Speech at high ambient air-pressure

Fant, G. and Sonesson, B.

journal: STL-QPSR
volume: 5
number: 2
year: 1964
pages: 009-021

http://www.speech.kth.se/qpsr
B. SPEECH AT HIGH AMBIENT AIR-PRESSURE*

1. Summary

The present study was initiated as an attempt to gain insight in the physiological and acoustical nature of the typical distortion of diver's speech at deep underwater levels. At a depth of more than 30 m, i.e. pressures in excess of 4 ata, the voice attains a typical "nasal" quality and spectrographic analysis shows an increase of formant frequencies and of voice fundamental frequency. Wathen-Dunn and Copel (6), Holywell and Harvey (5), and others have described these effects but they have been more concerned with problems of intelligibility and the toxic effects of absorbed gases than with the acoustic problems. As far as we know the mechanism underlying the "nasal quality" of over-pressure speech has not been satisfactorily explained before.

Our study started with speech recordings in the decompression chamber of the Swedish Marine in Karlskrona in 1960 and 1962. Spectrographic analysis showed that the frequency shift was proportionally greater in \( F_1 \) than in higher formants. A simple model of a linear frequency shift, such as associated with a change of the velocity of sound, did not fit these experiments. Furthermore it is known that the velocity of sound in air is almost independent of the pressure.

The most recent and conclusive study was undertaken in April 1964 in the decompression chamber on board HMS Belos in Stockholm. The latter facilities include an X-ray outfit which made it possible for us to study the velar function of a subject during phonation. Frontal and sagittal X-ray pictures showed a

* The experimental part of this study was supported by the Swedish Medical Research Council Grant T 312 and W 267 and by a grant for speech communication research from the Swedish Technical Research Council. A summary will be presented at the 68th Meeting of the Acoustical Society of America in Austin/Texas, October 1964.
normal status of the velum at 6 ata pressure. Although these negative results excluded velo-pharyngeal opening as the main cause of the observed spectrum distortion, the very nature of this distortion strongly indicates the presence of some kind of shunting mechanism in vocal transmission.

A theoretical analysis has now supplied conclusive evidence that the shunting mechanism is associated with the vibration of the walls of the vocal cavities, especially the soft parts of the throat. As a by-product of this study the role of the cavity walls in normal speech has been emphasized. These results tie in well with recent experiments and analysis of vocal transmission performed by Dr. O. Fujimura at the Speech Transmission Laboratory.

2. Spectrographic study

The recording of speech in the pressure tank of HMS Belos was carried out with a battery operated tape-recorder and a dynamic microphone. A speaking distance of 2 inches to the microphone was maintained. Recordings were made at 1 ata, i.e. normal atmosphere pressure, and at 6 ata, the latter corresponding to 50 meter diving depth. A frequency standard tone of 1000 c/s was recorded at each pressure level as a means of ensuring reliable frequency calibration.

Four subjects spoke a list of CV nonsense syllables comprising all possible combinations of C = [b], [d], [g], [m], [n], [l], [v], and V = [o:], [a:], [i:], [e:], [æ:] together with a sentence "I dag vill jag vila på min ö".

Representative spectrograms at 1 and 6 ata are shown in Figs. II-6 to II-9. From these it is apparent that at 6 ata F₁ is confined to a frequency range above a lower limit of 400-500 c/s and that the dynamic range of variation of F₁ is much restricted. Voiced consonants and the vowels [i] and [e] thus obtain almost the same F₁ at very high air-pressures.

