CALCULATION OF ACOUSTIC FREQUENCY RESPONSE OF VOCAL TRACT
CONTENTS

I. Introduction ................................................................................3

II. Speech production system ..........................................................4

III. Modeling sound as a propagating wave ......................................6

IV. Transfer function calculation for single tube .........................8
   a. Effect of losses ......................................................................11
   b. Effect of wall-vibration .........................................................14
   c. Effect of lip radiation ............................................................16
   d. Effect of glottal impedance ...................................................17

V. Transfer function calculation for multi-tube .........................19

VI. Transfer function calculation for fricatives and plosives ..........26

VII. Transfer function calculation for nasals .................................30

VIII. Feature work ........................................................................34

IX. References .............................................................................35
I- INTRODUCTION

This report is about the work done in the period December 2006 - May 2007. In our previous work we have analyzed the continuity of the visible resonance frequencies. It is observed that these frequencies depend on the basic structure of the model used for vocal tract. The well-known fact about resonance frequencies of vocal tract is that they are related to physical structure vocal tract. Hence, the reliable knowledge about VTRs (Vocal Tract Resonances) can be obtained by a detailed examination of the physical geometry of vocal tract. Therefore, this term we concentrate on modeling VT and finding resonance of acoustic system and comparing them with resonance obtained from the speech signal.

This report is organized as follows:

The second section of this report summarizes the speech production mechanism. The third section is related the modeling sound as a propagating wave. In section four and five calculation transfer function for single and multi-tube model are explained respectively. Section six and seven explain calculation transfer function for fricative and nasal sound respectively. The future work can be seen in section eight.
II- SPEECH PRODUCTION SYSTEM

Basic anatomy and physiology of speech

The speech production system contains mainly two parts: sub-glottal and supra-glottal system [1], which can be seen Figure-1.

Subglottal system: It is the system below the glottis, and consists of the organs responsible for providing and transporting the excitation energy of the phonological system, namely, lungs, bronchi, trachea, diaphragm and some other muscles. The lungs represent the air source, and their dynamics is controlled by the diaphragm.

Supraglottal system: These are the articulators and airways above the larynx. They will alter the spectrum of the forward signal, transforming it to an intended or target spectral distribution, according to the linguistic message that is to be transmitted. In short, palate, velum, jaw, tongue, teeth and lips, can modulate the excitation and produce sound which across the vocal and nasal tracts.

Figure-1 Speech production mechanism
Source–Filter model of speech production

The simple model for speech production is the source–filter model, which is based on an assumption of production system dynamics being linear and, specifically, separable into three main blocks: a glottal energy (source), the vocal tract (filter) and the effect of modeling radiation sound [5]. The glottal source roughly matches the subglottal systems previously presented, while vocal tract (VT) corresponds to the supra-glottal system. The radiation block can be considered as a converter, which converts volume velocity into acoustic pressure.

\[
S(\Omega) = U(\Omega) H(\Omega) R(\Omega)
\]

\(U(\Omega)\) represents excitation
\(H(\Omega)\) is the dynamic of the vocal tract filter
\(R(\Omega)\) models radiation effect

In Acoustic modeling the excitation \(u(t)\) is taken to be volume velocity or flow of sound. \(H(\Omega)\) is the ratio of the volume velocity at the output of the vocal system to that at the input to the tract the speech signal is usually
considered to be sound pressure at the output of the vocal system. \( H (\Omega) \) does not account for the flow-to-pressure conversion (radiation) function at the lip boundary, which is included in \( R (\Omega) \).

