



Introduction to Acoustic Terminology and Signal Processing

4

Christine Erbe, Alec Duncan, Lauren Hawkins,
John M. Terhune, and Jeanette A. Thomas

4.1 What Is Sound?

Most people think of sound as something they can hear, such as speech, music, bird song, or noise from an overflying airplane. There has to be a source of sound, such as another person, an animal, or a train. The sound then travels from the source through the air to our ears. Acoustics is the science of sound and includes the generation, propagation, reception, and effects of sound. The more scientific definition of sound refers to an oscillation in pressure and particle displacement that propagates through an acoustic medium (American National Standards Institute 2013; International Organization for Standardization 2017). Sound can also be defined as an auditory sensation that is evoked by such oscillation (American National Standards Institute 2013), however, more general definitions do not require a human listener, do allow for an animal receiver, or don't require a receiver at all.

Not all sounds produce an auditory sensation in humans. For example, ultrasound refers to sound at frequencies above 20 kHz, while infrasound refers to frequencies below 20 Hz. These definitions are based on the human hearing range of 20 Hz – 20 kHz (American National Standards Institute 2013). While sound outside of the human hearing range is inaudible to humans, it may be audible to certain animals. For example, dolphins hear well into high ultrasonic frequencies above 100 kHz. Also, inaudible doesn't mean that the sound cannot cause an effect. For example, infrasound from wind turbines has been linked to nausea and other symptoms in humans (Tonin 2018). As well, the effects of ultrasound on humans have been of concern (Parrack 1966; Acton 1974; Leighton 2018).

Noise is also sound, but typically considered unwanted. It therefore requires a listener and includes an aspect of perception. Whether a sound is perceived as noise depends on the listener, the situation, as well as acquired cognitive and emotional experiences with that sound. Different listeners might perceive sound differently and classify different sound as noise. One person's music is another person's noise. Noise could be the sound near an airport that has the potential to mask speech. It could be the ambient noise at a recording site and encompass sound from a multitude of sources near and far. It could be the recorder's electric self-noise (see also American National Standards

Jeanette A. Thomas (deceased) contributed to this chapter while at the Department of Biological Sciences, Western Illinois University-Quad Cities, Moline, IL, USA

C. Erbe (✉) · A. Duncan · L. Hawkins
Centre for Marine Science and Technology, Curtin
University, Perth, WA, Australia
e-mail: c.erbe@curtin.edu.au; A.J.Duncan@curtin.edu.au

J. M. Terhune
Department of Biological Sciences, University of New
Brunswick, Saint John, NB, Canada
e-mail: terhune@unb.ca

Institute 2013; International Organization for Standardization 2017). In contrast to noise, a signal is wanted, because it conveys information.

There are many ways to describe, quantify, and classify sounds. One way is to label sounds according to the medium in which they have traveled: air-borne, water-borne, or structure-borne (also called substrate-borne or ground-borne). For example, scientists studying bat echolocation work with air-borne sound. Those looking at the effects of marine seismic survey noise on baleen whales work with water-borne sounds. Some of the sound may have traveled as a structural vibration through the ground and is therefore referred to as structure-borne. Just as earthquakes can be felt on land, submarine earthquakes can be sensed by benthic organisms on the seafloor. In both cases, the sound is structure-borne (Dziak et al. 2004). Sound can cross from one medium into another. The sound of airplanes is generated and heard in air but also transmits into water where it may be detected by aquatic fauna (e.g., Erbe et al. 2017b; Kuehne et al. 2020).

Another way of grouping sounds is by their sources: geophysical, biological, or anthropogenic. Geophysical sources of sound are wind, rain, hail, breaking waves, polar ice, earthquakes, and volcanoes. Biological sounds are made by animals on land, such as insects, birds, and bats, or by animals in water, such as invertebrates, fishes, and whales. Anthropogenic sounds are made by humans and stem from airplanes, cars,

trains, ships, and construction sites. The distinction by source type is common in the study of soundscapes. These comprise a geophony, biophony, and anthropophony.

The following sections explain some of the physical measurements by which sounds can be characterized and quantified. The terminology is based on international standards (including, International Organization for Standardization 2007, 2017; American National Standards Institute 2013).

