ACOUSTIC TRANSMISSION-LINE ANALYSIS OF FORMANTS IN HYPERBARIC HELIUM SPEECH

Per Lunde
Norwegian Underwater Technology Center (NUTEC)
P.O.Box 6, 5034 Ytre Laksevaag, Bergen, Norway

ABSTRACT
Acoustic transmission-line vocal tract models are used to study formant frequencies, bandwidths and amplitudes of hyperbaric heliox speech versus those of speech in air at 1 ATA. The models account for energy losses due to glottal impedance, lip/nostrils radiation, wall vibration, viscous friction and thermal conduction. New wall impedance values are presented, matching measurements of the closed tract resonance. On basis of a uniform tube model, an extended version of the classic Fant-Lindquist formula describing formant frequency shifts is developed, and formulas for bandwidth and amplitude shifts are given. A multitube vocal tract model is applied for analysis of the effects of nonuniform vocal tract cross-sectional area on the formant shift.

INTRODUCTION
The speech of divers breathing a hyperbaric heliox gas mixture at large depths is known to be nearly unintelligible. Although the intelligibility is raised by speech correction through helium speech unscramblers, the quality of the corrected speech is not sufficient for safe and efficient operation. The intelligibility decreases rapidly when moving towards greater depths, even with the most advanced of today's unscramblers. This might be a reflection of the fact that the knowledge of diver's speech distortion is far from sufficient, based upon simplified acoustic models of the speech production system. The purpose of this work is to achieve extended knowledge of the distortion through enhanced speech production models.

Assuming that speech can be considered stationary over the time interval considered, also neglecting nonlinear interaction, the speech spectrum S(f) (measured by a microphone) can be described by a linear source-filter model:

\[ S(f) = G(f) V(f) L(f) M(f) \]

where \( G(f) \) is the source (excitation) signal spectrum, \( V(f) \) is the vocal tract transfer function, \( L(f) \) is the radiation characteristic and \( M(f) \) is the combined mask/microphone response. All of these terms may contribute to the total speech distortion under hyperbaric heliox conditions. In this paper, however, we will concentrate on \( V(f) \), discussing distortion due to changing formant structure as depth and gas mixture are changed.

\( V(f) \), defined as the ratio of volume velocity through lips (nostrils in case of nasals) to volume velocity through glottis \( U_{h}/U_{g} \) (\( U_{h}/U_{g} \)), determines the formant (vocal tract resonance) structure of speech. Based upon a uniform acoustic single-tube model of the vocal tract, assuming lumped elements and neglecting glottal, radiation, viscous and thermal losses, Fant and Lindquist [1] derived their classical formula describing the nonlinear shift of formant frequencies from air at 1 ATA to hyperbaric heliox conditions. In a distributed elements representation, Richards and Schafer [2] have extended the uniform tube model to account for the loss terms lacking in the Fant-Lindquist model. Shifts of formant frequencies, bandwidths and amplitudes were all discussed.

The first part of the present work is based upon Richards and Schafer's uniform tube model. However, an extended mathematical discussion is performed, leading to explicit expressions and shift formulas for formant frequencies, bandwidths and amplitudes. In addition, new wall impedance data are derived and applied, providing more realistic bandwidths of lower first formants. In the second part of the work, a multitube vocal tract model is applied for analysis of the influence of nonuniform vocal tract geometry on the formant shifts. With this model simulations of \( V(f) \) for 5 vowels and 2 nasal consonants were performed in air at 1 ATA as well as hyperbaric heliox conditions.

ACOUSTIC SINGLE-TUBE VOCAL TRACT MODEL
Transmission-Line Description
As a first approximation for studying vocal tract sound transmission, we neglected the effects of nonuniform vocal tract geometry. The vocal tract was modelled as a single, uniform (cylindrical), lossy tube of length \( L = 17.5 \text{ cm} \) and radius \( r = 1.26 \text{ cm} \), enabling derivation of explicit formant parameter formulas.