III- MODELING SOUND AS A PROPAGATING WAVE

The wave propagation in the vocal-tract is based on some laws of physics. These laws which describe the generation and propagation of sound in the vocal system are the fundamental laws of conservation of mass, conservation of momentum conservation of energy, the laws of thermodynamics and fluid mechanics.[9]

The complete acoustic theory of speech production must consider the following effects: 1-) excitation of sound in the vocal tract 2-) variation in time of the shape of vocal tract 3-) nasal coupling 4-) radiation of sound at the lips and nostrils 5-) losses due to viscous friction and heat conduction 6-) softness and vibration of the vocal and nasal tract walls. Using the laws governing the generation and propagation of sound in the vocal system a set of partial differential equations can be derived. Taking into account all the effects which appear in the vocal system the formulation and solutions of this set of differential equations is difficult to solve analytically. Therefore some simple assumptions have to be taken. For a simple configuration of the vocal system the vocal tract is modeled as a lossless uniform tube. Figure-3 denotes this model

![Figure-3](image-url)

Figure-3 Simple vocal tract confutation and single tube model
Single lossless Tube Analysis

Equations relating pressure and velocity of air particles in a lossless tube are [9]

\[-\frac{\partial u(x,t)}{\partial x} = \frac{A}{\rho c^2} \frac{\partial p(x,t)}{\partial t}\]  Conservation of mass

\[-\frac{\partial p(x,t)}{\partial x} = \frac{\rho}{A} \frac{\partial u(x,t)}{\partial t}\]  Conservation of momentum

Combining these equations are also known as one-dimensional wave equations.

where,
\(u(x,t)\) : volume velocity
\(p(x,t)\) : pressure
\(A\) : cross-sectional area of tube
\(c\) : speed of sound in air
\(\rho\) : density of air

From this equations an analogy is established with electrical wave propagation along transmission lines, where for a voltage \(v(x,t)\) and current \(i(x,t)\), the following relations hold [2]

\[-\frac{\partial i(x,t)}{\partial x} = C \frac{\partial v(x,t)}{\partial t}\]

\[-\frac{\partial v(x,t)}{\partial x} = L \frac{\partial i(x,t)}{\partial t}\]

where,
\(C \equiv \frac{A}{\rho c^2}\), \(L \equiv \frac{\rho}{A}\)

\(i(x,t)\) : current, \(v(x,t)\) : voltage,
\(C\) : acoustic capacitance per unit length of transmission line.
\(L\) : acoustic inductance per unit length of transmission line.

Transmission line model of lossless tube model can be seen in Figure-4
IV- ACOUSTIC TRANSFER FUNCTION CALCULATION FOR SINGLE TUBE

The transfer function of the single tube model can be found as a ratio of the output volume velocity to input volume velocity using some boundary conditions. In this part of the report we give direct result of the network representation of the acoustic tube model which is based on solution of one-dimensional wave equation to find transfer function of the acoustic system.

The well-known input-output characteristics of a four-terminal network are described by a matrix equation of the form [3]

\[
\begin{bmatrix}
    P_g(\Omega) \\
    U_g(\Omega)
\end{bmatrix}
= T(\Omega)
\begin{bmatrix}
    P_r(\Omega) \\
    U_r(\Omega)
\end{bmatrix}
\]

where,
- \( U_g(\Omega) \) : Fourier transform of volume velocity at glottis
- \( U_r(\Omega) \) : Fourier transform of volume velocity at radiation load
- \( P_g(\Omega) \) : Fourier transform of pressure at glottis
- \( P_r(\Omega) \) : Fourier transform of pressure at radiation load
\[ T(\Omega) = \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} 1 + \frac{Z_a}{Z_b} & 2Z_a + \frac{Z^2}{Z_b} \\ \frac{1}{Z_b} & 1 + \frac{Z_a}{Z_b} \end{bmatrix} \]

\[ Z_a = Z_0 \tanh\left(\frac{\gamma l}{2}\right) \text{ and } Z_b = \frac{Z_0}{\sinh(\gamma l)} \]

\[ \gamma = \sqrt{zy}, \ z = j\Omega L \text{ and } y = j\Omega C \]

Figure-5 denotes complete form of four terminal transmission networks. Using these parameters we can easily calculate transfer function of the system.

\[ H(\Omega) = \frac{U_r(\Omega)}{U_i(\Omega)} = \frac{1}{CZ_i + D} \]

and the input impedance \( Z_{IN} \)

\[ Z_{IN}(\Omega) = \frac{P_r(\Omega)}{U_i(\Omega)} = \frac{AZ_i + B}{CZ_i + D} \]