Formant frequencies for three of the subjects are tabulated below.
Fig. II-6. Spectrograms of syllables [va], [ve], [vi] uttered in a decompression tank at normal atmospheric pressure, 1 ata (above) and 6 ata (below).
Fig. 11-7. Spectrograms of syllables [ma], [me], [mi] at 1 and 6 ata pressure.
Fig. II-8. Spectrograms of syllables [go:], [ge:], [gi:] at 1 and 6 ata pressure.
Fig. II-8. Spectrograms of syllables [go:], [ge:], [gi:] at 1 and 6 ata pressure.
Fig. II-9. Spectrogram of a sample of connected speech "Idag vill jag vila" at 1 and 6 ata.
## TABLE II-1

Formant frequencies at 1 and 6 ata

<table>
<thead>
<tr>
<th>Vowel</th>
<th>Subject</th>
<th>( F_1 ) 1 ata</th>
<th>6 ata</th>
<th>( \Delta )</th>
<th>( F_2 ) 1 ata</th>
<th>6 ata</th>
<th>( \Delta )</th>
<th>( F_3 ) 1 ata</th>
<th>6 ata</th>
<th>( \Delta )</th>
</tr>
</thead>
<tbody>
<tr>
<td>[o:]</td>
<td>Ru</td>
<td>385</td>
<td>495</td>
<td>+110</td>
<td>694</td>
<td>780</td>
<td>+85</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ni I</td>
<td>430</td>
<td>560</td>
<td>+130</td>
<td>760</td>
<td>850</td>
<td>+90</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ga</td>
<td>390</td>
<td>530</td>
<td>+140</td>
<td>650</td>
<td>735</td>
<td>+85</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[ɛ:]</td>
<td>Ru</td>
<td>450</td>
<td>550</td>
<td>+70</td>
<td>1990</td>
<td>2105</td>
<td>+115</td>
<td>2550</td>
<td>2660</td>
<td>+110</td>
</tr>
<tr>
<td></td>
<td>Ni I</td>
<td>495</td>
<td>605</td>
<td>+110</td>
<td>1965</td>
<td>2005</td>
<td>+40</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ni II</td>
<td>460</td>
<td>590</td>
<td>+130</td>
<td>1925</td>
<td>1985</td>
<td>+60</td>
<td>2555</td>
<td>2580</td>
<td>+25</td>
</tr>
<tr>
<td></td>
<td>Ga [ɛ:]</td>
<td>575</td>
<td>705</td>
<td>+130</td>
<td>1730</td>
<td>1800</td>
<td>+70</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[ɔ:]</td>
<td>Ru</td>
<td>425</td>
<td>550</td>
<td>+125</td>
<td>1660</td>
<td>1655</td>
<td>-5</td>
<td>2325</td>
<td>2375</td>
<td>+50</td>
</tr>
<tr>
<td></td>
<td>Ni I</td>
<td>415</td>
<td>550</td>
<td>+135</td>
<td>1670</td>
<td>1685</td>
<td>+15</td>
<td>2500</td>
<td>2560</td>
<td>+60</td>
</tr>
<tr>
<td></td>
<td>Ni II</td>
<td>420</td>
<td>550</td>
<td>+130</td>
<td>1530</td>
<td>1587</td>
<td>+40</td>
<td>2490</td>
<td>2530</td>
<td>+40</td>
</tr>
<tr>
<td></td>
<td>Ga</td>
<td>415</td>
<td>550</td>
<td>+135</td>
<td>1775</td>
<td>1745</td>
<td>-30</td>
<td>2360</td>
<td>2415</td>
<td>+55</td>
</tr>
<tr>
<td>[ɑ:]</td>
<td>Ru</td>
<td>650</td>
<td>735</td>
<td>+85</td>
<td>960</td>
<td>1110</td>
<td>+150</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ni I</td>
<td>565</td>
<td>705</td>
<td>+140</td>
<td>925</td>
<td>942</td>
<td>+67</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ga</td>
<td>535</td>
<td>665</td>
<td>+130</td>
<td>920</td>
<td>955</td>
<td>+35</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[e:]</td>
<td>Ru</td>
<td>375</td>
<td>485</td>
<td>+110</td>
<td>2220</td>
<td>2315</td>
<td>+95</td>
<td>2630</td>
<td>2800</td>
<td>+170</td>
</tr>
<tr>
<td></td>
<td>Ni I</td>
<td>360</td>
<td>505</td>
<td>+145</td>
<td>2200</td>
<td>2375</td>
<td>+175</td>
<td>2690</td>
<td>2770</td>
<td>+80</td>
</tr>
<tr>
<td></td>
<td>Ga</td>
<td>390</td>
<td>545</td>
<td>+155</td>
<td>2045</td>
<td>2055</td>
<td>+10</td>
<td>2560</td>
<td>2540</td>
<td>-20</td>
</tr>
<tr>
<td>[i:]</td>
<td>Ru</td>
<td>330</td>
<td>475</td>
<td>+145</td>
<td>2280</td>
<td>2330</td>
<td>+50</td>
<td>3230</td>
<td>3300</td>
<td>+70</td>
</tr>
<tr>
<td></td>
<td>Ni I</td>
<td>305</td>
<td>500</td>
<td>+195</td>
<td>2345</td>
<td>2383</td>
<td>+38</td>
<td>3050</td>
<td>3160</td>
<td>+110</td>
</tr>
<tr>
<td></td>
<td>Ga</td>
<td>330</td>
<td>470</td>
<td>+140</td>
<td>2280</td>
<td>2270</td>
<td>-10</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Ru = Rundblom    Ni = Nilsson    Ga = Garner
Samples were taken at the middle or terminal part of the vowel whichever seemed more stationary in formant frequency pattern. Data for all consonantal environments have been averaged in Table II-1. The data on $F_3$ have been excluded in instances where they were judged to be less reliable. The series Ni I and Ni II pertain to the subject Ni at two different occasions.