The acoustic transmission-line analog of the uniform tract model is shown in Fig. 1. Here \( Z_g = R_g + i \omega L_g \) and \( Z_m = R_m + i \omega L_m \)

Assuming that speech can be considered stationary over the time interval considered, also neglecting nonlinear interaction, the speech spectrum \( S(f) \) (measured by a microphone) can be described by a linear source-filter model:

\[ S(f) = G(f) V(f) L(f) M(f) \]

where \( G(f) \) is the source (excitation) signal spectrum, \( V(f) \) is the vocal tract transfer function, \( L(f) \) is the radiation characteristic and \( M(f) \) is the combined mask/microphone response. All of these terms may contribute to the total speech distortion under hyperbaric heliox conditions. In this paper, however, we will concentrate on \( V(f) \), discussing distortion due to changing formant structure as depth and gas mixture are changed.

\[ V(f) \text{ defined as the ratio of volume velocity through lips (nostrils in case of nasals) to volume velocity through glottis } U_{h}/U_{g} = (U_{h}/U_{g}) \text{ determines the formant (vocal tract resonance) structure of speech. Based upon a uniform acoustic single-tube model of the vocal tract, assuming lumped elements and neglecting glottal, radiation, viscous and thermal losses, Fant and Lindquist [1] derived their classical formula describing the nonlinear shift of formant frequencies from air at 1 ATA to hyperbaric heliox conditions. In a distributed elements representation, Richards and Schafer [2] have extended the uniform tube model to account for the loss terms lacking in the Fant-Lindquist model. Shifts of formant frequencies, bandwidths and amplitudes were all discussed.} \]

The first part of the present work is based upon Richards and Schafer's uniform tube model. However, an extended mathematical discussion is performed, leading to explicit expressions and shift formulas for formant frequencies, bandwidths and amplitudes. In addition, new wall impedance data are derived and applied, providing more realistic bandwidths of lower first formants. In the second part of the work, a multitube vocal tract model is applied for analysis of the influence of nonuniform vocal tract geometry on the formant shifts. With this model simulations of \( V(f) \) for 5 vowels and 2 nasal consonants were performed in air at 1 ATA as well as hyperbaric heliox conditions.
are the glottal and lip radiation impedances, respectively. \( Z_1 \) and \( Z_2 \) are defined by \( Z_1 = \frac{Z_0}{\tanh(Y/2)} \) and \( Z_2 = Z_0/\sinh(Y/2) \), where \( Z_0 = \left(\frac{Z_0}{\lambda_0}\right)^{1/2} \) is the characteristic impedance and \( Y = \left(\frac{Z_0}{\lambda_0}\right)^{1/2} \) is the (complex) sound propagation constant \([3]\). The acoustic impedance \( Z_a = R_a + \text{i} \omega L_a \) is determined by mass and mass inerter of the gas. The admittance \( Y_a = G_a + \omega C_a + \text{i} (\omega L_a + 1/\omega C_a) \) is determined by thermodynamic gas compressibility, wall viscosity and wall mass vibration. For the glottal impedance, Flanagan's equivalent small-signal model \([3]\) was used. The viscous/thermal loss terms were also taken from \([3]\). For the lip radiation impedance, the piston-in-square and piston-in-infinite-baffle models were both used. What wall impedance (wall viscosity and mass vibration) concerns, the values given by \([3]\) (used by \([2]\)) did not perform sufficiently well, giving rise to extremely large bandwidths of formants in the vicinity of the closed tract resonance. Wall impedance values providing more realistic results therefore had to be determined. This point is discussed in a separate section below.

It should be noted that in the present transmission-line representation there is no assumption of lumped elements. This would have constricted the validity range of the analysis.