Figure-5 Two-port Network model
**Example:** As an example we can find resonance frequencies of the lossless single acoustic tube which is very rough representation vocal tract. Assuming $Z_R$ is equal to zero (that means there is no radiation at output). The resonance frequencies of the system can be found by denominator of the transfer function or input impedance $Z_{IN}$ by equating to zero

$$CZ_i + D = 0$$

Since we assumed that $Z_i$ is equal to zero, The resonance frequencies are obtained solution of following equation

$$D(\Omega) = 1 + \frac{Z_a}{Z_b} = 0$$

When we substitute $Z_a$ and $Z_b$ in equation above

We can find

$$D(\Omega) = \cos\left(\frac{\Omega \ell}{c}\right) = 0$$

$$\frac{\Omega \ell}{c} = \frac{\pi}{2} (2i - 1) \text{ for } i=1,2,3...$$

$$F_i = \frac{c}{4\ell} (2i - 1) \text{ for } i=1,2,3...$$

If we assume that the length of the vocal tract length $l=17$ cm and speed of sound $c=340$ m/s. The first four resonance frequencies can be found as; $F_1 = 500$Hz, $F_2 = 1500$Hz, $F_3 = 2500$Hz, $F_4 = 3500$Hz. The transfer function of the system can be seen in Figure-6
Previous section we obtained the transfer function of the acoustic system for lossless case. However, there exist many losses, which are heat, viscous losses and losses due to vibration at vocal tract. Hence in this section we model these losses and examine effect on transfer function. [3]
Heat and viscous losses effect on acoustic transfer function

The same four-terminal network can also be used to examine losses effect.

\[
\begin{bmatrix}
    P_e(\Omega) \\
    U_e(\Omega)
\end{bmatrix} = T(\Omega) \begin{bmatrix}
    P_i(\Omega) \\
    U_i(\Omega)
\end{bmatrix}
\]

where,

\[
T(\Omega) = \begin{bmatrix}
    A & B \\
    C & D
\end{bmatrix} = \begin{bmatrix}
    1 + \frac{Z_a}{Z_b} & 2Z_a + \frac{Z_a^2}{Z_b} \\
    \frac{1}{Z_b} & 1 + \frac{Z_a}{Z_b}
\end{bmatrix}
\]

The only difference is the calculation of the impedance \(Z_a\) and \(Z_b\). Some additional terms, which are \(R\), \(G\) for heat and viscous loss respectively, are needed. Transmission line model of losy system can be seen in Figure-6.

![Diagram of vocal tract configuration, single cylindrical tube approximation, and electrical analogy](image)

Figure-6 Single tube representation and its transmission line equivalent (Losy case)

\[
Z_a = Z_0 \tanh\left(\frac{\gamma L}{2}\right) \quad \text{and} \quad Z_b = \frac{Z_0}{\sinh(\gamma L)}
\]

\[
\gamma = \sqrt{zy}, \quad z = R + j\Omega L \quad \text{and} \quad y = G + j\Omega C
\]

where,

\[
R = \frac{S\sqrt{\rho\mu\Omega}}{2\sqrt{2A^2}} \quad \text{and} \quad G = \frac{(\eta - 1)S}{\rho c^2} \sqrt{\frac{\lambda\Omega}{2\zeta\rho}} \quad \text{where} \quad S = 2\sqrt{A\pi}
\]
where,

\[ A : \text{cross-sectional area of tube} \]
\[ c : \text{speed of sound in air} \]
\[ \rho : \text{density of air} \]
\[ \mu : \text{viscosity} \]
\[ \xi : \text{specific heat} \]
\[ \eta : \text{adiabatic gas constant} \]
\[ \lambda : \text{heat conduction of air} \]
\[ \Omega : \text{radian frequency} \]

The Transfer function of the system can be calculated similarly as follows

\[
H(\Omega) = \frac{U_i(\Omega)}{U_p(\Omega)} = \frac{1}{CZ_r + D}
\]

The effect of the losses in vocal tract can be seen in Figure-7. From the figure we observe that band-widths of the resonances frequencies are increased and the effect is not homogenous for all frequencies (The effect is more at high frequencies than low frequencies since the value of R and G are frequency dependent quantities)
b. VIBRATION-WALL (ELASTIC WALL) EFFECT

On the contrary to rigid wall assumption in vocal tract, there is a vibration in vocal tract wall as air pressure is varied inside the vocal tract. [1] This type of vibration is modeled as an second order system which is introduce additional terms (impedance $Z_w$) in transmission line model which can be seen in Figure -8.