4.2 Terms and Definitions

4.2.1 Units

A wide (and confusing) collection of units can be found in early books and papers on acoustics, but the units now used for all scientific work are based on the International System of Units, better known as the SI system (Taylor and Thompson 2008). In this system, a unit is specified by a standard symbol representing the unit itself, and a multiplier prefix representing a power of 10 multiples of that unit. For example, the symbol μPa (pronounced micro pascal) is made up of the multiplier prefix μ (micro), representing a factor of 10^{-6} (one one-millionth) and the symbol Pa (pascal), which is the SI unit of pressure. So, a measured pressure given as $1.4 \mu\text{Pa}$ corresponds to $1.4 \text{ times } 10^{-6} \text{ Pa}$ or 0.0000014 Pa . The SI base units are listed in Table 4.1. Other quantities and their units result from quantity equations that are

Table 4.1 SI base units (length, mass, time, electric current, temperature, luminous intensity, and amount of substance) and example derived units (frequency, pressure, energy, and power)

Quantity	Unit name	Unit symbol	Expressed in terms of base units
Length	meter	m	
Mass	kilogram	kg	
Time	second	s	
Electric current	ampere	A	
Temperature	kelvin	K	
Luminous intensity	candela	cd	
Amount of substance	mole	mol	
Frequency	hertz	Hz	1 / s
Pressure	pascal	Pa	kg / (m s ²)
Energy	joule	J	kg m ² / s ²
Power	watt	W	kg m ² / s ³

Table 4.2 SI multiplier prefixes

Prefix	Symbol	Factor	Prefix	Symbol	Factor
deci	d	10^{-1}	deka	da	10^1
centi	c	10^{-2}	hecto	h	10^2
milli	m	10^{-3}	kilo	k	10^3
micro	μ	10^{-6}	mega	M	10^6
nano	n	10^{-9}	giga	G	10^9
pico	p	10^{-12}	tera	T	10^{12}

based on these base quantities. The SI multiplier prefixes that go along with these units are listed in Table 4.2. Note that unit names are always written in lowercase. However, if the unit is named after a person, then the symbol is capitalized, otherwise the symbol is also lowercase. Examples for units named in honor of a person are kelvin [K], pascal [Pa], and hertz [Hz].

4.2.2 Sound

Sound refers to a mechanical wave that creates a local disturbance in pressure, stress, particle displacement, and other quantities, and that propagates through a compressible medium by oscillation of its particles. These particles are acted upon by internal elastic forces. Air and water are both fluid acoustic media and sound in these media travels as longitudinal waves (also called pressure or P-waves). A common misconception is that the air or water particles travel with the sound wave from the source to a receiver. This is not the case. Instead, individual particles oscillate back and forth about their equilibrium position. These oscillations are coupled across individual particles, which creates alternating regions of compressions and rarefactions and which allows the sound wave to propagate (Fig. 4.1¹). The line along which the particles

oscillate is parallel (or longitudinal) to the direction of propagation of the sound wave in the case of longitudinal waves.

Rock is a solid medium and here, vibration travels as both longitudinal (also called pressure or P-waves) and transverse waves (also called shear or S-waves). In S-waves, the particles oscillate perpendicular to the direction of propagation. It is again because of the coupling of particles, that the wave propagates. P-waves travel faster than S-waves so that P-waves arrive before S-waves. The P therefore also stands for “primary” and S for “secondary.”

4.2.3 Frequency

Frequency refers to the rate of oscillation. Specifically, it is the rate of change of the phase of a sine wave over time, divided by 2π . Here, phase refers to the argument of a sine (or cosine) function. It denotes a particular point in the cycle of a waveform. Phase changes with time. Phase is measured as an angle in radians or degrees. Phase is a very important factor in the interaction of one wave with another. Phase is not normally an audible characteristic of a sound wave, though it can be in the case of very-low-frequency sounds.