Formant Analysis

The vocal tract transfer function \( V(f) \) of the uniform, single tube was derived from the conditions of continuity of pressure and volume velocity for vocal tract sound transmission. The complex poles of \( V(f) \) were then found, from which the formant frequencies and bandwidths were readily determined. The details will not be given here. Under the assumptions \( |Z_0/Z_2| + |Z_2/Z_0| < 1 \), \( |Z_0/Z_2| < 1 \), \( R_a << \omega L_a \), \( G_a + \omega C_a << \omega C_a - 1/\omega L_a \) and 4 \((F_0/F_0) \omega L_a^2/3 \), \( 1 - \kappa g \) \( \ll 1 \), it is shown in \([4]\) that the formant frequencies \( F_n \), bandwidths \( B_n \) and amplitudes \( A_n \) are given by the formulas

\[
F_n = \left\{ F_0^2 \kappa g^2 + F_0^2 \left[ 1 - \kappa g \right]^2 \right\}^{1/2}
\]

\[
B_n = \frac{\kappa g \left[ B_0 + B_g + B_0 + B_0 \right]}{1 - \kappa g}
\]

(2) \hspace{1cm} (3)

where \( \kappa g \) is the low-frequency region.

Here \( c \) is the speed of sound, \( F_0 = (2n-1)c/4l \) is the n-th formant frequency of the lossless uniform tube, \( F_0 = 1/c \sqrt{(C_0 L_0)} \) is the closed tract resonance frequency, \( \kappa g \) \( = (4n \pi c_0 / \lambda_0) \) is the lip radiation correction factor (approximately equal to unity) and \( \kappa g \) \( = (4n \pi c_0 / \lambda_0) \) is a glottal correction factor (nearly equal to zero except for lower frequencies in hyperbaric gaseous conditions). The bandwidth contributions are all frequency dependent, given as \( B_0 = R_0 / 2 \pi L_0 \), \( B_g = G_g / 2 \pi C_0 \), \( B_0 + B_0 = \lambda c_0 / \pi L_0 \) and \( B_0 + B_0 = \lambda c_0 / \pi L_0 \). The main contributions to the formant frequencies are seen to descend from the lossless formant frequencies and the vibrating cavity walls. The lip radiation and glottal impedances contribute through small correction terms only. Thermal and viscous losses do not influence at all. For the formant bandwidths, however, all losses contribute to the total damping. The relative contribution to the total bandwidth from each loss term is shown as function of formant frequency in Fig. 2 for air at 1 ATA and heliox at 500 msw. Lower formant damping is dominated by wall vibration and glottal losses, whereas upper-frequency formants are largely damped by lip radiation losses. Note that when each loss term is considered isolated, eqs. (2) and (3) reduce perfectly to the formulas given by Flanagan \([3]\).

Formant Shift

Under the additional assumption that \( \kappa g \ll 1 \) (valid for \( f >> F_w \)) we are now able to obtain expressions relating formant frequencies, bandwidths and amplitudes in hyperbaric heliox (index 'he') to the corresponding formant parameters in air at 1 ATA (index 'a'):

\[
\frac{F_{he}}{F_a_\text{ca}} = \psi \left[ 1 + \frac{F_{he}}{F_a} \right]^{2} + \frac{2 \kappa g \psi_c a}{\left[ 1 - \psi_c a \right]^{1/2}}
\]

\[
B_{he} = \frac{B_a}{\psi_c a} \left[ B_0 + B_g + B_0 + B_0 \right] \text{he}
\]

(5) \hspace{1cm} (6)

where \( \psi \) is the density of the gas, proportional to pressure. We notice that for \( \kappa g \psi_c a = 0 \) and the formant frequency shift formula, eq. (5), reduces to the classic Fant-Lindquist formula, originally derived under assumption of no glottal, lip radiation nor viscous/thermal losses \([1]\). Here it is shown to correspond to the closed-glottis case, with some restrictions on \( Z_0 \) and \( Z_2 \). In general, however, the glottal/lip radiation correction term, though small, should be included.

The curves of Fig. 5 gives the shifts of formant frequencies and bandwidths from air at 1 ATA to heliox at 500 msw as predicted by eqs. (5) and (6). We notice several interesting features:

- In the low-frequency region the formant frequency shift is dominated by pressure and wall effects. The shift of sound velocity dominates in the mid-and-upper-frequency range. The glottal and lip radiation losses imply a slight modification of the Fant-Lindquist formula at low frequencies only.