Figure-7 Frequency response for single-losy–rigid tube

Figure-8 Figure-6 Single tube representation and its transmission line equivalent (Losy and vibration wall case)
Therefore the element of the transmission matrix $T(\Omega)$ can be calculated as follows [3]

$$Z_a = Z_0 \tanh\left(\frac{\gamma L}{2}\right) \quad \text{and} \quad Z_0 = \frac{Z_0}{\sinh(\gamma L)}$$

$$\gamma = \sqrt{zy}, \quad z = R + j\Omega L \quad \text{and} \quad y = G + j\Omega C + \frac{1}{Z_w}$$

$$Z_w = R_w + j\Omega L_w + \frac{1}{j\Omega C_w}$$

$$R_w = \frac{b}{S^2}, \quad L_w = \frac{m}{S^2}, \quad C_w = \frac{S^2}{k} \quad \text{and} \quad S = 2\sqrt{A_\pi}$$

- $b$: damping per unit length of the vocal tract
- $k$: stiffness per unit length of the vocal tract
- $m$: mass per unit length of the vocal tract

The transfer function of the system can be calculated same manner. Figure-9 denotes the effect of wall-vibration on transfer function of the system. First of all there is a high-pass effect on frequency response and resonance frequencies are shifted. Moreover, band-width of the resonance frequencies are more effected at low frequencies.
Figure-9  Frequency response for single-lossless–vibration wall tube

c. RADIATION LOAD EFFECT

The vocal tract tube terminates with the opening between lips. The sound waves spread real world. In order to model this effect, radiation impedance, $Z_r$ is used, which can be seen in Figure-5. The radiation impedance is defined as:[1]

$$Z_r = R_r + j\Omega L_r$$

$$R_r = \frac{128\rho c}{9\pi^2 A_m}, \quad L_r = \frac{8\rho}{3\pi\sqrt{\pi A_m}}$$

$A_m$, $c$ and $\rho$ are the Lip opening area, speed of sound and air density respectively.
The radiation effect for a single tube can be seen in Figure-10. There is a small decrease in location resonance frequencies and also there is an increase in band-width of the resonance frequencies.

![Frequency Response](image)

Figure-10  Frequency response for single-lossless–rigid-wall tube with radiation load

d.  **GLOTTAL IMPEDANCE EFFECT**

In general, the glottal impedance is considered as infinite value. However, in reality glottal impedance is time varying impedance its value changes according to glottal opening area. Glottal impedance is equal to infinite only when the glottis is closed. Average glottal impedance can be defined as fallows [1]
\[ Z_g = R_g + j\omega L_g \]
\[ R_g = \frac{12\mu dl^2}{A_g^3} + \frac{0.875\rho u_g}{2A_g^2}, \quad L_g = \frac{\rho d}{A_g} \]

\( A_g, l_g, \mu, d \) and \( \rho \) are the glottal opening area, glottis length, air viscosity, glottal depth (thickness), and air density respectively. The effect of glottal impedance can be seen in Figure-11. From figure, we see that there is an increase in resonance frequencies and their bandwidth and also lower frequencies more effected than higher frequencies.

**Figure-11** Frequency response for single-lossless–rigid-wall tube with glottal impedance
V- ACOUSTIC TRANSFER FUNCTION CALCULATION FOR MULTI-TUBE

It is clear that single tube model of the vocal tract is rough representation. Fine representation can be achieved by spatially discretizing vocal area function to model multi tube representation as can be seen in Figure-12 [9]

![Multi-tube representation of vocal tract](image)

Figure-12 Multi-tube representation of vocal tract

The transfer function the multi tube representation can be also found by four terminal networks model.