A simpler concept of frequency of a sine wave, as shown in Fig. 4.1, is the number of cycles per second. A full cycle lasts from one positive peak to the next positive peak. To determine the frequency, count how many full cycles and fractions thereof occur in 1 s. Note that pitch is an attribute of auditory sensation and while it is related to frequency, it is used in human auditory perception as a means to order sounds on a musical scale. As

¹ Dan Russell’s animations of particle motion during acoustic wave propagation: <https://www.acs.psu.edu/drussell/Demos/waves-intro/waves-intro.html>, of the amplitude at a fixed location: <https://www.acs.psu.edu/drussell/Demos/wave-x-t/wave-x-t.html>, and of longitudinal and transverse waves: <https://www.acs.psu.edu/drussell/Demos/waves/wavemotion.html>; accessed 12 October 2020.

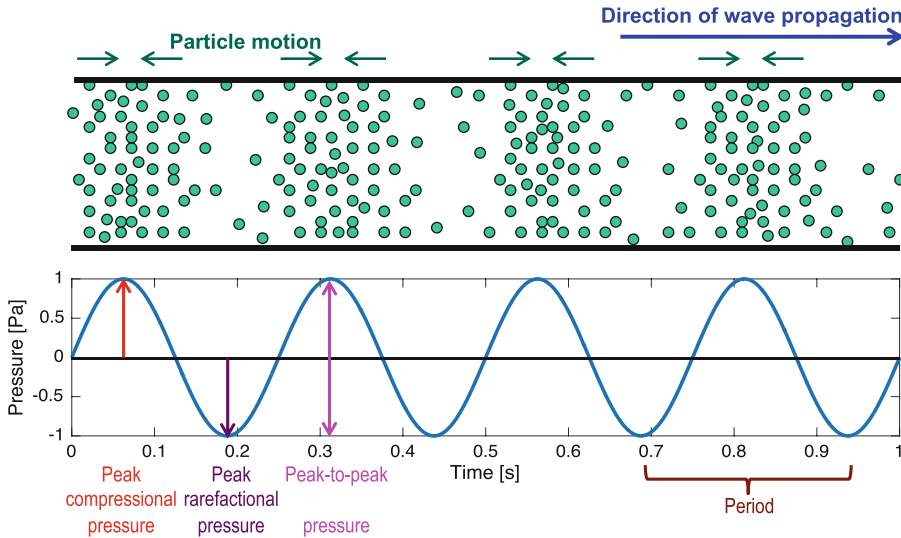


Fig. 4.1 A sinusoidal sound wave having a peak pressure of 1 Pa, a peak-to-peak pressure of 2 Pa, a root-mean-square pressure of 0.7 Pa, a period of 0.25 s, and a frequency of 4 Hz. The top plot indicates the motion of the particles of the medium; they undergo coupled oscillations back and forth, so that the sound wave

propagates to the right. At regions of compression, the pressure is high; at regions of rarefaction, it is low. The bottom plot shows the change in pressure over time at a fixed location. While the plots are lined up, the horizontal axes of the top and bottom plots are space and time, respectively

we know very little about auditory perception in animals, the term pitch is not normally used in animal bioacoustics.

The symbol for frequency is f and the unit is hertz [Hz] in honor of Heinrich Rudolf Hertz, a German physicist who proved the existence of electromagnetic waves. Expressed in SI units, $1 \text{ Hz} = 1/\text{s}$.

The fundamental frequency (symbol: f_0 ; unit: Hz) of an oscillation is the reciprocal of the period. The period (symbol: τ ; unit: s) is the duration of one cycle and is related to the fundamental frequency as (see Fig. 4.1):

$$\tau = \frac{1}{f_0}$$

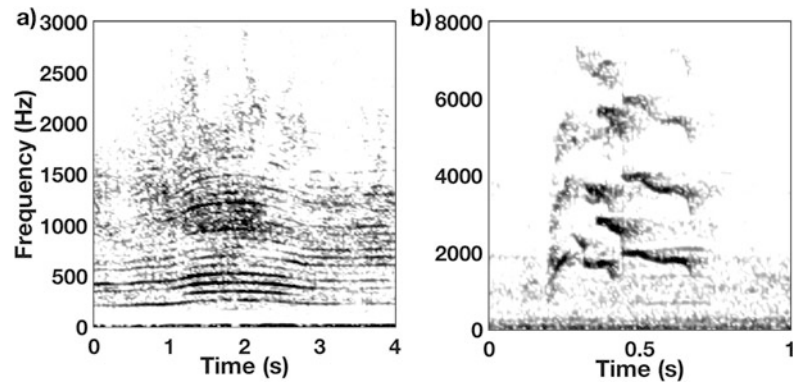
The wavelength (symbol: λ ; unit: m) of a sine wave measures the spatial distance between two successive “peaks” or other identifiable points on the wave.

