A note on the vocal tract wall impedance

Fant, G. and Nord, L. and Branderud, P.

journal: STL-QPSR
volume: 17
number: 4
year: 1976
pages: 013-020

http://www.speech.kth.se/qpsr
A. A NOTE ON THE VOCAL TRACT WALL IMPEDANCE
G. Fant, L. Nord, and P. Branderud

Abstract

Measurements of vibrational amplitude externally on the walls of the neck and face, when a low frequency sinewave is injected through the lips and the glottis is held closed, show a maximum amplitude at the level of the larynx and another but somewhat weaker maximum at the closed lips of the subject. This pattern suggests a nonuniform distribution of the equivalent distributed mass-loading, the major part of which may be lumped at the glottal end of the tract. Input impedance measurements through a tube held between the lips and with the tongue either raised to close off the mouth cavity or flat in the mouth to allow full coupling to the pharynx have made possible an estimate of an anterior and a posterior part of the massload as well as of the volumes of the front and back parts of the tract. The resonance frequency and bandwidth of the closed vocal tract were found to be of the order of $F_w = 190$ Hz and $B_w = 75$ Hz, respectively, for male subjects and $F_w = 220$ Hz and $B_w = 95$ Hz for female subjects which agree with the Fujimura-Lindqvist (1971) data. It has been found from electrical line analog simulation that the wall mass element not only affects the tuning of low frequency $F_1 = (F_0^2 + F_w^2)^{1/2}$ with the closed tract limit $F_1 \to F_w$ but also appears to be a significant factor in the tuning of pharyngeally constricted sounds such as [a] with high $F_1$ in which case the small back cavity volume is especially sensitive to the mass shunt.

Introduction

From several studies it is apparent that the vocal tract wall impedance is a major determinant of the frequency and bandwidth of very low frequency first formants, the extreme limit of $F_1$ being the closed tract resonance frequency. Under these conditions $F_1 = F_w$ is of the order of 150-250 Hz and $B_1 = B_w$ is of the order of 75-100 Hz. The equivalent circuit of the closed tract is simply a condenser $C_t = V_t/\rho c^2$, where $V_t$ is the total volume of the contained air connected in parallel with a branch containing the lumped mass element $L_w$ of the walls and an associated series resistance $R_w$ determining the bandwidth

$$F_w = \frac{1}{2\pi} \left( L_w C_t \right)^{-\frac{1}{2}}$$

$$B_w = \frac{1}{2\pi} \cdot \frac{R_w}{L_w}$$

With a finite opening of the mouth at the lips or the tongue passage the low frequency approximation is extended to include another parallel
branch containing an inductance \( L_1 \) with a rather small resistance element \( R_1 \). Under these conditions

\[
F_1 = \frac{1}{2\pi} \left( \frac{L_1 L_w}{L_1 + L_w} \right)^{-1/2}
\]

\[
B_1 = \frac{1}{2\pi} \cdot \frac{R_w}{L_w} \cdot \frac{L_1}{L_1 + L_w} + \frac{1}{2\pi} \cdot \frac{R_1}{L_1} \cdot \frac{L_w}{L_1 + L_w}
\]

which may be rewritten as

\[
F_1 = (F_w^2 + F_i^2)^{1/2}
\]

\[
B_1 = B_w \cdot \frac{F_w^2}{F_i^2} + B_i \cdot \frac{F_i^2}{F_i^2}
\]

where

\[
F_i = \frac{1}{2\pi} \left( \frac{L_1 C_i}{L_1 + L_w} \right)^{-1/2}
\]

\[
B_i = \frac{1}{2\pi} \cdot \frac{R_1}{L_1}
\]

is the frequency and bandwidth of the resonance with the wall impedance branch \( R_w + j\omega L_w \) neglected, i.e. hard walled conditions. As a consequence, with increasing mouth opening, \( F_1 \) rapidly approaches the hard wall value \( F_i \). In the low and also the medium frequency range of \( F_1 \) the bandwidth term associated with \( R_w \) is relatively dominant in spite of the inverse frequency square dependency. Fujimura and Lindqvist (1971), Fant and Pauli (1974), and Fant (1976).

