Dept. for Speech, Music and Hearing
Quarterly Progress and Status Report

Pressure and gas mixture effects on diver's speech

Fant, G. and Lindqvist-Gauffin, J.

journal: STL-QPSR
volume: 9
number: 1
year: 1968
pages: 007-017

KTH Computer Science and Communication

II. STUDIES RELATED TO DIVER'S SPEECH

A. PRESSURE AND GAS MIXTURE EFFECTS ON DIVER'S SPEECH

G. Fant and J. Lindqvist

Summary

The present study is a follow-up of that of Fant and Sonesson (1964). According to the theory outlined in the previous work the acoustic distortion of diver's speech can be derived from two factors. One is the pressure or rather density factor which causes a non-linear shift of low-frequency vocal resonances, subjectively perceived as "nasality". This effect originates from the participation of vocal cavity walls in vocal resonances. The other factor is the well-known linear transposition of vocal resonances in proportion to the velocity of sound for the particular gas mixture (Donald Duck-effect).

A number of experimental conditions have been investigated. In the pressure chamber of the Swedish Navy in Stockholm we have studied speech at a depth of 100 meter in air and also under the condition of the subjects breathing a helium-oxygen mixture through a mask. A supplementary material of tape-recorded helium speech at various depths was received from the Communication Sciences Laboratory of the University of Florida in Gainesville. This material was especially valuable since it represented speech without mask in a helium filled tank. Experimental data are in substantial agreement with theory. It is accordingly possible to predict the spectral distortion of a diver's speech under various conditions of depth and gas mixture. For any given gas mixture there exists an optimum diving depth for cancellation of the non-linear shifts of low-frequency formants.

The effects of high air pressures

A detailed account for the effects of high air pressures on speech was given by Fant and Sonesson (1964). These effects are primarily related to vocal cavity wall vibrations and can be derived by adding a distributed inductance shunting the equivalent network of the air passages. The size of this inductance is determined by the mass distribution of the walls in the first place and independent of external pressures and gas mixtures. The characteristic impedance of the air passages, $\rho c/A$, on the other hand, increases with the density ρ, i.e. with pressure. Thus at high pressures the mismatch between cavity wall impedance and gas impedance is reduced.

The effect of cavity wall vibrations is to increase the radiation of sound from the skin of the head and throat and to add reactive elements participating in the tuning of vocal resonances. The latter effect is the main cause of distortion. Another effect caused by a rise in pressure is that voiced sounds gain in amplitude over unvoiced sounds which also is predictable from theory.

The detuning of vocal resonances due to cavity wall vibrations can be calculated from a knowledge of the closed tract fundamental resonance. The total volume of air V has a capacitance

$$C_t = V/\rho c^2 \qquad (1)$$

which resonates with the net value of the vibrating wall inductance L_w

$$\frac{1}{L_w} = \sum \frac{1}{L_{wd}} \qquad (2)$$

where L_{wd} is the distributed inductance along the tract. The closed tract resonance frequency is accordingly

$$f_w = \frac{1}{2\pi} \sqrt{\frac{1}{C_t \cdot L_w}} = \frac{c}{2\pi}\sqrt{\frac{\rho}{L_w \cdot V}} \qquad (3)$$

which is of the order of f_w = 150-200 Hz in air at normal atmosphere pressure.

The density ρ is proportional to the atmospheric pressure P

$$\rho = \rho_o \cdot \frac{P}{P_o} \qquad (4)$$

where ρ_o is the density at the pressure P_o. The velocity of sound c is practically independent of the pressure of the gas.

Thus at increasing pressure and constant gas mixture the closed tract resonance becomes proportional to the square root of the pressure. In air

$$F_w = F_{w_o} \cdot \sqrt{P} \qquad (5)$$

In air at 100 meter depth and P = 11 atmospheres, F_w is of the order of 500-660 Hz. Evidently F_w sets the lower bound on the first formant frequency $F_1 \geq f_w$.

F_1 can be calculated from a lumped element circuit with L_w shunting C_t together with an inductance L_i of a finite mouth opening.

$$F_1 = \frac{1}{2\pi}\sqrt{\frac{L_i+L_w}{CL_i \cdot L_w}} = (F_i^2 + F_w^2)^{1/2} \qquad (6)$$

where

$$F_i = \frac{1}{2\pi}(L_i C_t)^{-1/2} \qquad (7)$$

is the ideal F_1 without the wall shunt. One can also conceive of the distributed inductive wall shunt L_{wd} as a decrease of the distributed capacitance of the gas from C_d to the effective value C_{de}

$$\left(\frac{1}{\omega C_d} - \omega L_{wd}\right) = \frac{1}{\omega C_{de}} \qquad (8)$$

For a single tube resonator

$$L_{wd} C_d = L_w C_t \qquad (9)$$

and the effective velocity of sound

$$c_e = (L_d C_{de})^{-1/2} = c_o(1-f_w^2/f^2)^{-1/2} \qquad (10)$$

the quarter wavelength resonance, i.e.

$$F_1 = \frac{c_e}{4\ell} = \frac{c_o(1-f_w^2/F_1^2)^{-1/2}}{4\ell} \qquad (11)$$

is accordingly

$$F_1 = \left[\left(\frac{c_o}{4\ell}\right)^2 + F_w^2\right]^{1/2} \qquad (12)$$

which is principally the same as Eq. (6) for the Helmholtz resonator resonance.

Since any formant frequency is proportional to c_e we can write

$$F_n = (F_{ni}^2 + F_w^2)^{1/2} \qquad (13)$$

where F_{ni} is the ideal formant frequency neglecting the wall shunt. This formula is only valid on an average basis since the wall shunt is not likely to be uniformally distributed.

Averaging the data of Fant and Sonesson (1964) from four males speaking at pressures corresponding to 0 meter and 50 meter, i.e. P = 1 and P = 6 atm provides a most convincing illustration of the validity of Eq. (13) as shown in Fig. II-A-1. At P = 1, F_w = 160 Hz and at P = 6,

$$F_w = (P)^{1/2} \cdot 160 = 390 \text{ Hz} \qquad (14)$$

The data are samples of F_1 F_2 and F_3 from six vowels in monosyllabic test words. A similar plot of one speaker's formant frequency shifts from 0 to 100 meter equivalent depth in air is shown in Fig. II-A-2. These data pertain to a more recent study. The spread is greater but the distribution fits the predicted line of Eq. (13) with $F_w = \sqrt{11} \cdot 200 = 600$ Hz. In the region of F_3 2000-3000 Hz the measured points lie about +4 % above the calculated values.

One aspect of the non-linear formant shifts at high pressures is that some of the contrast between close and more open voiced sounds is lost. As seen in Fig. II-A-2 the F_1-range from 200-400 Hz at 0-level is transposed to 660-750 Hz at the 100 meter level. The reduced F_1-range is accompanied by a reduced amplitude contrast of voiced consonants versus adjacent vowels, as follows from the general rules relating formant frequencies and formant levels, Fant (1960).

At 100 meter in air unvoiced consonants are reduced about 10 dB in level relative vowels. According to the theory developed by Fant and Sonesson (1964) this effect originates from a $(P)^{1/2}$ proportional increase in the amplitude of voiced sounds whereas unvoiced sounds are not expected to be influenced by pressure changes. The same effects holds true of high-altitude low pressure (P < 1) conditions where voiced sounds are known to be reduced in level versus unvoiced sounds.

