with the **LAPACK** (Linear Algebra PACKage) routine "**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_{sim}=48 kHz (output signals were then downsampled to f_s=12 kHz).
- The non-optimized C program runs (on a 1.8 GHz notebook PC) on the order of 10 times real-time. About **half** of the processing time is used by the **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_{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).

2. Aim

Reformulate the well-known VT acoustic model of Maeda (1982) to allow simulation with **any number of side-branches**.

The present work shows how the existing model that includes the main VT + nasal tract can be augmented with both the **left and right piriform fossae**. This poster summarizes the mathematical reformulation, and validates it with some examples of acoustic simulation with and without the extra side-branches.

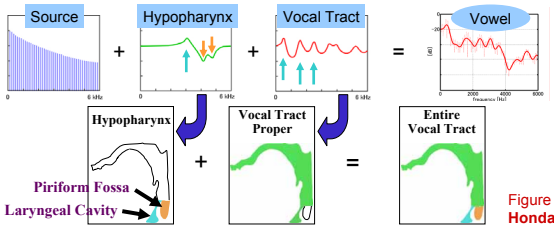


Figure taken from **Honda *et al*. (2004)**

4. Maeda's (1982) method

Four steps at every simulation sampling instant…

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

Step 2: From the circuit elements, compute the following variables:

Equivalent impedances "b_n" and "H_n":

$$b_n = \frac{1}{Y_{wn} + 2f_{sim}C_n}, \quad \text{where: } Y_{wn} = \frac{1}{2f_{sim}L_{wn} + R_{wn} + \frac{f_{sim}}{2C_{wn}}}$$

$$H_n = -2f_{sim}(L_{n-1} + L_n) - (R_{n-1} + R_n) - (b_{n-1} + b_n)$$

$$b_{M+1} = \frac{1}{G_{lip} + S_{lip}}, \quad b_{N+1} = \frac{1}{G_{nos} + S_{nos}}$$

Equivalent pressure (force) terms "F_n":

$$F_n[t] = b_n V_n[t-1] - b_{n-1} V_{n-1}[t-1] - Q_n[t-1] \text{ , where:}$$

$$Q_n[t-1] = 4f_{sim}U_n[t-1](L_{n-1}[t-1] + L_n[t-1]) - Q_n[t-2]$$

$$V_n[t-1] = Vc_n[t-1] - (Qwl_n[t-1] - Qwc_n[t-1])Y_{wn}$$

$$Vc_n[t-1] = 4f_{sim}C_n[t-1]P_n[t-1] - Vc_n[t-2]$$

$$Qwl_n[t-1] = 4f_{sim}L_{wn}[t-1]u_{wn}[t-1] - Qwl_n[t-2]$$

$$Qwc_n[t-1] = \frac{u_{wn}[t-1]}{f_{sim}C_{wn}[t-1]} + Qwc_n[t-2]$$

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

$$\begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_K \\ F_{JNT} \end{bmatrix} = \begin{bmatrix} H_1 & b_1 & & & \\ b_1 & H_2 & b_2 & & \\ & \ddots & \ddots & \ddots & \\ & & b_{K-1} & H_K & b_K \\ & & & b_K & H_{JNT} \end{bmatrix} - 1 \begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_K \\ U_{JNT} \\ P_{JNT} \end{bmatrix}$$

$$\begin{bmatrix} F_{K+1} \\ F_{K+2} \\ \vdots \\ F_M \\ F_{M+1} \end{bmatrix} = \begin{bmatrix} 1 & H_{K+1} & b_{K+1} & & \\ & b_{K+1} & H_{K+2} & b_{K+2} & \\ & & \ddots & \ddots & \ddots \\ & & & b_{M-1} & H_M & b_M \\ & & & & b_M & H_{M+1} \end{bmatrix} \begin{bmatrix} P_{JNT} \\ U_{K+1} \\ U_{K+2} \\ \vdots \\ U_M \\ U_{M+1} \end{bmatrix}$$

$$\begin{bmatrix} F_1^{NT} \\ F_2^{NT} \\ \vdots \\ F_N^{NT} \\ F_{N+1}^{NT} \end{bmatrix} = \begin{bmatrix} 1 & H_1^{NT} & b_1^{NT} & & \\ & b_1^{NT} & H_2^{NT} & b_2^{NT} & \\ & & \ddots & \ddots & \ddots \\ & & & b_{N-1}^{NT} & H_N^{NT} & b_N^{NT} \\ & & & & b_N^{NT} & H_{N+1}^{NT} \end{bmatrix} \begin{bmatrix} P_{JNT} \\ U_1^{NT} \\ U_2^{NT} \\ \vdots \\ U_N^{NT} \\ U_{N+1}^{NT} \end{bmatrix}$$