There are typical trends to be observed. The $F_1$-shift is with few exceptions greater in magnitude than the shifts in $F_2$ and $F_3$. In vowels $[\phi]$ and $[i]$ the observed $F_2$-shifts are of the same order of magnitude as the standard deviation, 50 c/s, or less. It is interesting to note that the distinction between the $F_1$ of $[i]$ and $[e]$ tended to be eliminated at the high pressure for subjects Ni and Ru and that accordingly the auditive distinction between these phonemes was almost lost.

The typical distortion of back vowels $[o]$ and $[\alpha]$ at high pressures is that of a raise in both $F_1$ and $F_2$ but generally more in $F_1$ so that $F_1$ comes rather close to $F_2$. These attributes account for the especially apparent nasal quality of the vowel $[\alpha]$. Two of the four speakers showed a moderate increase in voice fundamental frequency $F_0$.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Average $F_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 ata</td>
</tr>
<tr>
<td>Ru</td>
<td>124</td>
</tr>
<tr>
<td>Ni</td>
<td>114</td>
</tr>
<tr>
<td>Ga</td>
<td>127</td>
</tr>
<tr>
<td>An</td>
<td>187</td>
</tr>
</tbody>
</table>

Other general observations of speech at high pressures are an increase in overall sound pressure level of voiced sounds and a relative loss of spectrum intensity level at high frequencies. There is a pronounced weakening of the energy of all unvoiced consonants especially of the burst interval of stop sounds. No actual measurements of formant bandwidths were undertaken but the general observation from the broad-band spectrograms was that bandwidths did not increase except in the low $F_1$-range. The speaking tempo was substantially reduced. Some but not all of the
speakers showed typical signs of disturbed voice source mechanism in terms of a randomization of subsequent pitch pulse positions and a general "noisiness" superimposed on the spectrogram.

3. Theory

It is of interest to note that observations of quite opposite signs have been made on speech at high altitudes. K.C. Clark et al (2) reported a reduction of the free field sound pressure level of voiced sounds of the order of 10 dB at 35000 feet altitude, whilst the sound pressure level of unvoiced consonants increased 5 dB and thus gained 15 dB relative to voiced sounds. A lack of nasality was pronounced.

Thus there is a consistent set of variations of observed acoustic characteristics as a function of the air-pressure, i.e. the density from deep underwater to high altitude conditions.

a. Frequency shifts

Since density $\rho$ is proportional to pressure $P$ the velocity of sound

$$c = \sqrt{\frac{\gamma \cdot P}{\rho}}$$

is dependent on $\gamma$, the ratio of specific heats at constant pressure and volume, only. A change of $P$ from 1 ata to 6 ata is associated with $+0.7\%$ change in $\gamma$ (1.407 to 1.417), and thus merely $0.35\%$ in $c$.