The theoretical reasoning goes as follows. Assume a constant glottal pressure drop Δp independent of density. With notation A(t) for time-varying glottal area the glottal volume velocity flow is

$$U_q(t) = A(t) \sqrt{2\Delta p/\rho} \quad , \qquad (15)$$

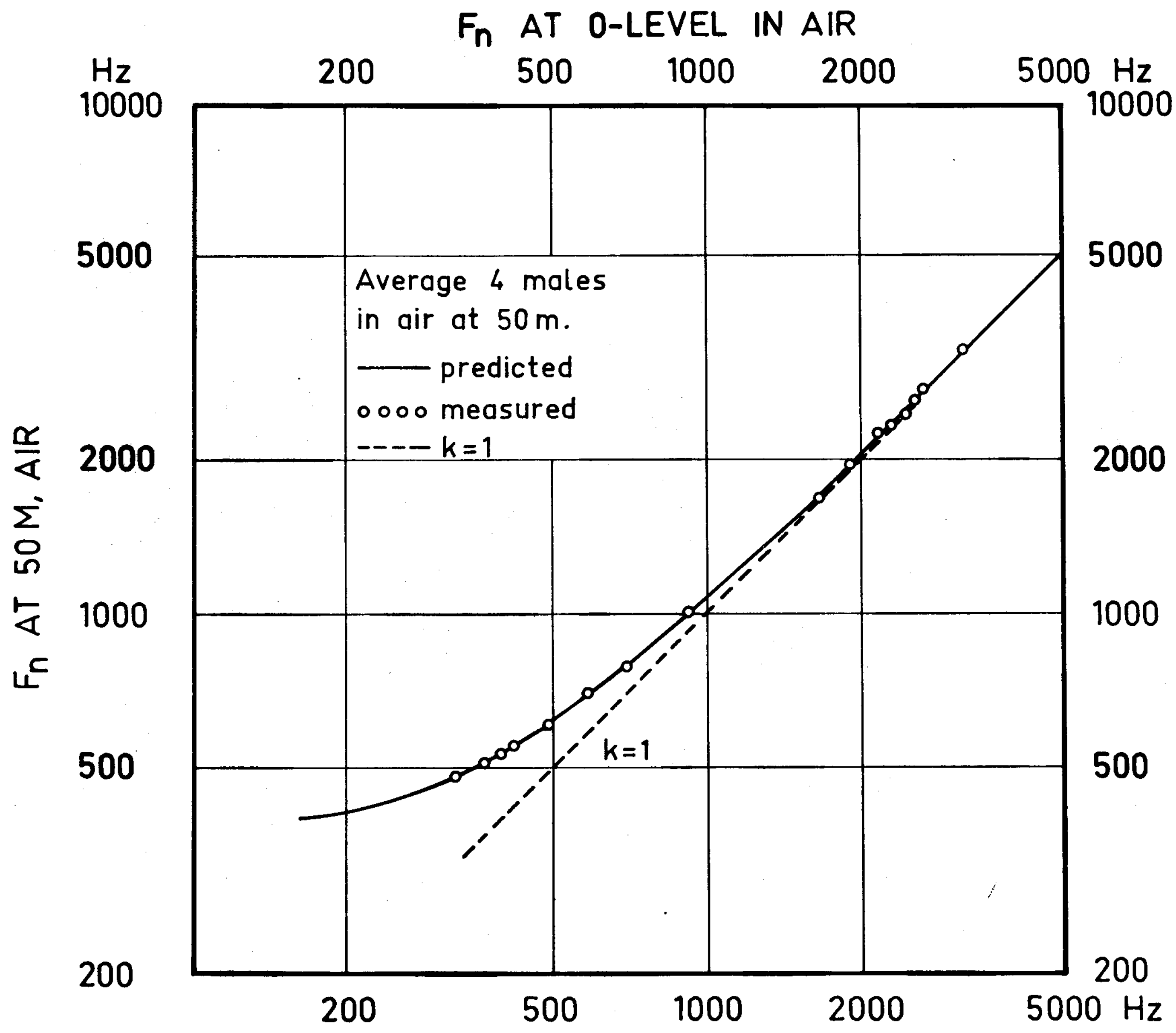

Fig. II-A-1. Predicted and measured formant frequency transpositions in air at 50 meters equivalent depth. Average of four subjects.