Step 4: Compute the remaining variables (then go to Step 1):

$$P_n = b_n(U_n - U_{n-1} + V_n[t-1])$$

$$u_{wn}[t] = Y_{wn}(P_n + Qwl_n[t-1] - Qwc_n[t-1])$$

5. New Formulation with a Single Matrix

(only 1 side-branch)

Expand Maeda's 3 matrix equations:

(a) $F_1 = H_1 U_1 + b_1 U_2$

(b) $F_n = b_{n-1}U_{n-1} + H_n U_n + b_n U_{n+1}, \quad n = \{2,\ldots,K-1\}$

(c) $F_K = b_{K-1}U_{K-1} + H_K U_K + b_K U_{JNT}$

(d) $F_{JNT} = b_K U_K + H_{JNT} U_{JNT} - P_{JNT}$

(e) $F_{K+1} = P_{JNT} + H_{K+1}U_{K+1} + b_{K+1}U_{K+2}$

(f) $F_n = b_{n-1}U_{n-1} + H_n U_n + b_n U_{n+1}, \quad n = \{K+2,\ldots,M\}$

(g) $F_{M+1} = b_M U_M + H_{M+1}U_{M+1}$

(h) $F_1^{NT} = P_{JNT} + H_1^{NT}U_1^{NT} + b_1^{NT}U_2^{NT}$

(i) $F_n^{NT} = b_{n-1}^{NT}U_{n-1}^{NT} + H_n^{NT}U_n^{NT} + b_n^{NT}U_{n+1}^{NT}, \quad n = \{2,\ldots,N\}$

(j) $F_{N+1}^{NT} = b_N^{NT}U_N^{NT} + H_{N+1}^{NT}U_{N+1}^{NT}$

Constraint at NT junction: $U_{JNT} = U_{K+1} + U_1^{NT}$

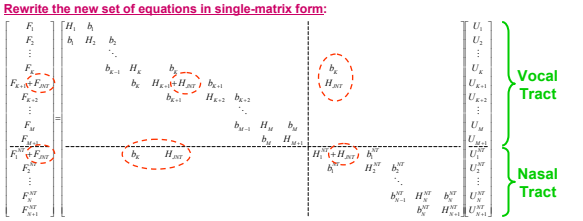
eliminate U_{JNT} → new Eqn. (c): $F_K = b_{K-1}U_{K-1} + H_K U_K + b_K U_{K+1} + b_K U_1^{NT}$

eliminate U_{JNT} → $P_{JNT} = b_K U_K + H_{JNT}U_{K+1} + H_{JNT}U_1^{NT} - F_{JNT}$

eliminate P_{JNT} → new Eqn. (e): $F_{K+1} + F_{JNT} = b_K U_K + (H_{JNT} + H_{K+1})U_{K+1} + b_{K+1}U_{K+2} + H_{JNT}U_1^{NT}$

eliminate P_{JNT} → new Eqn. (h): $F_1^{NT} + F_{JNT} = b_K U_K + H_{JNT}U_{K+1} + (H_{JNT} + H_1^{NT})U_1^{NT} + b_1^{NT}U_2^{NT}$

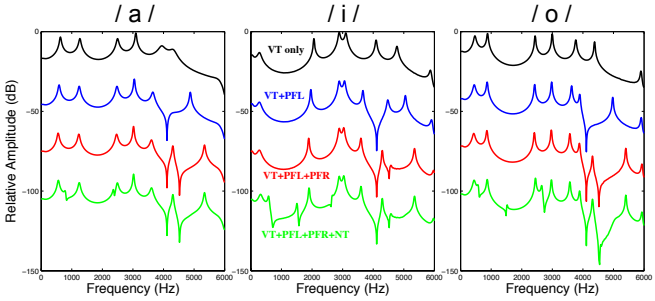
Rewrite the new set of equations in single-matrix form:



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

8. Acoustic Validation

- A_g: step-down function simulating a sudden glottal closure (f_{sim} = 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 velopharyngeal port = 0.01cm^2]
  - 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. The 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_g) modelling, and increase the overall *naturalness* of the synthesizer to include *personal characteristics*.