![Two-port network representation for multi-tube](image)

Figure-13 Two-port network representation for multi-tube
\[
\begin{bmatrix}
P_g(\Omega) \\
U_g(\Omega)
\end{bmatrix} = T_r(\Omega) \ldots T_N(\Omega)
\begin{bmatrix}
P_r(\Omega) \\
U_r(\Omega)
\end{bmatrix}
\]

Let \( T_F(\Omega) \) be a over all transmission matrix, we can calculate \( T_F(\Omega) \) as fallows [3]

\[
T_r(\Omega) = \begin{bmatrix}
A_f & B_f \\
C_f & D_f
\end{bmatrix} = T_r(\Omega) \ldots T_N(\Omega)
\]

\[
H(\Omega) = \frac{U_r(\Omega)}{U_g(\Omega)} = \frac{1}{C_f Z_r + D_f}
\]

and the input impedance \( Z_{IN} \)

\[
Z_{IN}(\Omega) = \frac{P_g(\Omega)}{U_g(\Omega)} = \frac{A_f Z_r + B_f}{C_f Z_r + D_f}
\]

**Example: Two-tube model**

In this section we shows some example to model some vowel sound with \( N=2 \) tube. Figure-14, Figure-15 and Figure-16 denotes modeling vowel u, a and i with two tube respectively. In these figure there are some speech spectrogram of corresponding vowel for comparison of resonance location purpose.
Figure-14 The modeling phoneme -u-

Two-tube representation

Figure-15 The modeling phoneme -a-
Two-tube representation

F1=249, F2=1850, F3=2833, F4=3817

Figure-16 The modeling phoneme -i-

Example: N=44 Tube each tube is same length l=0.396cm

In this section we shows some example to model some vowel sound with N=44 tube. Figure-17, Figure-18 and Figure-19 denotes modeling vowel a, i and o with forty-four tube respectively. The areas values of vowel sound are taken from MR image [6].
Area Function

Distance from glottis cm

Area cm²

Area function for -a-

Frequency response

Magnitude(dB)

Figure-17 Frequency response -a- for forty-four tube
F1= 781, F2=1171, F3= 2793, F4=3323, F5= 4431
Figure 18 Frequency response for -i- forty-four tube

\[ F_1 = 218, \quad F_2 = 2417, \quad F_3 = 3465, \quad F_4 = 3876, \quad F_5 = 4662 \]
Area Function

Area from glottis cm

Area cm²

Distance from glottis cm

Area function for -o-

Frequency response

Magnitude(dB)

Figure-19 Frequency response for -o- forty-four tube
F₁ = 373, F₂ = 845, F₃ = 2417, F₄ = 3643, F₅ = 4737
VI- ACOUSTIC TRANSFER FUNCTION CALCULATION FOR EXCITATION SOURCE LOCATED AT SUPRA-GLOTTAL SYSTEM

Until this section we consider that source (excitation) signal is located at the glottis. However, for some phoneme like fricative, plosive, excitation location is somewhere at the supra-glottal system [3]. In this section, therefore, we explain the calculation frequency response of the system using two examples.

Example: Acoustic filter for fricative (§ in Turkish)

We use three-tube for modeling fricative sound (§). Lb=12.5 cm, Lp=1.5 cm and Lf=3 cm and The areas of tubes are Ab=6 cm², Ap=0.5 cm², Af=2 cm². The star denotes the location of excitation.

Figure-20 Three-tube model for fricative and plosive sound
Two-port network representation of the system and its reduced equivalent can be seen in Figure-21, 22 respectively.