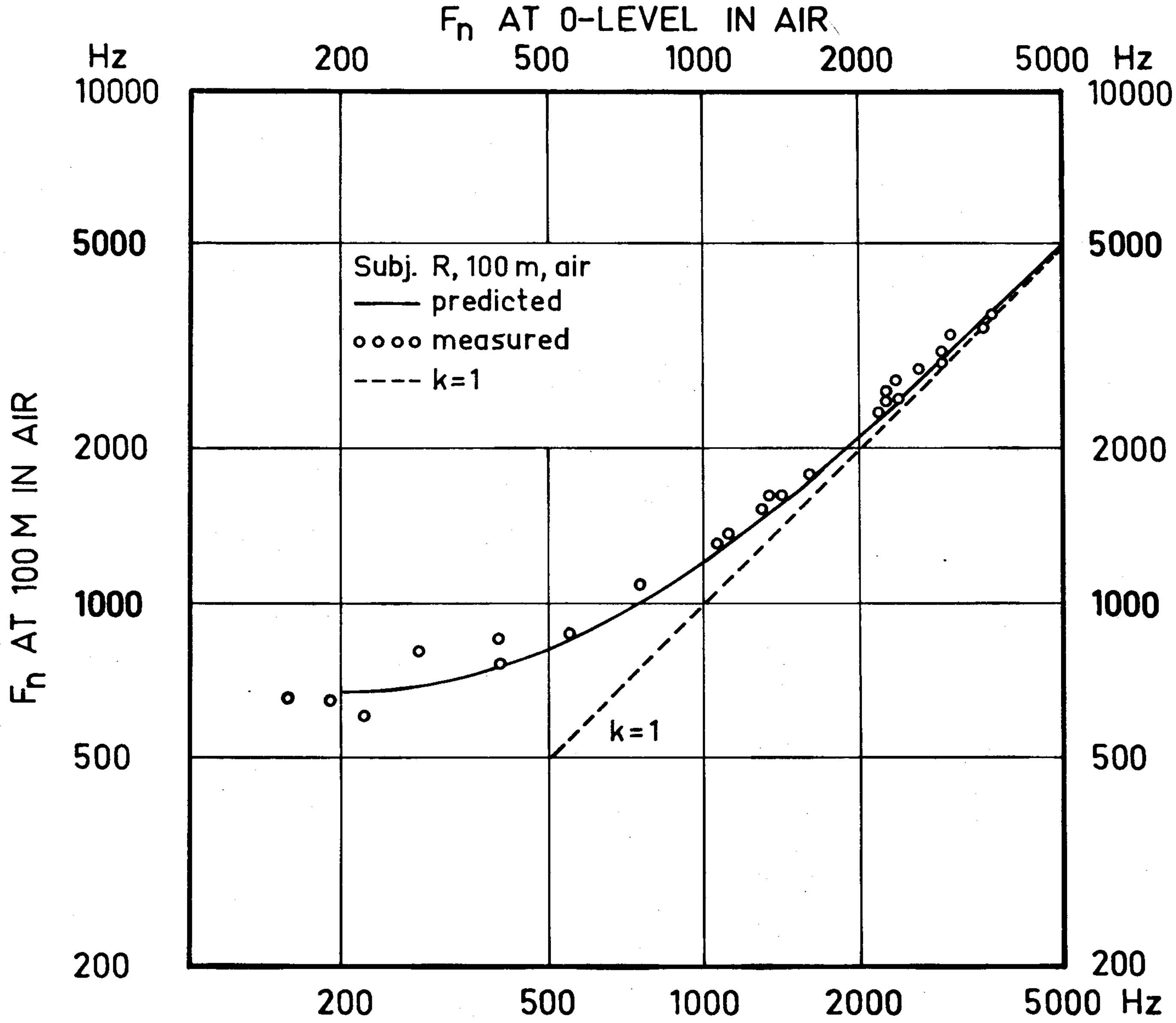

Fig. II-A-2. Predicted and measured formant frequency transpositions in air at 100 meters equivalent depth. One subject (R).

see Fant (1960). The transfer $H(\rho) = U_o(\omega)/U_q(\omega)$ from glottis to the lips is independent of ρ except for the cavity wall effects earlier discussed. The radiation transfer

$$\frac{P_\ell(\omega)}{U_o(\omega)} = \frac{\rho\omega}{4\pi\ell} \qquad (16)$$

relating sound pressure at a distance ℓ to the volume velocity at the lips introduces a ρ-proportionality. The net effect is thus

$$P_\ell(\omega) \sim \sqrt{\rho} \sim \sqrt{P} \qquad (17)$$

The sound pressure of fricatives and other unvoiced sounds is proportional to the square of the particle velocity U_c/A of the generating air stream at the constriction,

$$\Delta p = \frac{\rho}{2}\left(\frac{U_c}{A_k}\right)^2 \qquad (18)$$

At constant pressure drop Δp the output flow of unvoiced sound at the lips is

$$U_o(\omega) \sim U_c^2 \sim \frac{1}{\rho} \qquad (19)$$

This $1/\rho$ factor is cancelled by the ρ-proportionality of radiation.

The bandwidth of vocal resonances is not radically changed with increasing pressure. Low F_1-formants of ordinary speech are largely damped by losses from cavity wall vibrations. At high pressures the bandwidth B_1 of the transposed formant is about the same as that of the same formant under normal pressure or somewhat higher. The frequency shift of a low F_1-formant is thus accompanied by an increase in its $Q = F_1/B_1$. Thus close (low F_1) vowels may obtain a gain in amplitude due to pressure transposition in addition to that of Eq. 17 providing the Q-gain is not compensated by a lower source spectrum level at the higher F_1. This is generally the case below F_1 = 600 Hz. However, an initially low F_2 high F_1-vowel as [ɑ] may gain in overall sound intensity by a decreased F_2/F_1 ratio which increases the amplitude of each formant and adds to the "nasal" quality.

The $(P)^{1/2}$ proportional increase of the level of voiced sounds should be considered when designing audio links for large diving depths. A tape-recorder adjusted for optimal input level at P = 1 ata may be severely overloaded at P > 10 ata. The gain in voice level might also be a point to consider in speaker training since the voice effort might be reduced accordingly. To the normal behavior of the subject in air at high pressures is a markedly reduced tempo of speaking. Phonation and pauses are prolonged about 100 % at 11 ata which aids the intelligibility. This is probably not a learned behavior. It could be related to the increase of glottal flow resistance in proportion to $(P)^{1/2}$ at constant glottal pressure drop, see Eq. (15), or it could be related to the oxygen partial pressure.

One question we posed was whether the vocal tract cavity wall vibrations could be damped out by placing a suitable protection externally on the subject's throat. A segmented lead sheet was worn by the subject in order to mass-load the walls of his throat and chins and thus lower his closed tract resonance F_w. However, this experiment was a failure. The articulation scores of our rhyme-type word lists, which were of the order of 50 %, did not improve significantly and F_w was not affected either. The reason for the failure may be incomplete coupling between the throat tissue and the lead and the existance of other competing wall shunts, e.g. between the soft palate and the nasal cavities.

Helium speech

In a rigid resonator the frequency of the resonances are proportional to the velocity of sound. The equation for the velocity of sound in an ideal gas (PV = RT) is

$$c = \sqrt{\frac{\gamma \cdot P}{\rho}} \qquad (20)$$

where

$$\gamma = \frac{c_p}{c_v}$$

P = pressure

ρ = density

c_p = specific heat at constant pressure

c_v = specific heat at constant volume

The relation of the velocity of sound in the He-mixture c to that of the air c_o of the same pressure is then:

$$k = \frac{c}{c_o} = \frac{f}{f_o} = \sqrt{\frac{\gamma \cdot \rho_o}{\rho \cdot \gamma_o}} \qquad (21)$$

Calculating the gamma for a gas mixture involves the weighted mean of c_p over the weighted mean of c_v.

$$\gamma = \frac{\sum (p_i \cdot m_i \cdot c_{pi})}{\sum (p_i \cdot m_i \cdot c_{vi})} \qquad (22)$$

p_i = volume per cent of gas i
m = molecular weight of gas i

The density is obtained from the expression:

$$\rho = \frac{\sum (p_i \cdot m_i)}{\sum 22.41} \left(\frac{P}{760}\right) \left(\frac{273 \cdot 16}{T}\right) \qquad (23)$$

where P is the barometric pressure in millimeters of mercury and T is the absolute temperature in degrees Kelvin.

The ratio of this limiting frequency under the condition of helium mixture and a pressure of P ata to that of reference air conditions at P = 1 ata is accordingly

$$\frac{F_w}{F_{wo}} = \frac{c}{c_o} \sqrt{\frac{\rho}{\rho_o}} \qquad (24)$$

where c_o and ρ_o pertain to reference air conditions. The density of the gas mixture ρ at pressure P in ata is according to Eq. (18)

$$\rho = P \cdot \rho_{go} \qquad (25)$$

where ρ_{go} is its density at zero level, i.e. at P = 1 ata. From Eqs. (20) (23) (24) and (25) we obtain

$$\frac{F_w}{F_{wo}} = \sqrt{\frac{\gamma}{\gamma_o} \cdot P} \qquad (26)$$

which at high helium contexts is approximately $1.1 \cdot (P)^{1/2}$, see Fig. II-A-3. It is interesting to note that a subject's F_w is approximately independent of gas mixture and varies as a function of P only. It is therefore possible to find a depth and a pressure P where the closed tract resonance shift Eq. (26) is the same as that of the overall transposition Eq. (21) thus avoiding the non-linear F_1-shifts. The