Since all equations for calculating resonance frequencies of an arbitrarily complex vocal cavity system contain the factor $c$ and in addition functions of cavity dimensions only, it is obvious that the detuning of cavity resonances as a result of a change in air pressure enters through $c$ alone and is insignificantly small. Thus the expression for the resonance frequency of a Helmholtz resonance is

$$F = \frac{c}{2\pi} \sqrt{\frac{A}{l_e V}}$$

where $V$ is the volume, $A$ the cross-sectional area of the neck and $l_e$ its effective length. The frequencies of standing wave
resonances in a tube terminated differently at the two ends (open circuit at one end and short circuit at the other) are

\[ F_n = \frac{c}{4l_e} (2n-1) \]  

(3)

and when terminated equally at both ends,

\[ F_n = \frac{c}{2l} \cdot n \]  

(4)

Now to the effect of a finite cavity wall impedance. It was originally postulated by van den Berg (1) that the soft parts of the vocal cavity walls behave like a mass element to the first approximation with a resistive element to account for dissipation. In Fig. 11-10 the equivalent network elements of the cavity walls are denoted \( L_w \) and \( R_w \). The radiation impedance is denoted \( R_0 \).

Assuming that the vibrating walls occupy an area of \( A_w = 50 \text{ cm}^2 \) along a pharynx length of 8 cm and an internal volume of 80 cm\(^3\) representative of a palatal tongue position, the pharynx wall inductance is

\[ L_w = \frac{\rho_w \cdot 1}{A_w} \]  

(5)

where \( \rho_w \approx 1 \text{ g/cm}^3 \) is the density and \( l_w = 1 \text{ cm} \) is the average thickness of the walls.

At complete closure of the vocal tract the mass element \( L_w \) resonates with the capacitance \( C \) of the entire air volume. The limiting resonance frequency is thus

\[ F_{1w} = \frac{1}{2\pi} \sqrt{\frac{1}{L_w C}} \]  

(6)

where \( L_w = \frac{1}{A_w} \) as assumed above, and

\[ C = \frac{V}{\rho c^2} \]  

(7)

At a pressure of 1 ata and normal speaking conditions \( c = 35000 \text{ cm/sec} \) and \( \rho = \rho_1 = 1.2 \cdot 10^{-3} \text{ g/cm}^3 \). At a pressure of \( P \) ata the density is \( \rho = P \cdot \rho_1 \) and the limiting value of \( F_1 \) is
which amounts to

\[ F_{1w} = \sqrt{P} \cdot 150 \text{ c/s} \]  

(9)

Thus at \( P = 6 \text{ ata} \) the fundamental resonance \( F_1 \) must exceed 370 c/s, which conforms with our observations. It is also known that the "voice bar" \( F_v \) of voiced consonants never goes below 150 c/s in normal speech.

The first formant of a voiced consonant or of a close or half-open vowel is apparently tuned by \( L_w \) in parallel with \( L_1 \). If, for example, \( L_1 = L_w \), \( F_1 \) would equal \( \sqrt{2} \cdot F_{10} = 210 \text{ c/s} \). In general denoting the resonance of the system with \( L_w \) excluded as \( F_{1a} \), and \( F_1 \) with due respect to both \( L_1 \) and \( L_w \), and \( F_{1w} \) with \( L_1 \) excluded there holds the relation

\[ F_{1}^2 = F_{1w}^2 + F_{1a}^2 \]  

(10)

At a pressure of \( P \) this may be written

\[ F_{1P}^2 = P \cdot F_{1w(P=1)}^2 + F_{1a}^2 \]  

(11)

By combining equations of this type for \( P = 1 \) and \( P = 6 \text{ ata} \) we obtain

\[ F_{1w(P=1)}^2 = (F_{16}^2 - F_{11}^2)/P \]  

(12)

which relates the fundamental resonance of the closed vocal tract at 1 ata \( F_{1w(P=1)} \) to the observed frequencies of the first formant \( F_{16} \) at \( P = 6 \text{ ata} \), and \( F_{11} \) at \( P = 1 \text{ ata} \).