- In the upper frequency range the formant bandwidth shift is dominated by the glottal and lip radiation losses. Towards lower frequencies wall losses become more influential, while
The specific wall impedance, defined as \( z = \frac{R_w}{G_w} \), is studied in the context of the acoustic multtube vocal tract model, discussed below.

For low-frequency formants the wall impedance values used in their analysis (given by [3]) imply serious overestimation of the frequency and especially the bandwidth shift. Furthermore, in [4] it is shown that their formant bandwidth expression is equivalent to

\[
B_w = k_m \left[ B_n + B_g + B_w + (1 - (F_w/F_n)^2)(B_v + B_m) \right] \tag{5}
\]

where \( k_m = 1 - G_m/F_m^2 \). This implies unrealistic large bandwidths as well as additional overestimation of bandwidth shifts for formants that are the vicinity of the closed tract resonance frequency. The requirements of consistency with the assumptions already made in the derivation of eq. (5) implies that eq. (3) is the correct formant bandwidth expression.

Validity Range

It can be shown [4] that the approximations leading to eqs (2) - (7) are valid in the formant frequency range

\[
F_w << F_n << \frac{3c}{2\pi r_w}, \quad n = 1,2,\ldots
\]

where \( r_w \) is the lip opening radius. The lower limit constraining the validity range is a result of restrictions based on the glottal and wall impedances, while the upper limit arises because of restrictions laid upon the lip radiation impedance. Therefore, for typical tube parameters, the formulas are valid for 190 Hz \( << \) \( F_n << 2650 \text{ Hz} \) in air at 1 ATA. For heliox at 500 msw, however, the validity range is constricted to 1550 Hz \( << \) \( F_n << 7500 \text{ Hz} \). From these considerations, we realize that there is need for a theory valid in a wide frequency range which is not constricted at large depths. This is solved by introducing the multtube vocal tract model, discussed below.

Wall Impedance

The specific wall impedance, defined as \( z_w = r_w + 1/G_w \), is essential for the present analysis, and the values for \( z_w \) given by [3] produce unrealistic large bandwidths for formants in the lower frequency range [4]. A comment on the values used here should therefore be given. In order to obtain realistic estimates of \( z_w \), it was matched to measurements of the resonance frequency and bandwidth of the closed vocal tract in air at 1 ATA. Eqs. (2) - (7) were derived under approximations not valid in the closed-glotis case and could not be used for this purpose. As a better approach, we assumed lumped elements representation, valid for \( \gamma_1 << 1 \). Under the additional assumptions of no viscous nor thermal losses, closed glottis (\( |Z_g| \rightarrow \infty \)) and lips closed (\( |Z_l| \rightarrow \infty \), \( \omega |Z_w| \rightarrow 0 \)), the specific wall resistance and inductance were found to be

\[
r_w = \frac{2\pi l_0}{l_0}(F_w) \tag{10}
\]

\[
1_w = \frac{c^2/2\pi f^2}{l_0}(F_w^2 + B_w^2(F_w)) \tag{11}
\]

where \( F_w \) and \( B_w(F_w) \) are the closed tract resonance frequency and bandwidth, respectively. By matching this to \( F_w = 190 \text{ Hz} \) and \( B_w = 75 \text{ Hz} \), we obtained \( r_w = 6500 \text{ kg/m}^2 \) and \( l_w = 13.8 \text{ kg} \cdot \text{m}^2/\text{s}^2 \). These are the values used in all calculations of this work. Note that the wall impedance is not uniformly distributed over the vocal tract and the values derived are valid only on an average basis.