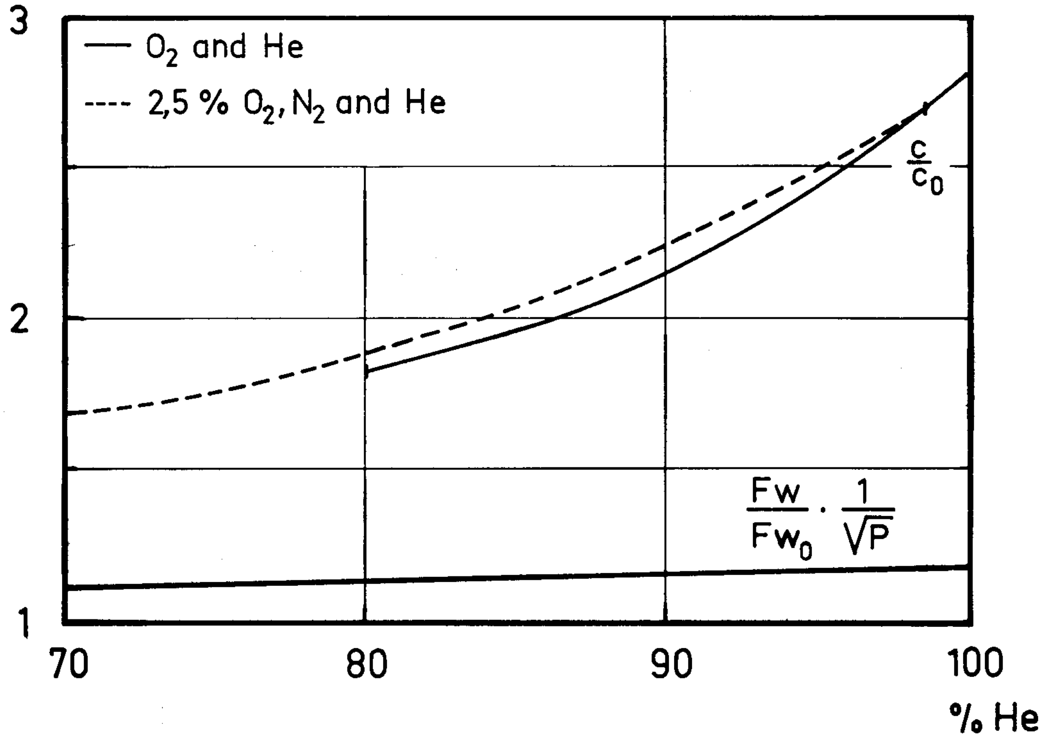

Fig. II-A-3. The transposition of the lower limiting frequency F_w/F_{w_0} and of the sound velocity c/c_0 as a function of the helium content.

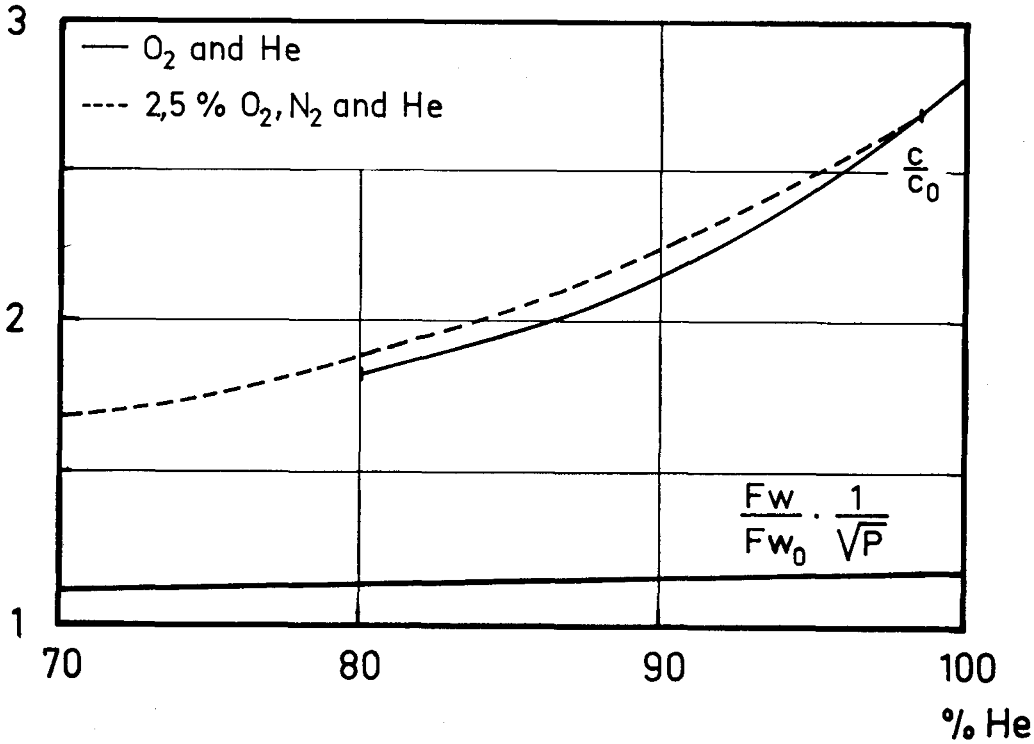

Fig. II-A-3. The transposition of the lower limiting frequency F_w/F_{w_0} and of the sound velocity c/c_0 as a function of the helium content.

condition is apparently

$$P = \rho_o/\rho_{go} \qquad (27)$$

where ρ_o is the density of air at 0-level and ρ_{go} the density of the gas mixture at 0-level. In terms of diving depth d in meters

$$(1 + \frac{d}{10}) = \rho_o/\rho_{go}$$

$$d = 10\,(\frac{\rho_o}{\rho_{go}} - 1)$$

$$= 10(k^2\,\frac{\gamma_o}{\gamma} - 1) \qquad (28)$$

where k is the transposition factor. Thus with $k = 2$ the optimum depth for the mixture is of the order of 35 meters, with $k = 3$, $d = 90$ meters and with $k = 4$, $d = 180$ meters. In order to reach optimum conditions at high depths it is therefore necessary to use a very light gas. Providing a sufficiently wide-band microphone is used the speech can then be restored by a single spectrally uniform retransposition system. It should be noted that the optimum condition, Eq. (27), also implies that at the depth d and pressure P the density $\rho = P \cdot \rho_{go}$ of the gas equals ρ_o, the density of air at 0-level. Under this condition the flow resistance for breathing and phonation is also close to normal values. Hydrogen mixtures would be needed at large depths.

During our experiments in Stockholm we did not have access to helium for filling the pressure tank. Instead we used a light mask breathing unit with gas container. The contents of the expired gas was automatically recorded. After the nitrogen had been ventilated away at 100 meters equivalent depth there remained 97.5 % He and 2.5 % 0_2 and a negligible content of carbon dioxide and water vapor. The calculated shift in sound velocity $k = c/c_o$ was 2.67*. An experimental check by standing-wave measurements in a tube filled with expiration gas from the subject gave k = 2.63**.

* The influence of the carbon dioxide content (about 0.4 %) in the expired gas is negligible. If the humidity at 100 meter is taken into account this value will be lowered with about 1 %.

** These techniques of velocity of sound determinations are reported on separately by F. Fransson in this issue of STL-QPSR.