It is the purpose of our study to provide more accurate data on \( F_w \) and \( B_w \) and also to study the gross distribution within the tract of the wall inductance \( L_w \). We have also had a suspicion that the wall inductance in the pharyngeal area could have a noteworthy influence on \( F_1 \) of \([a\) type vowels. The acoustic impedance method we developed provides indirect measurements of the elements of the equivalent circuit, whilst the measurements performed by Ishizaka, French and Flanagan (1975) involve direct measurements of the wall impedance by external measurements. The compliance of the walls which rightly enters as a condenser \( C_w \) in series with \( L_w \) has been estimated by these authors to provide a series resonance with \( L_w \) at 30-70 Hz, which is well below the range of \( F_w \) and, accordingly, less important.
Method

In an introductory study we measured the distribution of vibrational amplitude externally on the walls of the neck and face of a male subject who closed his lips around a narrow sound-emitting plastic tube connected to a loudspeaker driven by a low frequency signal. The subject was instructed to keep his glottis closed. A piezoelectric transducer was used for the pick-up. Lines of equal vibrational amplitude were constructed. As shown in Fig. II-A-1 these display two dominant regions of vibration, one at a level just above the larynx and the other somewhat less intense at the lips. These would suggest that the shunting effect of the distributed mass load might be lumped into two inductance elements, one at the lips and one at the larynx, where the walls are thinner than at other places.

Our next experiment aimed at a direct measurement of the closed tract resonance frequency and bandwidth. The same sound-emitting system was used together with a pick-up probe tube also inserted through the closed lips. A variable frequency oscillator was tuned by hand whilst tracing the response curve. The results were encouraging. For male subjects we measured $F_w$ close to 190 Hz and $B_w = 75$ Hz with standard deviations for repeated measurements within a subject as low as 4 Hz and 6 Hz, respectively, and inter-subject variations of 15 Hz in $F_w$ and 6 Hz in $B_w$ for 5 male subjects.

We next designed an improved experimental set-up aiming at not only a recording of the input resonance curve but also a calculation of acoustic circuit elements. The method is illustrated in Fig. II-A-2. The input and output probes and a short metal tube acting as a known inductance are inserted through a plexiglass adapter shaped to be held comfortably between half-open lips with good acoustic scaling.

The tube of length 35 mm and diameter 9 mm was supplied with a shutter to allow it being closed off acoustically. In Fig. II-A-2 the tube is represented by its inductance $L_1$ in series with a switch. The inductance element was calculated from standard formula as

$$L_1 = \frac{\rho l_e}{A_1} \left[ 1 + \frac{S}{A} \left( \mu / (2 w) \right)^{1/2} \right]$$

where the effective length $l_e = 4.2$ cm is the sum of the physical length 3.5 mm and the endcorrections at both ends. The correction for the frictional
Fig. II-A-1. Equal vibrational amplitudes along the face and neck measured externally with an accelerometer. Low frequency sound is injected from a narrow probe tube through the lips of the subject who holds his lips closed.
Fig. II-A-2. Equivalent electric circuit of the VT low frequency input impedance measurements. $L_1$ is the inductance of a tube held between the lips of the subject.
layer (Fant, 1960) has been included in Eq. (8). The probes were 3 mm thick acoustically damped plastic tubes inserted in separate holes through the plexiglass adapter so as to minimize the interference with the impedance structure.

The theory of operation is simple. The resonance frequency is measured under four conditions. The lip tube is either shut or open and this is repeated in two articulations. One is with the tongue against the hard palate as in the occlusion of the syllable [ga]. The other is with neutral tongue articulation to allow for a free coupling between the mouth and the pharynx. In the low frequency approximation we may thus neglect the tongue hump inductance $L_h$ in series with the switch for closing off or opening the tongue passage. With this switch open the two measurements involving the front cavity provide the measures

$$L_{w1} = L_1 \left( \frac{F_1^2}{F_{w1}^2} - 1 \right) \quad (9)$$

$$C_1 = \frac{1}{4\pi^2 L_1 (F_1^2 - F_{w1}^2)} \quad (10)$$

where $F_{w1}$ pertains to the lip tube shut and $F_1$ to the lip tube open, see Eqs. (1) and (3).