ACOUSTIC MULTITUBE VOCAL TRACT MODEL

Transmission-Line Description

As a more realistic approximation of the vocal tract, a multtube model of the tract was developed and implemented. The purpose of this analysis was to obtain a model which, first, is valid in a wide frequency range also at large depths (to test the reliability of the uniform tube formulae, eqs. (2) - (7)) and second, enables investigation of the influence of nonuniform cross-sectio-
The acoustic transmission-line analog of the multi-tube tract model is shown in Fig. 3. The pharynx, mouth and nose are all approximated by a number of cascaded, lossy, cylindrical tubes of arbitrary radii and lengths, all accounting for wall, viscous and thermal losses. Each tube is represented in the transmission line as an equivalent distributed T-section. The network is terminated by impedances due to glottis and radiation from lips and nostrils. Network theory was applied to transform the conditions of continuity in pressure and volume velocity into matrix representation, suitable for computer simulation. The resulting transmission equations are given in Fig. 3. The general outline of the model is sketched in [3]. The recursion schemes for solving the impedance determinants involved, however, are derived in [4], and all the impedances involved are represented as distributed elements. The simulation system provides a flexible instrument where the model parameters (gas, glottal, lips, nostrils, wall and tube parameters) can be varied independently of each other. For vowel simulations, no nasal coupling was assumed. For simulation of nasals, the main transmission path is the pharyngeal/nasal cavities, with the mouth acting as a parallel and nonuniform vocal tract area on the formant shift. The simulations are therefore valid in the approximate range 0 - 7 kHz in air and 0 - 20 kHz in pure helium gas.

**Simulation Results**

Using cross-sectional vocal tract area data given by Fant [6], the formant structures \( V(f) \) of 5 non-nasalized Russian vowels (\( /a/, /o/, /u/, /N/, /e/ \)) and 2 Russian nasal consonants (\( /m/, /n/ \)) were simulated for various breathing gas mixtures and pressures corresponding to depths 0 - 500 ms. 34 - 40 tubes were used for modelling the vowels. 59 tubes were used for the nasals. Each tube was 0.5 cm long. In order to evaluate the reliability of eqs. (2) - (7), \( V(f) \) for the single, uniform tube of length 17.5 cm (the "neutral vowel") was also simulated. Formants were extracted from the computed data by a peak-picking routine. This provided data for shifts in formant parameters for each individual phoneme. Details are given in [4].

Fig. 4 gives the transfer function \( V(f) \) for \( /a/ \) and \( /m/ \), simulated for conditions corresponding to 500 ms. In Fig. 5 the formant shift data obtained from the simulations of the uniform tube, the 5 vowels and the 2 nasals are plotted together with the predicted results of eqs. (5) and (6). We conclude with the following:

- The formant shift formulas, eqs. (5) and (6), are valid in a wider frequency range than their derivation procedure suggests.
- The extended Fant-Lindquist formula, eq. (5), describes the formant frequency shift independent of vocal tract geometry.
- The shift of formant bandwidth depends on articulation, i.e. vocal tract geometry. The "neutral vowel" has a bandwidth shift satisfactorily described by eq. (6), while nonuniform cross-sectional vocal tract area seems to decrease the mean formant bandwidth shift.
- Except for the lower frequency range, the mean bandwidth shift of nasals seems to be in the lower edge of the mean bandwidth shift of vowels.

**CONCLUSIONS**

Our analysis provides new insight into the shift of formant frequencies, bandwidths and amplitudes from air at 1 ATA to hyperbaric helium conditions. For both uniform and nonuniform vocal tract geometry, the formant frequency shift is well described by the extended Fant-Lindquist formula, eq. (5). For nonuniform vocal tract (as is the realistic configuration for speech) our simulations indicate that the mean formant bandwidth shift is less than predicted by the uniform tube vocal tract model. For both uniform and nonuniform vocal tract the bandwidth shift is less than the formant shift, at low frequencies (in the range of the first formant) the difference is large. The shift of formant amplitudes is inversely proportional to the formant bandwidth shift.

**ACKNOWLEDGEMENTS**

I wish to thank my colleagues M. Ye sterheim and K. Dalland, both NUTEC, for numerous and valuable discussions on the subject of this paper. The work was carried out on a fellowship from The Royal Norwegian Council for Scientific Research.

**REFERENCES**