J. (Jenkins) and 180 Hz for subject C. (Cannon). At a depth of 200 feet (60 meter) the limiting frequency was transposed to F_w = 460 Hz and 500 Hz for the two subjects and at 450 feet (135 meter) to 680 and 740 Hz, respectively for J. and C. Measured formant frequency shifts are in substantial agreement with predicted values (see Figs. III-A-6 to III-A-9). There is, however, a tendency of F_2 and F_3 from the range above 2000 Hz of reference conditions being transposed 5-10 % more than predicted. Articulatory differences, such as an elevated larynx, could account for this difference.

The agreement between theory and experiments, demonstrated in this article, should constitute a final proof that the non-linear F_1-shifts observed at large diving depths differ from the overall transposition rule not because of compensatory articulations, as suggested by a recent report (4) but because of the reduced mismatch between vocal cavity wall impedance and the impedance of the expiration gas at high pressures.

There is a tendency of the subjects to use a higher voice fundamental frequency at higher pressures. The behavior of subject C. is typical in this respect; his mean F_o varying from 124 Hz at zero level to 135 Hz at 200 feet and 146 Hz at 450 feet. It is interesting to note that the slowing down in speaking tempo was about the same for both speakers at both levels, 38 % for subject C. at 200 feet and 38 % for subject J. At 450 feet the stretching of the time scale was 30 % for C. and 38 % for J. Those values do not correlate well with density which was 2.5 ρ_o at 200 feet and 3.5 ρ_o at 450 feet. Incidentally the correlation is excellent with the partial pressure of oxygen which under both conditions was 42 % higher than at 0 level and air. This suggests that the amount of oxygen could affect the speaking tempo. In the Swedish study, however, the speaker variation was large. Intelligibility scores reported by Hollien ran from 90 % at zero level to 47 % at 200 feet and 12 % at 450 feet.

It should be noted that since nitrogen and oxygen have similar gas constants it is possible to add nitrogen to a helium-oxygen mixture in order to retain a relatively low transposition factor at a prescribed oxygen content. The effect on the velocity of sound of substituting oxygen for nitrogen in a He, N_2, 0_2 mixture is small as illustrated by Fig. II-A-3.

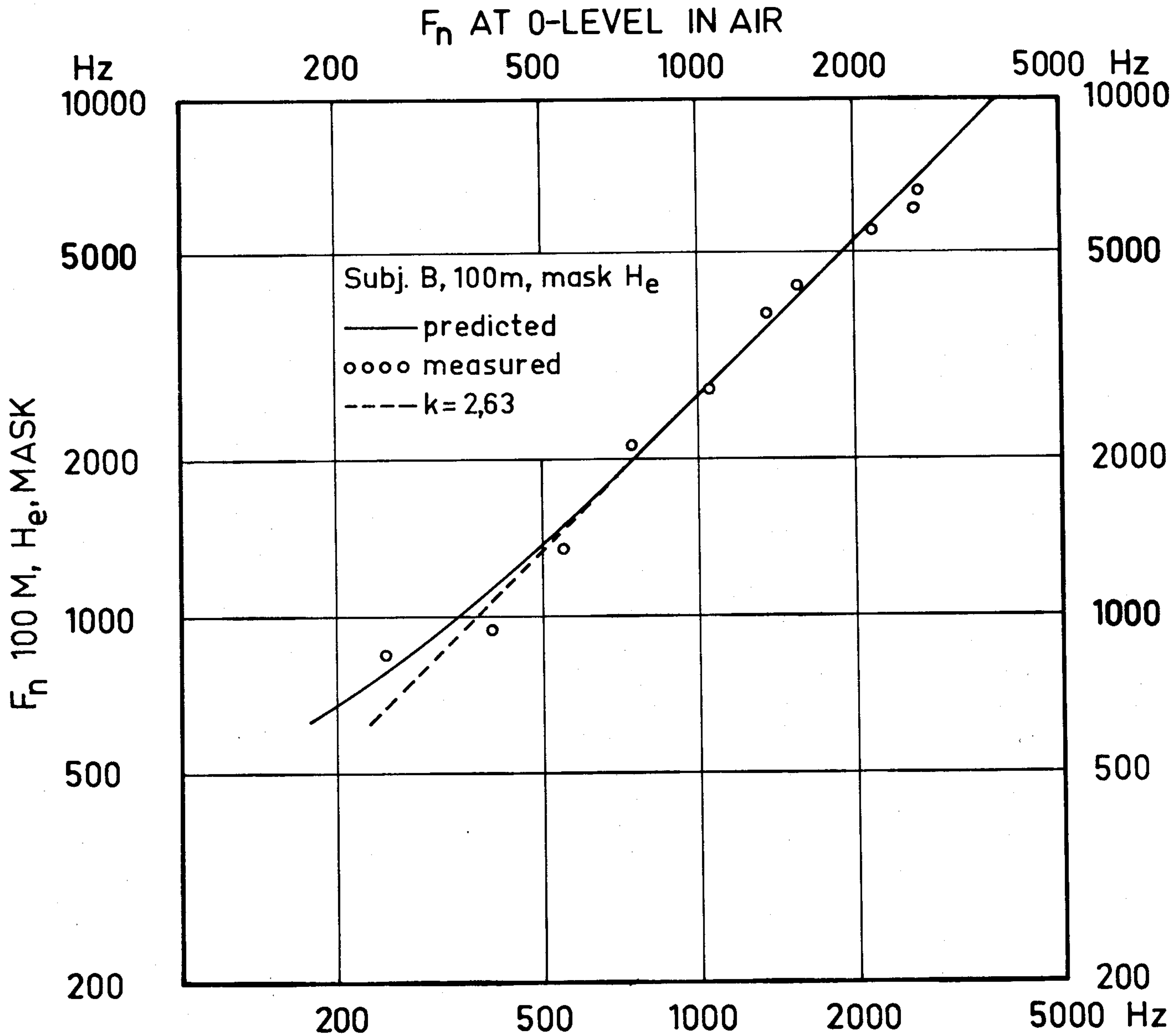

Fig. II-A-4. Predicted and measured formant frequency transpositions of subject B when breathing a 2.5 % O_2, 97.5 % H_e mixture through a light diving mask at 100 meters equivalent depth.