In the open tongue articulation we assume that the entire vocal tract be regarded as a single Helmholtz resonator with

$$C_t = C_1 + C_2$$
$$L_w = \frac{L_{w1} \cdot L_{w2}}{L_{w1} + L_{w2}}$$

which can be calculated by the same procedure from Eqs. (1)(3)(9) and (10). Thus

$$C_2 = C_t - C_1$$

$$L_{w2} = \frac{L_w \cdot L_{w1}}{L_{w1} - L_w}$$

(12)
Discussion of experimental techniques

A few words should be said about the recording technique. We designed an automatic linear up and down frequency sweep of the oscillator with a total period of about 1 sec and a log amplitude display of the microphone signal on a persistant screen oscillograph. To aid the subject in maintaining an elevated velum we inserted a microphone probe tube in the nose. This set-up allowed a convenient check of the short time stability of a subject's performance.

The sound emitter and receiver units were tested in a closed cavity calibrator to provide a -6 dB/octave fall of the response within ±1 dB from 65 Hz to 650 Hz. This conforms with the linearity requirement for the sound pressure developed in a cavity as a response to a constant volume velocity source and accordingly a symmetric resonance curve (conjugate pole with zero at origin) when tracing the fundamental mode of the VT input impedance. It was also possible to trace higher modes of the VT input impedance and this technique might have been utilized systematically for ensuring stability of VT configuration during the experiments.

Some subjects gave reproducible measurements from one day to another, others were less stable and some showed a drift of data from one sweep to the next. Such changes can depend on the instability of tongue articulation as well as on insufficient glottal closure and tension built-up in the vocal wall muscles. The nasal monitor appeared to be quite effective in avoiding an open nasopharyngeal part but some minor variations could still influence the data. Also some variations could be caused by how tight the subjects closed their lips around the mouthpiece.

Results and discussion of data

In the following tabulation we have summarized the results from measurements of the closed tract frequency resonance $F_w$ and bandwidth $B_w$ and the derived values of lumped inductance $L_{w1}$, $L_{w2}$ and volume $V_1$, $V_2$ of the front and back parts and of the whole VT, $L_w$, and $V_t$. The resonance frequency of the back part $F_{w2} = \frac{1}{2} (L_{w2} C_{w2})^{-\frac{1}{2}}$ is also included.
The measurements were made with 5 male and 7 female subjects. The intersubject standard deviations are noted.

TABLE II-A-1.

<table>
<thead>
<tr>
<th></th>
<th>$F_w$</th>
<th>$B_w$</th>
<th>$F_{w2}$</th>
<th>$L_{w1}$</th>
<th>$L_{w2}$</th>
<th>$L_w$</th>
<th>$V_1$</th>
<th>$V_2$</th>
<th>$V_t$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hz</td>
<td>Hz</td>
<td>Hz</td>
<td>gcm$^{-4}$</td>
<td>gcm$^{-4}$</td>
<td>gcm$^{-4}$</td>
<td>cm$^3$</td>
<td>cm$^3$</td>
<td>cm$^3$</td>
</tr>
<tr>
<td>5 males average</td>
<td>191 76</td>
<td>183</td>
<td>0.038</td>
<td>0.015</td>
<td>0.011</td>
<td>18</td>
<td>80</td>
<td>98</td>
<td></td>
</tr>
<tr>
<td>st. dev.</td>
<td>15 4</td>
<td>0.006</td>
<td>0.007</td>
<td>0.004</td>
<td>7</td>
<td>36</td>
<td>39</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 females</td>
<td>218 94</td>
<td>202</td>
<td>0.032</td>
<td>0.024</td>
<td>0.014</td>
<td>21</td>
<td>32</td>
<td>53</td>
<td></td>
</tr>
<tr>
<td>average</td>
<td>27 11</td>
<td>0.006</td>
<td>0.008</td>
<td>0.006</td>
<td>8</td>
<td>22</td>
<td>20</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The data on $F_w$ and $B_w$ appear to be reliable within 4%. It may be noted that Fujimura and Lindqvist (1971) measured $F_w = 189$ Hz and $B_w = 73$ Hz for a Swedish subject articulating the vowel $[\text{u}]$. A value of $F_w = 180-200$ Hz was derived by Fant and Lindqvist (1968) from a study of formant shifts in divers' speech under different conditions.

In an earlier study by Fant and Sonesson (1964) somewhat lower values were calculated, $F_w = 150-180$ Hz, for the particular male divers. The lowest value of $F_w$ observed in our test series above was 170 Hz.