A sound that consists of only one frequency is commonly called a pure tone. Very often, sounds contain not only the fundamental frequency

but also harmonically related overtones. The frequencies of overtones are integer multiples of the fundamental: $2 f_0$, $3 f_0$, $4 f_0$, ... Beware that there are two schemes for naming these tones: f_0 can be called either the fundamental or the first harmonic. In the former case, $2 f_0$ becomes the first overtone, $3 f_0$ the second overtone, etc. In the latter case, $2 f_0$ becomes the second harmonic, $3 f_0$ the third harmonic, etc.

Musical instruments produce harmonics, which determine the characteristic timbre of the sounds they produce. For example, it is the differences in harmonics that make a flute sound unmistakably different from a clarinet, even when they are playing the same note. Animal sounds also often have harmonics as they use similar basic mechanisms to musical instruments. Most mammals have string-like vocal cords and birds have string-like syrinxes. Fish have muscles that contract around a swim bladder to produce percussive-type sounds. Insects and invertebrates stridulate or rub body parts together to produce a percussive sound.

Fig. 4.2 Spectrograms of (a) a jet ski recorded under water Erbe 2013 and (b) a Carnaby's Cockatoo (*Calyptorhynchus latirostris*) whistle, both displaying frequency modulation



The frequency or frequencies of a sound may change over time, so that frequency is a function of time: $f(t)$. This is called frequency modulation (abbreviation: FM). If the frequency increases over time, the sound is called an upsweep. If the frequency decreases over time, the sound is called a downsweep. Sounds without frequency modulation are called continuous wave. The sound of jet skis under water is frequency-modulated due to frequent speed changes (Erbe 2013). Whistles of animals such as birds or dolphins (e.g., Ward et al. 2016) are commonly frequency-modulated and often exhibit overtones (Fig. 4.2).

The acoustic features of frequency-modulated sounds such as whistles can identify the species, population, and sometimes individual animal that made them (e.g., Caldwell and Caldwell 1965). Such characteristic features include the start frequency, end frequency, minimum frequency, maximum frequency, duration, number of local extrema, number of inflection points, and number of steps (e.g., Marley et al. 2017). The start frequency is the frequency at the beginning of the fundamental contour, the end frequency is the frequency at the end of the fundamental contour (Fig. 4.3). The minimum frequency is the lowest frequency of the fundamental contour and the maximum frequency is the highest. Duration measures how long the whistle lasts. Extrema are points of local minima or maxima in the contour. At a local minimum, the contour changes from downsweep to upsweep; at a local maximum, it changes from upsweep to downsweep. Mathematically, the first derivative of the whistle

contour with respect to time is zero at a local extremum, and the second derivate is a positive number in the case of a minimum or a negative number in the case of a maximum. At an inflection point, the curvature of the contour changes from clockwise to counter-clockwise or vice versa. Mathematically, the first derivative of the whistle contour with respect to time exhibits a local extremum and the second derivative is zero at an inflection point. Steps in the contour are discontinuities in frequency. There is no temporal gap but the contour jumps in frequency. The frequency measurements are taken from the fundamental contour. The duration, number of local extrema, number of inflection points, and number of steps are the same in fundamental and overtones and can therefore be measured from any harmonic contour. This is beneficial if the fundamental is partly masked by noise.

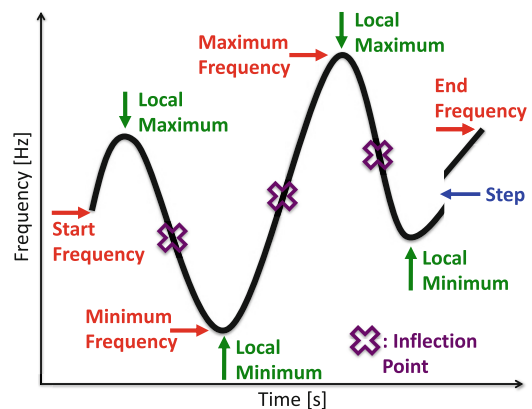


Fig. 4.3 Spectrogram of a frequency-modulated sound, identifying characteristic features

4.2.4 Pressure

Atmospheric pressure is the static pressure at a specified height above ground and is due to the weight of the atmosphere above. Similarly, hydrostatic pressure is the static pressure at a specified depth below the sea surface and is due to the weight of the water above plus the weight of the atmosphere.