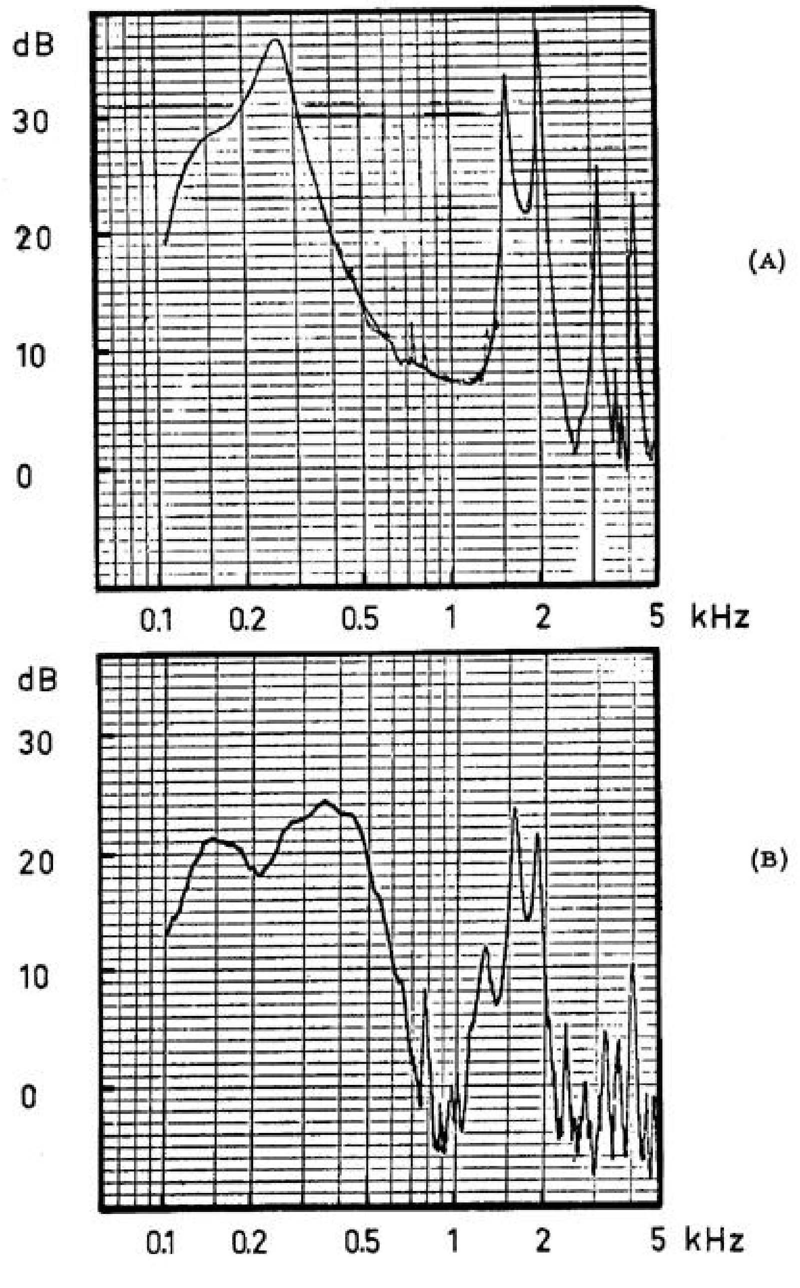

Fig. II-A-5. Sweep frequency measurements of the vocal tract transfer function, (A) without a mask, (B) with a mask. 0-depth in air.

Acknowledgments

This work was supported in part by the Swedish Delegation for Medical Defence Research, project Nos. 24:071/65 and 54:071/66. The work under the grant was carried out in cooperation with Dr. B. Sonesson of the Anatomical Institution, University of Lund and with Mr. F. Fransson of the Department of Speech Communication, KTH. We are most grateful for the help given us by Swedish Navy personnel and in particular by Dr. J. Adolfsson and Dr. A. Muren.

We are much grateful to Dr. H. Hollien of the Communication Sciences Laboratory for making available to us samples of speech data recorded under Sealab III training conditions at the Experimental Diving Unit, Washington, D.C. Valuable comments were received by Miss Patti Paul.

References:

Fant, G.: Acoustic Theory of Speech Production ('s-Gravenhage 1960).

Fant, G. and Sonesson, B.: "Speech at high ambient air-pressure", STL-QPSR No. 2/1964, pp. 9-21.

Fant, G. and Sonesson, B.: "Diver's speech in compressed-air atmosphere", Military Medicine 132 (1967), pp. 434-436.

Fujimura, O. and Lindqvist, J.: "The sinewave response of the vocal tract", STL-QPSR No. 1/1964, pp. 5-10.

Gerstman, L.J.: "Breathing mixture and depth as separate effects on helium speech", J.Acoust.Soc.Am. 40 (1966), p. 1283(A).

HRB-SINGER, Inc.: "The correction of helium speech distortion", Final Report, 4007-F-1, prepared by R.S. Brubaker and J.W. Wurst, May 1967.

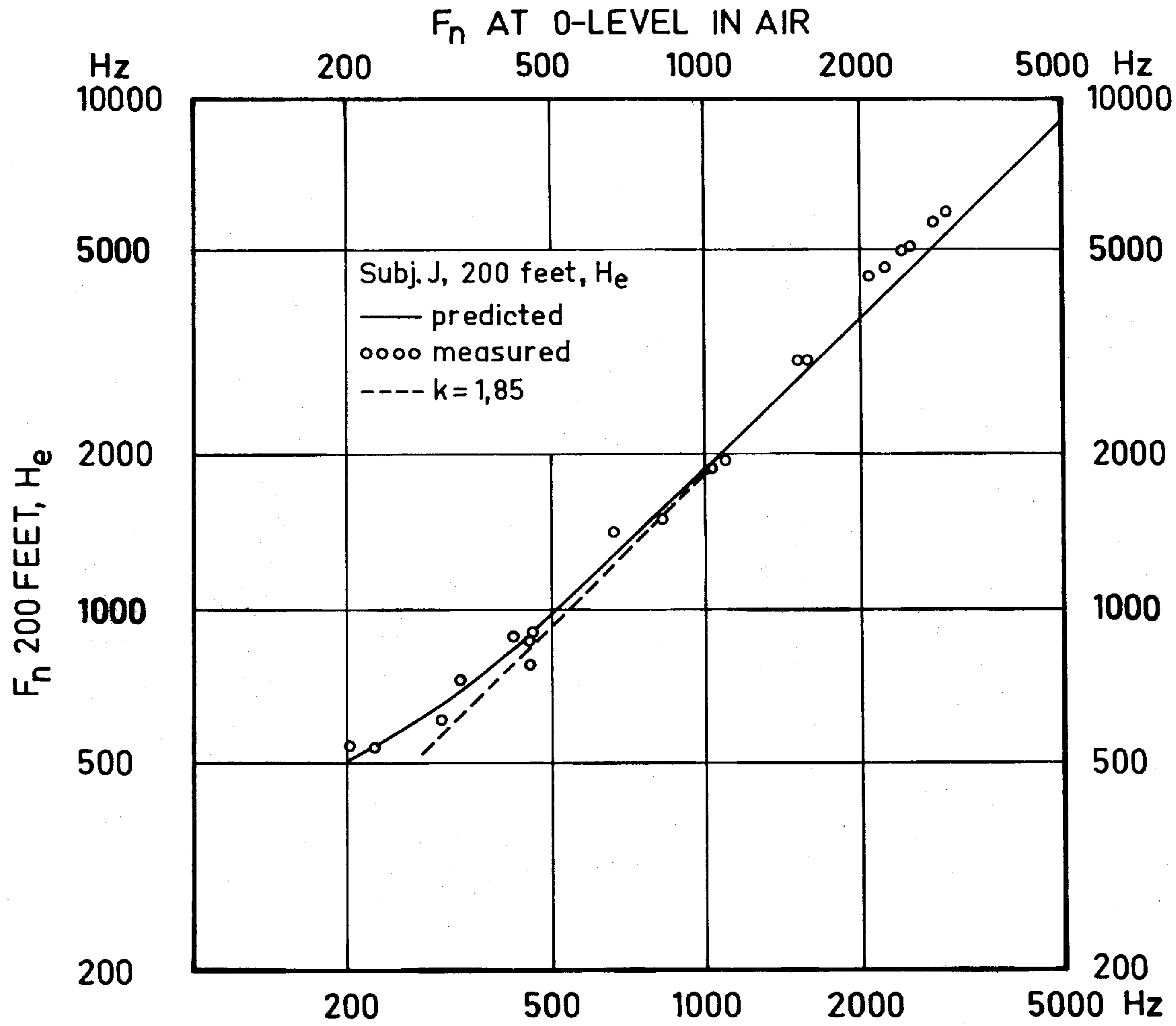

Fig. II-A-6. Predicted and measured formant frequency transposition of subject J breathing a 79.25 % H_e, 4.25 % 0_2, 16.5 % N_2 mixture in "free-field" conditions at a depth of 200 feet. Speech recording (Hollien) from U.S. Navy Experimental Diving Unit, Washington, D.C., USA.