If the theory holds it should be possible to calculate reasonable values of \( F_{1w} \) at 1 ata from the observed frequency shifts. Calculations on our material gave the following data:
TABLE II-2.

Limiting first formant frequency at 1 ata in c/s

<table>
<thead>
<tr>
<th>Vowel</th>
<th>Ru</th>
<th>Ni I</th>
<th>Ni II</th>
<th>Ga</th>
<th>Mean value</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ɔ:]</td>
<td>138</td>
<td>160</td>
<td>160</td>
<td>153</td>
<td></td>
</tr>
<tr>
<td>[ɛ:]</td>
<td>141</td>
<td>157</td>
<td>167</td>
<td>183</td>
<td>162</td>
</tr>
<tr>
<td>[œ:]</td>
<td>155</td>
<td>161</td>
<td>157</td>
<td>160</td>
<td>157</td>
</tr>
<tr>
<td>[ɑ:]</td>
<td>155</td>
<td>190</td>
<td>155</td>
<td>156</td>
<td></td>
</tr>
<tr>
<td>[œ:]</td>
<td>141</td>
<td>151</td>
<td>170</td>
<td>156</td>
<td></td>
</tr>
<tr>
<td>[i:]</td>
<td>153</td>
<td>178</td>
<td>150</td>
<td>160</td>
<td></td>
</tr>
</tbody>
</table>

Mean value 147 168 162

Of the sounds tabulated above it is only [œ], [ɛ], [œ], and [i] that should permit an application of the simple Helmholtz resonator model for $F_1$. The mean value of $F_{1w}(F=1)$ is 160 c/s for these vowels. Similar calculations performed on the earlier material from 1962 gave the following results:

<table>
<thead>
<tr>
<th>Subject</th>
<th>$F_{1w}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>G.J.</td>
<td>180</td>
</tr>
<tr>
<td>B.B.</td>
<td>150</td>
</tr>
<tr>
<td>H.K.</td>
<td>165</td>
</tr>
<tr>
<td>G.E.</td>
<td>180</td>
</tr>
</tbody>
</table>

$F_{1w}$ the low frequency limit of $F_1$ in normal speech thus varies only little with the particular speaker.

What about $F_2$ and higher formants? Can the effect of the atmospheric pressure be calculated in this frequency range?

Fig. II-10.A shows the necessary modification of the transmission line representation of the vocal tract with a distributed shunting inductance $L_s(x)$ per unit length, paralleling the distributed capacitance $C$ of the inhomogeneous line. The effect of $L_s$ can be included as a frequency dependent modification $C_s$ of the distributed capacitance $C = C(1 - 1/\omega^2 L_s C)$ or as a modification of the speed of sound propagation down the line.
Fig. II-10. A. Transmission line analog of the vocal tract with distributed inductance $L(x)$, capacitance $C(x)$, and wall inductance $L_w(x)$ per unit length of the resonator at a coordinate $x$.

B. Helmholtz resonator with total inductance $L_w$ and resistance $R_w$ of the cavity walls included. $R_0$ is the radiation resistance.

C. Equivalent circuit of B for calculation of $F_1$ of voiced consonants and close front vowels.
where $w_w$ is the cutoff frequency

$$w_w^2 = 1/L_s c$$  \hfill (14)

With notations $A$ for cross-sectional area of the resonator and $A_s$ for the area per unit length of the vibrating wall of thickness $d_s$ and density $\rho_s$

$$\begin{cases} L_s = d_s \rho_s / A_s \\ C = A / \rho c^2 \end{cases}$$  \hfill (15)

thus

$$w_w^2 = \frac{A_s}{A} \cdot \frac{\rho c^2}{d_s \cdot \rho_s} = \frac{A_s}{A} \cdot \frac{\rho}{d_s} \cdot \frac{c^2}{\rho_s}$$  \hfill (16)

The cutoff frequency $w_w$ is proportional to the density of the air $\rho$, or to the pressure $P$ in at. Under the special circumstances of a uniform distribution of the wall impedance along a single tube model it is found that $w_w = 2\pi F_1^w$.