An estimate of the loss resistance $R_{w2}$ may be made from Eq. (4) by assuming $R_{w2}/L_{w2} = R_w/L_w$.

$$R_{w2} = 2\pi L_{w2} B_{w2} = 18 \text{ gcm}^{-4} \text{ sec}^{-1} \text{ (acoustical ohms)}$$

(13)

for the male group and 15 acoustical ohms for the female vocal tract.

The values of the lumped inductances and volumes have a greater degree of uncertainty than the resonance frequencies and bandwidths which can be seen from the tabulation. This is especially true of $L_{w2}$ and $V_2$ which are subject to two sources of error. One is due to the difference term in the denominator of Eq. (12), the other is the rather disputable assumption that the sum of $V_1$ and $V_2$ is independent of articulation. The total volume $V$ pertains strictly to the neutral articulation only whereas $V_2$ is merely the difference between the total volume of the neutral articulation and the front cavity volume of the $[\text{ga}]$ articulation. The accuracy in the estimation of $V$ is of the order of 20%.
How do these values compare with other data in the literature? Ishizaka, French and Flanagan (1975) report a distribution of mass of the order of 2 g/cm² which, converted to acoustical impedance by division with a total assumed surface of the vocal tract of 160 cm², (corresponding to an elliptical cross-sectional VT shape of 2 by 4 cm, 16 cm length and 100 cm³ volume, see Table II-A-1) leads to a total lumped inductance of 0.0135 g cm⁻⁴, which is equal to the mean of our male and female data.

Ishizaka et al measure a somewhat smaller inductance at the cheeks than at the neck which is contrary to the trend in our data. However, in view of the different measurement techniques and the apparent non-uniform distributions revealed by our Fig.II-A-1 and considering the lack of information of the actual places of measurements in the Ishizaka et al data we cannot expect to find a closer agreement.

What about the calculated volumes? Are they realistic? The order of magnitude appears to be correct as judged by the study of the Russian vowels analyzed by Fant (1960). Volumes ranged from 68 to 100 cm³ with 98 cm³ for the open vowel [ə] which is the same as that of our male average for neutral articulation. From an unpublished study of an X-ray tomographic material analyzed by Edholm (see Fant, 1964), we can quote ratios of male to female vocal tract total volumes for the vowels [a], [i], and [u] to be 86/36, 46/24, and 60/32, respectively, which indicate that the volume of a female vocal tract is about one half of that of a male subject. Because of the unnatural positionings of the subjects in the tomographic study, there may have entered systematic deviations from normal vocal tract volumes but the female/male ratios are the same as what we have reported in Table II-A-1. The front cavity volume \( V_f \) was 18 cm³ for our males and 21 cm³ for our females. This might be compared with the \( V_f = 21 \) cm³ and 19 cm³ for the male and female front cavity volume of the vowel [u] in the Edholm study. These comparisons support the representability of our volume data, but it should be kept in mind that our intersubject variations were of the order of 40% and our group means have an uncertainty of 20%.

The wall inductance \( L_{w2} \) of our model, representing a shunt in the lower pharyngeal region, cannot be expected to be independent of the particular vowel articulated. We do not have sufficient data on such
variations but it appears that the increase of $L_{w2}$ with pharyngeal narrowing as in the vowel [a] is not very large. Assuming a value $L_{w2}$ according to Table II-A-I we were able to connect a corresponding inductance in parallel to our vocal tract electrical line analogue LEA, Fant (1960), at the laryngeal end. This addition raised $F_1$ of a vowel [a] with about 100-150 Hz depending on the articulation whilst other formants were not much affected. This effect is perceptually significant. Moreover, in simulations we have done of [a]-type vowels from articulatory data, we have generally come out with too low $F_1$, which makes the vowel sound centralized unless an inductance shunt is added. However, it has not yet been proved that $L_{w2}$ is substantially independent of articulation and a similar increase in $F_1$ could also be accounted for by a small nasal coupling which often occurs in maximally open vowels. Incomplete vocal cord closure would also have a similar effect.

We welcome attempts from other research groups to clarify these problems. With proper monitoring and control the impedance measurement technique we have developed is potentially useful for this type of investigations.

References