Figuer-21 Two-port network representation for fricative and plosive sound

The transfer function of the system can be defined as follows:

\[ H(\Omega) = \frac{U_r(\Omega)}{U_i(\Omega)} \]

In order to find \( H(\Omega) \), first we need \( Z_{\text{SUB}} \)

\( T_{\text{SUB}} \) can be calculated using following formula.

\[ Z_{\text{SUB}}(\Omega) = \frac{A_{\text{sub}}Z_g + B_{\text{sub}}}{C_{\text{sub}}Z_g + D_{\text{sub}}} \]

Where,

\[
\begin{bmatrix}
A_{\text{sub}} & B_{\text{sub}} \\
C_{\text{sub}} & D_{\text{sub}}
\end{bmatrix} = \begin{bmatrix}
A_p & B_p \\
C_p & D_p
\end{bmatrix} \begin{bmatrix}
A_b & B_b \\
C_b & D_b
\end{bmatrix}
\]

Using Figure-22 the transfer function of the system can be calculated as follows.
Figure-22 Reduced (equivalent) Two-port network representation for fricative and plosive sound

\[ Z_p(\Omega) = \frac{A_f Z_r + B_f}{C_f Z_r + D_f} \]

\[ H(\Omega) = \frac{U_r(\Omega)}{U_i(\Omega)} = \frac{Z_{sub}}{Z_{sub} + Z_p} \cdot \frac{1}{(C_f Z_r + D_f)} \]

Figure-23 denotes the frequency response of system, it can be seen that there is an peak in about frequency 3500Hz and some zeros canceling by poles.
**Example: Acoustic filter for fricative (-f-)**

We use again three-tube for modeling fricative sound (f). For this times we choose \(L_b=15\) cm, \(L_p=1.5\) cm and \(L_f=0.5\) cm and The areas of tubes are \(A_b=6\) cm\(^2\), \(A_p=0.5\) cm\(^2\), \(A_f=2\) cm\(^2\). Using similar calculation method, we obtain frequency response of the system which can be seen in Figure-24. We that frequency response is relatively flat and there is no prominent spectral peak. There are some poles which are cancelled by zeros.

![Figure-24 Frequency response of sound -f-](image-url)
VII- ACOUSTIC TRANSFER FUNCTION
CALCULATION INCLUDING NASAL TRACT

In this section, we examine calculation of frequency response of nasal sound. We use following example.

Example: Acoustic filter for fricative (-n-)
We use tube model which can be seen in Figure-25 for modeling nasal sound (n). The parameters are Ln=12.5 cm, Lp=8.5 cm and Lm=5 cm and the areas of tubes are An=4 cm², Ap=5 cm², Am=6 cm².

![Figure-25 Tube model for nasal sound](image-url)
Two-port network equivalent of the model can be seen in Figure-26

![Two-port network representation for nasal sound](figure)

We can calculate $Z_p$ as follows

$$Z_p(\Omega) = \frac{A_n Z_r + B_n}{C_n Z_r + D_n}$$

$Z_r = \infty$ (Closed lips), then

$$Z_p(\Omega) = \frac{A_n}{C_n}$$

Figure-27 denotes the reduced equivalent two-port network model
The transfer function of the system can be calculated as follows:

\[ H_{\text{Nose}}(\Omega) = \frac{U_N(\Omega)}{U_g(\Omega)} = \frac{1}{(C_{\text{all}}Z_{\text{NR}} + D_{\text{all}})} \]

\[ T_{\text{Nose}}(\Omega) = \begin{bmatrix} A_b & B_b \\ C_b & D_b \end{bmatrix} \begin{bmatrix} 1 \\ 1/Z_p \end{bmatrix} \begin{bmatrix} A_v & B_v \\ C_v & D_v \end{bmatrix} \]

\[ = \begin{bmatrix} A_{\text{all}} & B_{\text{all}} \\ C_{\text{all}} & D_{\text{all}} \end{bmatrix} \]

Figure-28 denotes our simulation result and Figure-29 is the real spectrum of nasal sound -n- which is given comparison purpose.
Figure-28 Frequency response of sound -n-

Figure-29 FFT and LPC for a sound -n-
VIII- FEATURE WORK

We observe that there is a close and non-linear relation between vocal tract area function, vocal tract resonance frequencies and speech formant frequencies. Our future work plan is to combine these relations and try to give better explanation for speech phenomena.
IX- REFERENCES