As a specific example assume a pharynx length of 8 cm and a high prepalatal articulation in which case $F_2$ of [i] could be approximately calculated as

$$F_2 = \frac{c}{21} = \frac{25000}{16} \cdot \frac{1}{(1-1.16/2200)^2}$$  \hfill (17)

The cavity wall correction factor at 1 ata is 0.26 % which is negligible and 1.6 % or 37 c/s at $P = 6$ at. The calculated difference $37 - 6 = 31$ compares well with the measured data.

As another example the vocal tract will be considered as a single homogeneous tube loaded with distributed wall inductance and $w_w = 2\pi \cdot 160 \sqrt{P}$ as before. Assuming a frequency of the first formant $F_1 = c/4d_e = 500$ c/s the addition of $L_s$ causes a shift of $\Delta F_1 = 135$ c/s at 6 at which also reflects the correct order of magnitude according to measurements.

In short the effects described above are attributable to the high density $\rho$ increasing the characteristic impedance level $\rho c/A$ of the vocal cavity system at high air-pressures thus making this system more susceptible to the shunting effects of the cavity
walls, the latter being independent of the air-pressure. The increase of the air column load in series with the mechanical impedance of the vocal folds could account for the phonatory source disturbances observed.

There remains to map the relative conductivity of the cavity walls. Thin flesh obviously has a more severe shunting effect than bony and thick structures. One would thus primarily conceive of the sides of the throat and the cheeks of the mouth to permit vibrations but also the soft velum itself. The relative influence of a shunt on the frequency of vocal resonance is also dependent on the spatial distribution of sound pressure in the vocal tract. In the frequency range of the first formant a cavity wall shunt is thus the more effective the closer it comes to the glottal end of the system. In general the effect of a shunt is large wherever the sound pressure is high.

b. Formant bandwidths

A theoretical study of the extent to which the density of air enters expressions of formant bandwidth can be made from ref. (3), pp. 300-310. The conclusion is that bandwidths are always related to expressions of the form \( \frac{R}{\omega L} \) or \( \frac{1}{\omega RC} \). Acoustic inductance as well as resistance, even the radiation resistance, are proportional to density and acoustic capacitance is inversely proportional to density. Thus density is cancelled out in all expressions above.

Energy losses associated with vocal cavity vibrations are of a greater interest. Recently Fujimura has determined the closed condition resonance \( F_1 \) of the vocal tract experimentally and found frequencies of the order of 150-200 c/s and bandwidths of the order of 100 c/s. The bandwidth is attributed to a resistance element \( R_w \) in series with the wall inductance \( L_w \) as in Fig. 11-10 and we thus conclude that

\[
\frac{R_w}{\omega L_w} = 100
\]

At formant frequencies of interest \( \omega L_w > R_w \) and we compute a
With a finite mouth opening of inductance $L_1$ parallel to $L_w$ we have approximately

$$R_p = \frac{(\omega L_w)^2}{R_w}$$  \hspace{1cm} (19)

But

$$B_1 = \frac{1}{2\pi R_w C} = \frac{R_w}{2\pi (\omega L_w)^2 C}$$  \hspace{1cm} (20)

$$\frac{L_1 L_w}{L_1 + L_w} C = \frac{1}{\omega^2} = \frac{1}{4\pi^2 F_1^2}$$  \hspace{1cm} (21)

$$B_1 = B_w \cdot \frac{1}{(1 + L_w/L_1)}$$  \hspace{1cm} (22)

The contribution of cavity wall dissipation to first formant bandwidth $B_1$ is thus inversely proportional to the square of frequency. If for example $L_1 = L_w$ and $F_1 = \sqrt{2} \cdot F_{10}$ which is of the order of 225 c/s, the bandwidth contribution is $B_1 = B_w/2$ or of the order of 50 c/s. The formant data of Fujimura, ref. (4), confirms this frequency dependency of $B_1$. Normally $B_1(f)$ has a minimum at $F_1 = 500$ c/s of about $B_1 = 30$ c/s.