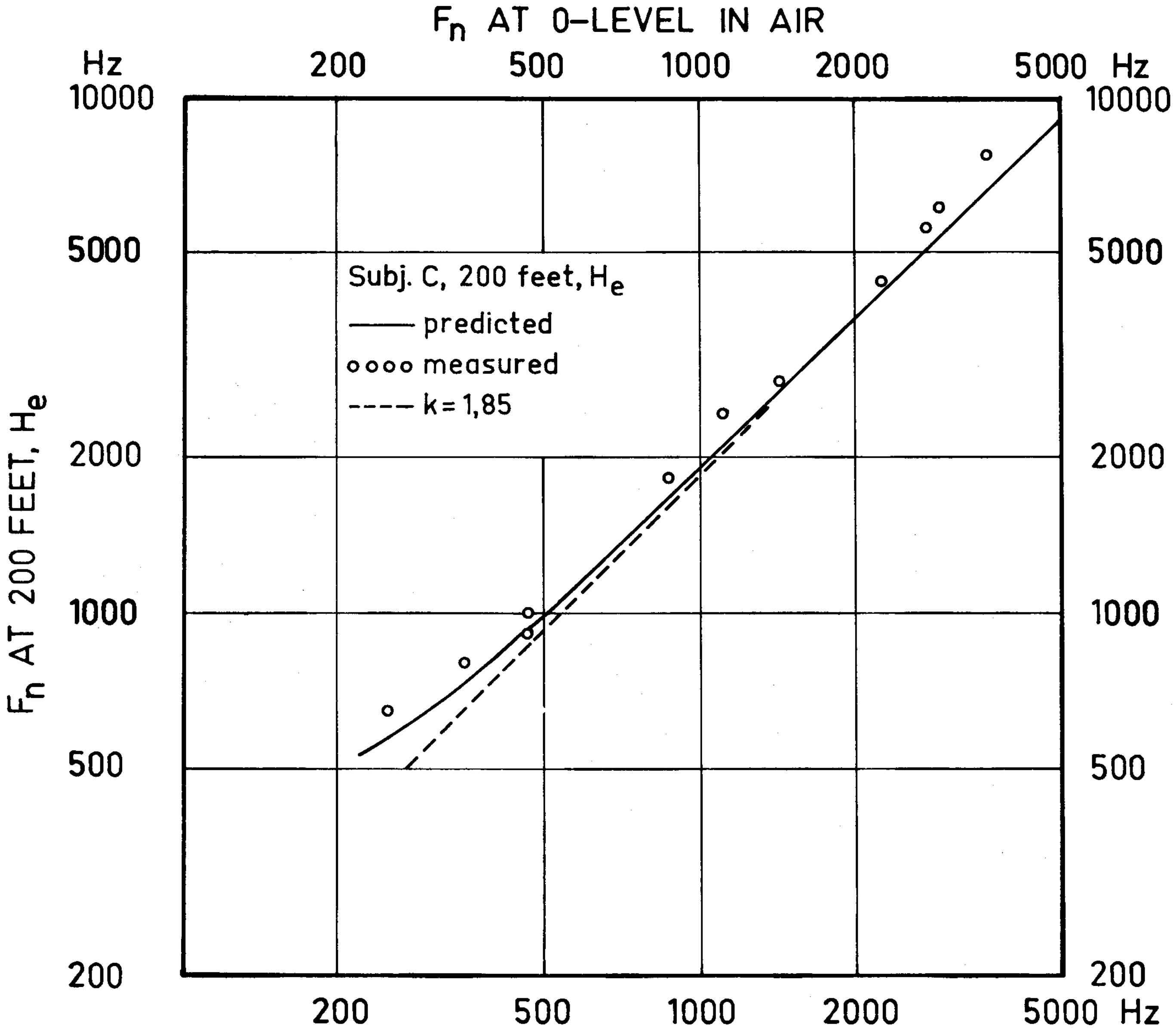

Fig. II-A-7. Same as Fig. II-A-6 for subject C.