Sound pressure (or acoustic pressure) is caused by a sound wave. Sound pressure (symbol: p ; unit: Pa) is dynamic pressure; it varies with time t (i.e., p is a function of t : $p(t)$). It is a deviation from the static pressure and defined as the difference between the instantaneous pressure and the static pressure. Air-borne sound pressure is measured with a microphone, water-borne sound pressure with a hydrophone. The unit of pressure is pascal [Pa] in honor of Blaise Pascal, a French mathematician and physicist. Some of the superseded units of pressure are bar and dynes per square centimeter, which can be converted to pascal: $1 \text{ bar} = 10^6 \text{ dyn/cm}^2 = 10^5 \text{ Pa}$. Mathematically, pressure is defined as force per area. Pascal in SI units is

$$1 \text{ Pa} = 1 \text{ N/m}^2 = 1 \text{ J/m}^3 = 1 \text{ kg}/(\text{m s}^2)$$

where N symbolizes newton, the unit of force, and J symbolizes joule, the unit of energy.

The pressure in Fig. 4.1 follows a sine wave: $p(t) = A \sin(2\pi ft)$, where A is the amplitude and f the frequency. In the example of Fig. 4.1, $A = 1 \text{ Pa}$, $f = 4 \text{ Hz}$. In general terms, the amplitude is the magnitude of the largest departure of a periodically varying quantity (such as sound pressure or particle velocity, see Sect. 4.2.8) from its equilibrium value. The magnitude is always positive and commonly symbolized by two vertical bars: $|p(t)|$. These are the same values as $p(t)$, but without the sign (i.e., the magnitude is always positive). The amplitude may not always be a constant. When it changes as a function of time $A(t)$, the signal undergoes amplitude modulation (abbreviation: AM).

The signal in Fig. 4.4 is both amplitude- and frequency-modulated:

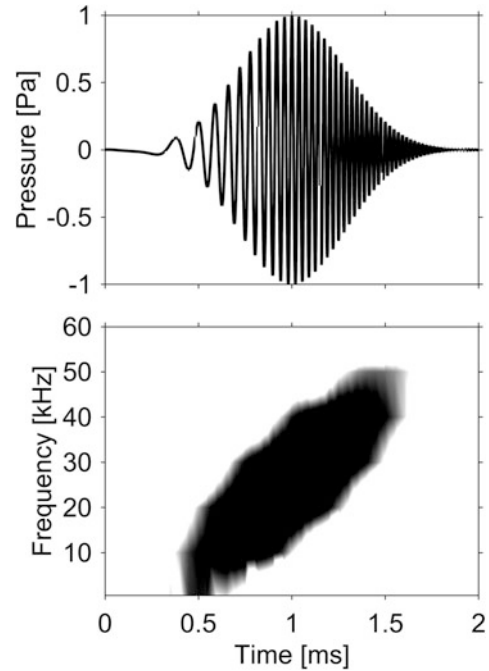


Fig. 4.4 Gabor click similar to a beaked whale click. The signal is based on a sine wave; the amplitude is modulated by a Gaussian function, and the frequency is swept up with time. The corresponding spectrogram is shown in the bottom panel

$$p(t) = A(t) \sin(2\pi f(t) \times t)$$

The amplitude function changes exponentially with time:

$A(t) = e^{-(t-t_0)^2/2\sigma^2}$, where the peak occurs at $t_0 = 1 \text{ ms}$, and σ is the standard deviation of the Gaussian envelope. Such signals (sine waves that are amplitude-modulated by a Gaussian function) are called Gabor signals. Echolocation clicks are commonly of Gabor shape (e.g., Kamminga and Beitsma 1990; Holland et al. 2004). In several species of beaked whales, the sine wave is frequency-modulated (Baumann-Pickering et al. 2013) as in the example in Fig. 4.4, where the frequency changes linearly with time, sweeping up from 10 to 50 kHz.