At the higher atmospheric pressures the only significant change of the equivalent circuit is the decrease of the capacitance of the air and thus the increase in resonance frequency, whilst the bandwidth is that of the corresponding resonance at the lower pressure, i.e., lower frequency. At $L_1 = L_w$ and $P = 6$ ata assuming $F_{1w} = 160$ we thus have $F_{16} = \sqrt{6} \cdot \sqrt{2} \cdot 160 = 555$ and $B_{16} = B_w/2 = 50$ c/s. In addition there is the radiation and frictional damping adding some 10-20 c/s more to $B_1$. Spectrograms confirm the theory. At 6 ata the voice bar $F_1$ of voiced consonants is a frequency transposed replica of the 160 c/s voice bar at 1 ata with the weak and thin voicing striations typical of a large bandwidth formant.

c. Sound pressure levels

A positive shift of a formant of constant bandwidth is associated with an increase of its $Q$ and of the spectrum level at
frequencies above $F_1$. This is one factor contributing to an increase of sound pressure levels at higher ambient pressures. It affects primarily voiced sounds of a narrow articulation.

There are also reasons for expecting an overall increase of the intensity of voiced sounds because of more efficient radiation. The basis for this consideration is a phonation at constant subglottal overpressure $\Delta p$. According to ref. (3), p. 268, the volume velocity of the source is

$$U_q(t) = A(t) \sqrt{2\Delta p/\rho}$$

(23)

To this we add the vocal tract transfer function $H(\omega) = \frac{U_q(\omega)}{U_q'(\omega)}$ and the radiation transfer $p_1(\omega)/U_0$ relating sound pressure at 1 cm in front of the speaker to the volume velocity at the lips.

$$\frac{p_1(\omega)}{U_0} = \frac{g(\omega)}{4\pi l}$$

(24)

Thus the sound pressure $p_1(\omega)$ is proportional to $\sqrt{\rho}$ or to $\sqrt{\rho}$. A factor of $\sqrt{\rho} = \sqrt{5}$ from $P = 1$ to 6 ata, would thus cause an increase of the sound pressure level of $\sqrt{6}$ or 8 dB. Our experimental data from HMS Belos showed an increase of on the average 5 dB, but this is not conclusive since the microphone distance was not sufficiently well controlled. Sounds of low $F_1$ gained much more than sounds of high $F_1$ as could be conceived from theory.

d. Unvoiced sounds

What is the reason for the decreased relative level of stop bursts and fricatives at high ambient pressures? Here the theory is not so well developed but it may be of interest to refer to p. 273 of ref. (3). The sound pressure of fricatives is proportional to the square of the velocity of the generating air stream in a vocal tract constriction. The relation of this air velocity $u = U_0/A$ to the pressure drop $\Delta p$ is

$$\Delta p = \frac{n}{2} \left( \frac{U_0}{Ak} \right)^2$$

(25)

where $k$ is a constant. Again assuming a constant $\Delta p$ the $U_0^2$ is
inversely proportional to $\rho$. This effect is probably counteracted by the $\rho$ proportionality of the radiation transfer, eq. (24). At high ambient air-pressures there remains a decrease of the sound pressure level of fricatives and stops relative to the average level of voiced sounds.

G. Fant and B. Sonesson

References:

(1) van den Berg, Jw.: Physics van do stemvorming met toepassingen, thesis Univ. of Groningen ('s-Gravenhage 1953).


*Department of Anatomy, University of Lund, Sweden, and the Office of the Surgeon General, Naval Staff, R.S.N., Stockholm, Sweden.