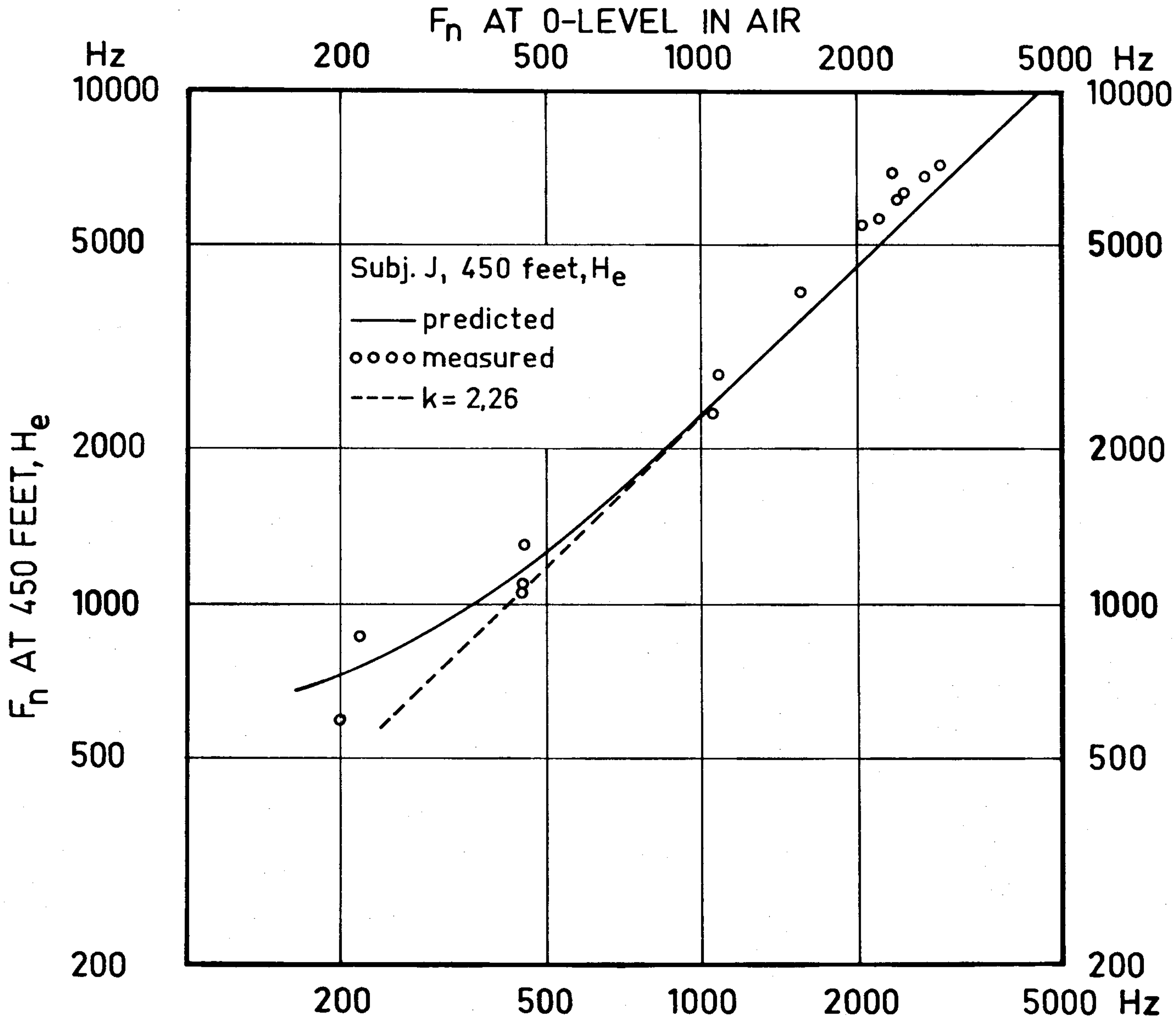

Fig. II-A-8. Predicted and measured formant frequency transposition of subject J breathing a 90.2 % H_e, 2.06 % 0_2, 7.8 % N_2 mixture in "free-field" conditions at a depth of 450 feet. Speech recording (Hollien) from U.S. Navy Experimental Diving Unit, Washington, D.C., USA.

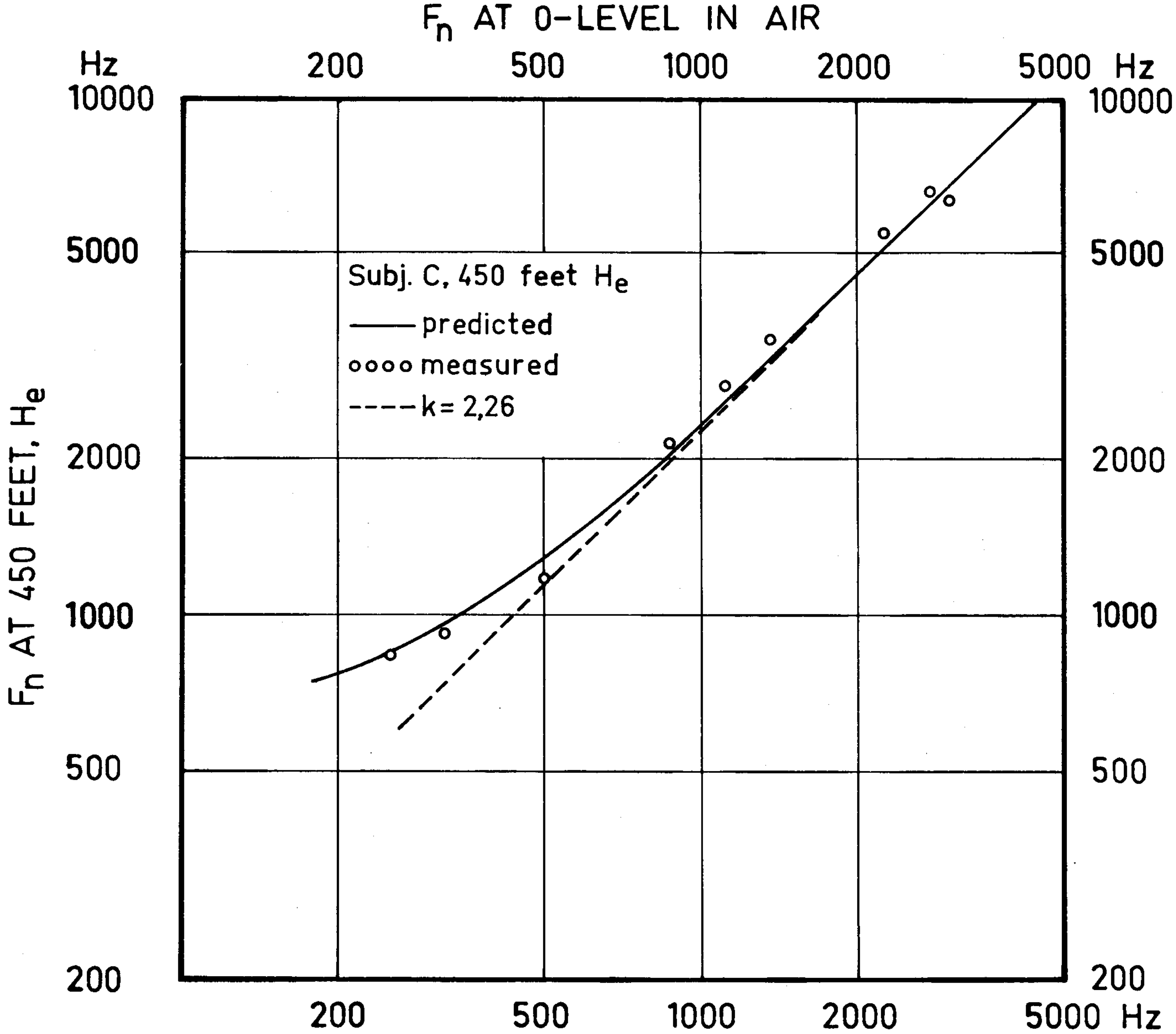

Fig. II-A-9. Same as Fig. II-A-8 for subject C.

B. MEASUREMENTS OF THE SOUND VELOCITY IN GAS MIXTURES

F. Fransson

As a part of a project on the study of human speech under conditions of high pressures and various gas mixtures a method has been worked out for determining the velocity of sound in the expiration gas. During experiments in a pressure tank at the Stockholm Navy Yard the subject diver was provided with gas tubes containing a helium-oxygen mixture and an outfit for the inhalation of the mixture. The air pressure in the tank was raised to 10 kg/cm^2. The subject's expiration gas was collected in a test tube.

The test tube

The test tube was a cylindric steel tube with internal length of 31.4 cm and diameter 3.0 cm, provided with valves for in- and outlet and a pressure meter. A small earphone of type Oticon at each end was used as sound transmitter and receiver.

Measurements

Preliminary measurements of two resonance frequencies f_1 and f_2 were made in the pressure tank. The test tube was then transported to the laboratory where three to five resonance frequencies were measured at a temperature of 32-38°C. The sound velocity obtained from these measurements was corrected to 35° C.

Determination of the effective tube lengths of the test tube

On account of the presence of valves, pressure meter and earphones, the test tube cannot be regarded as a simple homogeneous tube. These difficulties can be circumvented by evaluating a set of effective tube lengths. Therefore a similar tube without fittings was made. It was filled in one experiment with air in another with nitrogen at atmospheric pressure and constant temperature. Six resonance frequencies were measured using the ionophone (G) as sound source. The ionophone electrodes were placed at 0.3 cm from one end of the closed tube and at the other closed end a sond microphone was used as receiver.

The transmission function for this auxiliary tube is: