Sound, hearing and the human voice

In this chapter we discuss the basic physics of sound transmission in air, how sound vibrations are received by the ear, and how we generate sounds with our voices. This is a rich area of study, and by no means fully understood even today.

1. Sound propagation in gases

The key equations we shall need from our previous study of fluids are conservation,

\[ \frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \vec{v}) = 0 \]

and Euler’s equation\(^1\)

\[ \rho \left( \frac{\partial \vec{v}}{\partial t} + \vec{v} \cdot \nabla \vec{v} \right) = -\nabla p \]

where we have neglected viscosity and external forces (such as gravity).

Imagine the gas has equilibrium pressure \(\overline{p}\), density \(\overline{\rho}\), and zero macroscopic velocity\(^2\). Then we may picture a sound wave as a propagating train of small excursions from equilibrium \(\delta p, \delta \rho\) and \(\vec{v}\). Ignoring terms of second or higher order in small quantities, we reduce the conservation law and Euler’s equation to

\[ \frac{\partial \delta \rho}{\partial t} + \overline{\rho} \nabla \cdot \vec{v} = 0 \]

and

\[ \overline{\rho} \frac{\partial \vec{v}}{\partial t} = -\nabla \delta p. \]

Taking the time derivative of the first, and the divergence of the second equation, and eliminating \(\delta \vec{v}\) between them, we obtain

\[ \frac{\partial^2 \delta p}{\partial t^2} = \nabla^2 \delta p. \]

This is almost a wave equation—in fact it is one since almost certainly \(\delta p\) is proportional to \(\delta \rho\). However if we want to find out what the wave velocity is (that is, the speed of sound) so we can compare with experiment, we need to know the constant of proportionality between \(\delta p\) and \(\delta \rho\).

Such a relation follows from knowledge of the thermodynamics of the gas. Sir Isaac Newton assumed that the temperature remains constant during the fluctuations of density and pressure in a sound wave. That is, he assumed Boyle’s Law, \(pV = \text{constant}\) or

\[ \delta (pV) = 0, \]

which yields the relation

\[ \frac{\delta p}{\overline{p}} = \frac{\delta \rho}{\overline{\rho}}. \]

In fact, Newton was wrong—the fluctuations are adiabatic (sometimes called isentropic), meaning no heat flows in or out of the small volumes we have been considering. Under these conditions we have\(^3\)

\[ \delta Q = 0 = \delta U + p \delta V \]

where

1. That is, Newton’s second law of motion.
2. Since the molecules move quite rapidly, but randomly, there is no overall flow.
3. This is conservation of energy, which Newton did not know about—in fact the concept of energy had not yet been invented!
\[ U = \frac{f}{2} RT \]

is the internal energy of the gas (per gram-mol). Here \( R \) is the gas constant, 8.3 J/mol/°K and \( T \) the absolute temperature. The number \( f \) is the number of (active) degrees of freedom of the gas molecules\(^4\). Thus, e.g., monatomic gases have \( f = 3 \), diatomic gases have \( f = 5 \), and tri-(or more) atomic gases have \( f = 6 \).

Now if we combine the equation for energy conservation with the equation of state of a perfect gas,

\[ p \delta V + V \delta p = R \delta T \]

we have

\[ \frac{f}{2} (p \delta V + V \delta p) + p \delta V = 0 \]

and since

\[ \frac{\delta V}{V} = -\frac{\delta p}{\rho} \]

we see that

\[ \left( \frac{2}{f} + 1 \right) \frac{\delta p}{\rho} = \frac{\delta p}{\rho} \]

or in other words,

\[ \delta p = \gamma \frac{\rho}{\mu} \delta p \]

where \( \gamma \) is the ratio of specific heats at constant pressure and volume:

\[ \gamma = \frac{c_p}{c_v} = \frac{2}{f} + 1. \]

When the above relation is substituted into our previous equation we get

\[ \frac{\partial^2 \delta p}{\partial t^2} = \left( \frac{\rho}{\gamma} \frac{\mu}{\rho} \right) \nabla^2 \delta p. \]

This is a wave equation—a solution of the form

\[ \delta p = \delta p_0 f(x - ut) \]

satisfies it, with \( u \) (the speed of sound) given by

\[ u^2 = \left( \frac{\rho}{\gamma} \frac{\mu}{\rho} \right). \]

We now compute the speed of sound in air. From the perfect gas law the preceding formula can be re-expressed as

\[ u = \left( \frac{\gamma RT}{\mu} \right)^{\frac{1}{2}}. \]

where \( \mu \) is the molar mass (grams per gm-mol). Hence the speed of sound in dry air at 20 °C is

\[ u = (1.4 \times \frac{8.3 \times 293}{0.029})^{\frac{1}{2}} = 343 \text{ m/sec}. \]

Since Newton's theory omitted the factor \( \gamma (\frac{5}{2} \text{ for monatomic gases and } \frac{7}{2} \text{ for diatomic ones}) \) the observed speed of sound in air (a mixture of diatomic gases) is about 20% greater than his prediction.

Newton was convinced of the general correctness of his ideas, and proposed several hypotheses to explain the discrepancy; chief among them, that he hadn't considered humidity (which he thought would increase the speed of sound\(^5\)). Because he did this, Broad and Wade accuse Newton of fraud\(^6\).

---

4. The number of active degrees of freedom depends on temperature. At higher temperatures the vibrational modes of the molecules can be excited, whereas at temperatures below 75 0K the rotational modes of hydrogen are no longer excited.

5. ...since sound travels faster in water than in air. Actually water vapor is triatomic so this would decrease \( \gamma \), thereby decreasing the speed of sound relative to dry air.

What Newton did was not scientific fraud: in fact it is entirely appropriate for theorists to suggest alternative hypotheses and new lines of investigation when their numbers are a bit off. In accusing him of fraud, therefore, Broad and Wade display their ignorance of how science is actually pursued, as well as their inability to distinguish good science from bad.

We now clear up several outstanding issues before leaving the theory of sound in gases. First, since
\[ \frac{\partial \delta \rho}{\partial t} + \bar{\rho} \nabla \cdot \vec{v} = 0 \]
and
\[ \bar{\rho} \frac{d\vec{v}}{dt} = -\nabla \delta \rho, \]
we see that if the sound wave is propagating in the x direction,
\[ \delta \rho = \delta \rho_0 f(x - ut), \]
then the velocity field will be
\[ \vec{v} = \hat{x} u \delta \rho_0 \frac{f(x - ut)}{\rho}. \]
This means that the displacement field (of the air from its equilibrium position) points along the direction of propagation. We call such a wave longitudinal.

The next issue is, “How good was the adiabatic approximation” (in which we neglected heat flow from one parcel of gas to its neighbor)? We can answer this by considering the equation of heat conduction,
\[ c_p \frac{\partial T}{\partial t} + \kappa \nabla^2 T = 0, \]
where \( c_p \) is the specific heat at constant pressure and \( \kappa \) is the thermal conductivity.

This is like the diffusion equation which we have already studied. The time scale for heat to diffuse a distance \( \lambda \) (that is, a wavelength of sound, which is the characteristic distance scale) is
\[ \tau_H \sim \frac{c_p \lambda^2}{\kappa} \]
and we must compare this with the period of the sound wave,
\[ \tau_S = \frac{1}{v} = \frac{\lambda}{u}. \]
That is, we want the ratio
\[ \frac{\tau_H}{\tau_S} \sim \frac{c_p \lambda}{\kappa u}. \]
For air \( \kappa \approx 25 \times 10^{-3} \text{ W/m}^2/\text{°K} \) and the (volumetric) specific heat is
\[ c_p = \left( \frac{f}{2} + 1 \right) \frac{R}{\Omega} \]
where \( \Omega \) is the molar volume, 0.0224 m\(^3\) at STP.

Thus for sound of frequency 20 KHz (with \( \lambda = 2 \text{ cm} \)) we find
\[ \frac{\tau_H}{\tau_S} \sim 2.6 \times 10^5. \]
In other words, the time for heat to be transmitted from one small parcel of air to another is much greater than the time for the pressure and density in the parcels to oscillate. Hence we may neglect heat transmission. The third and final issue we need to discuss is the power transmitted by a sound wave. That is, how much energy flows across a unit surface per unit time. We would like to express this in terms of pressure fluctuations \( \delta p \) since they are more easily measured than density fluctuations.

The energy density (energy per unit volume) in a sound wave is

7. Contrast this with a surface wave on liquid, or light, where the displacements are always transverse to the direction of propagation.
\[ U = \frac{1}{2} \delta \dot{\nu}^2 + \frac{\gamma}{2} \delta \left( \nabla \cdot \zeta \right)^2 \]

where the displacement field \( \zeta \) is related to the velocity field by

\[ \frac{\partial \zeta}{\partial t} = \nu. \]

By looking at the equations for small displacements we find, ultimately, that

\[ U = \frac{1}{\gamma \rho} \left( \delta \rho \right)^2 = \frac{1}{u_s \rho} \left( \frac{RT}{\gamma \mu} \right)^{1/2} \left( \delta \rho \right)^2 \]

so that the sound intensity is then

\[ I = U u_s = \frac{1}{\rho} \left( \frac{RT}{\gamma \mu} \right)^{1/2} \left( \delta \rho \right)^2. \]

2. \textbf{Hearing and the ear}

The ear is a remarkable device for sensing sound. It consists of the outer ear, which guides sound onto the eardrum, a thin membrane. This in turn presses on three small bones (the ossicles)—in fact the smallest bones in the human body—located in the middle ear. The vibrations transmitted to the three ossicles are then conveyed in turn to the oval window of the cochlea. All these are shown schematically in the figure above. The cochlea is the organ that actually resolves sound into low- and high-frequency components. The middle ear is the region between eardrum and cochlea. Note the eustachian tube that connects the middle ear to the nasal passages. It serves both for drainage and pressure-equalization. The reason for equalizing pressure between outer and middle ear regions is to prevent excessive strain from bursting the rather delicate eardrum. The drainage function reduces the incidence of middle ear infections by sweeping.

---

8. Infections of the middle ear are far less frequent than in earlier generations. This is partly due to increased use of antibiotic and sulfa drugs, and partly to the almost complete elimination of the tonsillectomy as a routine surgical procedure. Virtually all members of the author's generation had their tonsils removed, whereas virtually none of his students have been so treated. That this surgery was known to be unnecessary, possibly deleterious, and far too frequent even as early as 1949 is attested by the remark: "The function of the tonsils is lymphocyte production. They are organs of major surgical importance: they can be removed easily, with little harm to the patient and much benefit to the surgeon." — J. Nonidez and W.F.
bacteria out into the nasal passages where they can be expelled.

The ossicles—named malleus (hammer), incus (anvil) and stapes (stirrup)—are not mere sound conductors. They provide a frequency dependent acoustic impedance that conditions complex sounds (that is, ones comprising superpositions of many frequencies). They also serve as a pressure multiplying device, since as levers they have a mechanical advantage of about 1.5; whereas the areas in contact with eardrum and oval window are approximately in the ratio of 20:1. Thus the pressure on the eardrum is multiplied 30-fold in its trip to the cochlea.

The cochlea is (primarily) a coiled tapering cylinder filled with a lymph-like fluid that transmits sound from the oval window along the basilar membrane that divides the cylinder into two halves. At the far end of the basilar membrane is a pressure-equalizing hole (the helicotrema); and moreover, the membrane ends just short of the far end of the tube so that sound waves can pass from one section to the other.

At the far end of the acoustical path, on the opposite side of the dividing membrane from the oval window is a round window. Each of these windows is closed with a thin membrane; the oval window admits vibrations from the stapes into the interior fluid of the cochlea. The round window presumably serves to prevent sound from reflecting from the end of the acoustical path. The overall length of the acoustical path is about 3 cm.

The basilar membrane is stiffer near the oval window than at its far end. Thus short wavelength sounds create the largest motion near the oval window; whereas sounds of lower pitch cause greatest motion at the far end. In other words, the basilar membrane’s structure spatially separates sounds of different pitch, thereby enabling the subsequent information processing to distinguish them on the basis of which nerves were excited.

In contact with the basilar membrane along its length is the organ of Corti, the transducer that converts sound to nerve impulses. It is covered with hair cells. When the cilia of these cells are set in motion by the motion of the basilar membrane, the cells fire off pulses that travel along the auditory nerve to the auditory centers of the brain.

Just how the sound information is encoded into pulses is not fully understood. Nevertheless there are some general principles the encoding mechanism must obey. Nerve impulses are (temporally) sharp spikes in the action potential that have nearly constant magnitude. That is, since each spike on a given axon is very much like its predecessor and its successor, there is no possibility of conveying information by varying the size of the spike (amplitude modulation). Thus the information must be transmitted by varying the frequency of pulses (frequency modulation).

One simple neural encoding scheme maps intensity of stimulation to increased frequency of pulses. If this is the way Nature actually encodes, it would explain both why prolonged listening to loud sounds leads to fatigue and diminished response (the nerves run out of fuel faster at higher intensity levels); as well as to the logarithmic perception of sound (see §3 below).

3. **Response to sound**

The approximate threshold of hearing (that is, the softest sound we can hear) involves a pressure variation of

\[ \delta p \approx 2 \times 10^{-5} \text{ N/m}^2 \]

which is about \(10^{-10}\) atmospheres.

The response of the ear to sound intensity is approximately logarithmic, that is, we perceive loudness approximately as

\[ N = \log_{10} \left( \frac{I}{I_0} \right) \]

where the reference intensity is

\[ I_0 = 10^{-12} \text{ W/m}^2, \]

the approximate threshold of hearing. This scale divides perception of loudness into bels, each representing a decade increase in actual intensity. That is, a sound that is 1 bel louder is 10 times more intense. The bel is inconveniently large so we measure in units of 0.1 bel, or decibels. The approximately logarithmic response of the ear to sound is called the Weber-Fechtner law.

A sound that is 120 decibels relative to the threshold (that is, intensity 1 W/m²) is about as loud as we can tolerate without sustaining damage to our ears. Sounds of 140 dB are guaranteed to damage our ears if listened to for sustained periods of time. Many rock concerts are actually louder than this.

An important function of the the muscles that hold the eardrum and ossicles in place is to protect against excessively loud sounds and sudden pressure changes. In the presence of sounds above 85 dB these muscles tighten and pull the stapes away from the oval window of the cochlea. This acoustic reflex provides about 20 dB worth of protective isolation of the inner ear from very loud sounds. However the response time of the muscles is about 150 msec, so loud sounds with rapid rise-times (gunfire, for example) can damage the ears permanently. (One also suspects that during prolonged exposure to rock concerts the muscles tire and become ineffective.)

Logarithmic response implies that the mapping of intensity to frequency must be something like

\[ \frac{dl}{I} = \frac{df}{f_{\text{max}}}. \]

That is, there is obviously a maximum possible frequency at which nerve impulses can be fired off. On dimensional grounds the ratio of change in frequency (response) to the maximum frequency must equal the change in intensity divided by some other intensity. But in the absence of a reference level of sound to compare with, the only quantity with the dimensions of sound intensity (available to divide by) is the current level of intensity itself. In other words, it would be hard to envision any other response curve than a logarithmic one.

In the figure below are the Fletcher-Munson curves of equal loudness. They exhibit clearly the logarithmic character of perception, over the range of audible frequencies.
4. **Frequency response of the ear**

Below is a schematic representation of the outer and middle ears. If the ear tube is treated as a Helmholtz resonator its resonant frequency would be

\[ f = \left( \frac{u_s}{2\pi} \right) \sqrt{A/V} L \]

where \( A \) is the area of the tube, \( L \) is its length and \( V \) the volume of the chamber. Plugging in the numbers shown in the Table below

<table>
<thead>
<tr>
<th>Parameters of Helmholtz resonators</th>
<th>Orangina bottle</th>
<th>Human ear</th>
</tr>
</thead>
<tbody>
<tr>
<td>L</td>
<td>4 cm</td>
<td>2.5 cm</td>
</tr>
<tr>
<td>V</td>
<td>313 cm³</td>
<td>2 cm³</td>
</tr>
<tr>
<td>A</td>
<td>2.8 cm²</td>
<td>0.096 cm²</td>
</tr>
</tbody>
</table>

and taking 343 m/sec for the speed of sound we get \( f \approx 760 \) Hz. To see how good this picture is I have measured the frequency response of an Orangina™ bottle whose dimensions are also given in the Table. The measured frequency was 215 Hz, whereas that calculated from the Helmholtz formula was 260 Hz. If the ear’s Helmholtz response were depressed similarly, we should estimate it at 630 Hz.

Clearly the main Fletcher-Munson dip at 3500 Hz cannot be this Helmholtz resonance\(^9\). On the other hand, the resonant frequency of an organ pipe closed at one end and open at the other is given by the condition that the pipe should hold \( \frac{1}{4} \) of a standing wave (see below). If we take the resonant tube to be the outer ear, of length 2.5 cm, we see the wavelength (corresponding to the resonant frequency) is 10 cm, giving a fundamental frequency of 3430 Hz. We note also that the first harmonic of the quarter-wave organ pipe lies at three times the fundamental frequency, namely 10,300 Hz. In fact, the Fletcher-Munson curves exhibit a less-pronounced dip at about 11 Khz.

\[ \text{1/4-wave organ pipe showing displacement vs. position of standing wave} \]

The best frequency response of the ear is the deepest minimum of the Fletcher-Munson curves, in the frequency range 3000-4000 Hz, but as noted above, there are also less pronounced dips around the Helmholtz frequency as well as at the first harmonic of the organ pipe. We conclude the ear is more like an organ pipe than like a Helmholtz resonator, but that both aspects of its geometry act to determine its frequency response.

5. **Spatial discrimination of sounds**

A sound in the optimal frequency range for hearing has a wavelength of 10 cm. If a source of a pure tone were placed some distance away, clearly the wavelength would determine our ability to locate the source through phase comparisons of the two

\[ \text{1/4 wave organ pipe showing displacement vs. position of standing wave} \]

However, note the less-pronounced dip at 500-600 Hz.
ears. Or to put it another way, the angular resolution—our ability to discriminate between the positions of two similar sources—would be given by the Rayleigh criterion of optics (see Ch. 11, p. 102):

\[ \theta_{\text{min}} \approx 1.22 \frac{\lambda}{D} \]

where \( D \) is the diameter of the “lens” (that is, the width of the human head, about 20 cm). Thus if our ability to localize the origins of sounds were diffraction-limited, the minimum angle we could resolve (at 4 KHz) would be about 35°.

A simple classroom experiment with a blindfolded subject revealed that we can localize sounds with much greater precision—within \( \pm 2^o \)—if they are impulsive sounds, such as the striking of a tuning fork. On the other hand, we do not localize steady sounds of definite pitch very well.10

There are two other types of information that allow us to discriminate angle: loudness and time of arrival (for impulsive sounds). Imagine sound emanating from a point source. Then assuming it is at an angle \( \theta \) with respect to a line perpendicular to the forehead, we see that the distances from the source to the ears are given by (see figure to the right)

\[ r^2 = R^2 \pm RD \sin \theta + \frac{D^2}{4} = R^2 \pm RD \theta; \]

since intensities vary inversely with the square of the distance, we see that the intensity difference between the two ears is

\[ \frac{\Delta I}{I} \approx \frac{2D}{R} \theta. \]

We can discriminate differences in loudness of—say—0.1 dB. This means that we can discriminate relative differences of

\[ \frac{\Delta I}{I} \approx 0.1 \times \frac{\ln(10)}{10} = 0.0230... \]

which translates to an angular difference, for sound originating a distance 10 m away, of about 33°. This is comparable to the diffraction limitation. However, for sounds of frequency 4000 Hz or more (that is, with wavelengths shorter than \( D/2 \)) the head produces a “sound shadow”. That is, the ear on the far side of the head receives about 8 dB less sound than that on the near side. This improves angular discrimination eightfold, reducing the angular error to \( \Delta \theta \approx \pm 8^o \).

However we can also discriminate between times of arrival of sound in our two ears. If the ears are a distance \( D \) apart, then the angular precision we can achieve is clearly

\[ \theta = \frac{u_s \Delta t}{D}, \]

where \( \Delta t \) is the time needed to analyze the sound. Clearly this must be the time for sound to travel down the 3 cm length of the cochlea. Assuming the speed of sound in cochlear fluid is the same as that in water, 1500 m/sec, we find \( \Delta t = 2 \times 10^{-5} \) sec. Taking this for \( \Delta t \) we find

10. Nonetheless we do it better than \( \pm 35^o \). Probably we are comparing the relative loudness of the signals in our two ears.
This is, in fact, about what we can do in resolving impulsive sounds.

6. **The human voice**

Evolution has endowed the human species with the ability to use language, both for thought and communication, at a level that greatly exceeds the abilities of any other known species. The human vocal apparatus can produce a rich variety of sounds whose production is controlled by a significant fraction of the brain.

Here we concern ourselves with the physical acoustics of voice production. As we see above, human anatomy combines voice production with the organs of breathing and ingestion. Since we must open the trachea (to provide airflow that vibrates the vocal cords) in order to speak, our mothers were right to warn us not to talk with food in our mouths. Besides being socially incorrect, such a solecism carries its own penalty, in the form of a serious danger of choking. In fact a significant number of people die each year from this cause—the Heimlich maneuver notwithstanding.

Supposing that you are obeying parental injunctions and are vocalizing without oral obstructions, the process goes something like this: air from the lungs, expelled by contraction of the diaphragm and chest muscles, passes over the vocal cords, setting them into vibration. The vocal cords are not strings but rather folds of ligament within the cartiliginous larynx.

The waveform produced by this process is a sawtooth—that is, our fundamental sound generator produces buzzes and hisses—as shown on p. 96 below. Also shown in the figure are partial sums of the Fourier decomposition of the sawtooth wave. To understand what this is about requires a small digression on Fourier’s series.

---

**Digression**

A function that is periodic in time,

\[ f(t + \tau) = f(t) \]

can, under quite general conditions, be expanded as an infinite series of trigonometric functions

---


12. The danger increases with age because the muscles of the throat and mouth lose their tone; the effects of alcohol also increase the risk.


14. The larynx can also produce glottal clicks, which are featured in African and Austrasian languages, although not in English.
\[
f(t) = \sum_{n=0}^{\infty} \left[ a_n \cos\left(\frac{2\pi nt}{\tau}\right) + b_n \sin\left(\frac{2\pi nt}{\tau}\right) \right]
\]

The coefficients \(a_n\) and \(b_n\) can be expressed as integrals (\(b_0\), of course, is identically zero):

\[
\begin{align*}
(a_n) &= \frac{2}{\tau} \int_0^{\tau} f(t) \cos\left(\frac{2\pi nt}{\tau}\right) dt \\
(b_n) &= \frac{2}{\tau} \int_0^{\tau} f(t) \sin\left(\frac{2\pi nt}{\tau}\right) dt
\end{align*}
\]

The curve labelled “0’th and 1st harmonics” is the sum of the above series, with terms \(n=0\) and \(n=1\). The two other curves, “+ 2nd harmonic” and “+ 3rd harmonic” mean the terms with \(n=2\) and \(n=3\) have been added, respectively. Clearly, the sum is converging rapidly to the original sawtooth function.

The Fourier decomposition makes clear that the sawtooth function contains components of many frequencies besides the fundamental tone, whose frequency is

\[f = \frac{1}{\tau}.
\]

Also, the vocal cords do not produce a single fundamental frequency, but rather a superposition of sawtooth signals superimposed upon each other. That is, even ignoring the presence of harmonics\(^{15}\), there is some spread to the frequency spectrum.

Seen from this point of view, the mechanism of voice production generates a range of fundamental and harmonic tones, which are selectively amplified by the various resonating chambers and organ pipes of the airway. Thus, for example, if we view the trachea as a quarter-wave pipe around 20 cm long, its resonant frequency will be about 440 Hz.

---

15. A harmonic (or as the Brits would say, an harmonic) is simply a sound whose frequency is an integer multiple of the fundamental tone’s.
On the other hand, if we think of the chest cavity as a Helmholtz resonator about 40 cm long, with volume 2,000 cm$^3$ connected to a tube (trachea) about 5 cm in diameter, the formula

$$f = \frac{u_s}{2\pi} \sqrt{\frac{A}{V_L}}$$

predicts a resonance at about 850 Hz.

Finally, the oral and nasal cavities further selectively amplify sound at frequencies that can be varied by varying such factors as their shape, the pressure of the air in them, the muscle tone in their walls, etc. In some instances (the formation of sibilants, for example) the tongue, hard palate and teeth act as components of a whistle.

In other words, it should not surprise us to find that the frequency range of human speech is 500–8,000 Hz—this is, after all, the range of typical frequencies and overtones of our vocal apparatus.

End of digression

A last remark: the frequency is proportional to the speed of sound. Thus if the breathing apparatus is filled with Helium gas instead of air, the speed of sound increases by the factor 2.93.

Exercise

Show that the speed of sound in He is 2.93 times that in air. (The molar mass of air is 29 gm, while that of He is 4 gm.) Do not forget that whereas air is diatomic, He gas is monatomic.

Thus the frequency range of someone breathing the He-O$_2$ mixture of deep divers will be 1500-24,000 Hz—in a word, like Donald Duck.