The peak-to-peak sound pressure (symbol: p_{pk-pk} ; unit: Pa) is the difference between the maximum pressure and the minimum pressure of a sound wave:

$$p_{pk-pk} = \max(p(t)) - \min(p(t))$$

In other words, it is the sum of the greatest magnitude during compression and the greatest magnitude during rarefaction.

The peak sound pressure (symbol: p_{pk} ; unit: Pa) is also called zero-to-peak sound pressure and is the greatest deviation of the sound pressure from the static pressure; it is the greatest magnitude of $p(t)$:

$$p_{pk} = \max(|p(t)|)$$

This can occur during compression and/or rarefaction. In other words, p_{pk} is the greater of the greatest magnitude during compression and the greatest magnitude during rarefaction (Fig. 4.1).

The root-mean-square (rms) is a useful measure for signals (like sound pressure) that aren't simple oscillatory functions. The rms of any signal can be calculated, no matter how complicated it is. To do so, square each sample of the signal, average all the squared samples, and then take the square root of the result. It turns out that the rms of a sine wave is 0.707 times its amplitude, but this is only true for sinusoidal (sine or cosine) waves. The units for rms are the same as those for amplitude (e.g., Pa if the signal is pressure or m/s if the signal is particle velocity). The root-mean-square sound pressure (symbol: p_{rms} ; unit: Pa) is computed as its name dictates, as the root of the mean over time of the squared pressure:

$$p_{rms} = \sqrt{\frac{\int_{t_1}^{t_2} p^2(t) dt}{t_2 - t_1}}, \text{ or in discrete form :}$$

$$p_{rms} = \sqrt{\frac{\sum_{i=1}^N p_i^2}{N}} \quad (4.1)$$

This computation is practically carried out over a time interval from t_1 to t_2 .

The mean-square is the mean of the square of the signal values. The mean-square of a signal is always equal to the square of the signal's rms. Its units are the square of the corresponding amplitude units (e.g., Pa² if the signal is pressure or

(m/s)² if the signal is particle velocity). The mean-square sound pressure formula is similar to (Eq. 4.1) but without the root.

The sound pressure level (abbreviation: SPL; symbol: L_p) is the level of the root-mean-square sound pressure and computed as

$$L_p = 20 \log_{10} \left(\frac{p_{rms}}{p_0} \right)$$

expressed in dB relative to (abbreviated: re) a reference value p_0 . The standard reference value is 20 μPa in air and 1 μPa in water.

The peak sound pressure level (also called zero-to-peak sound pressure level; abbreviation: SPLpk; symbol: $L_{p,pk}$) is the level of the peak sound pressure and computed as

$$L_{p,pk} = 20 \log_{10} \left(\frac{p_{pk}}{p_0} \right)$$

It is expressed in dB relative to a reference value p_0 (i.e., 20 μPa in air and 1 μPa in water). Similarly, the peak-to-peak sound pressure level is the level of the peak-to-peak sound pressure:

$$L_{p,pk-pk} = 20 \log_{10} \left(\frac{p_{pk-pk}}{p_0} \right)$$

Example sound pressure levels in air and water are given in Tables 4.3 and 4.4. Sources can have a large range of levels and only one example is given for each source. Animal sounds and their levels may vary with species, sex, age, behavioral context, etc. Animals in captivity may produce lower levels than animals in the wild. Ship noise depends on the type of vessel, its propulsion system, speed, load, etc. The tables are intended to give an overview of the dynamic range of source levels across the different sources.

Loudness is an attribute of auditory sensation. While it is related to sound pressure, loudness measures how loud or soft a sound seems to us. Given that very little is known about auditory perception in animals, the term loudness is rarely used in animal bioacoustics.

Table 4.3 Examples of sound pressure levels in air. All levels are broadband; the hearing thresholds are single-frequency. Nominal ranges from the source are given in

meters. Note that the different sources listed can have a range of levels and only one example is given

	Pa	dB re 20 μPa
Explosion at 1 m	63,246	190
Airplane take-off at 25 m	632	150
Human pain threshold at 1 kHz	200	140
Lion roar at 1 m	13	116
Human discomfort threshold at 1 kHz	10	114
Diesel lawn mower at 1 m	1	94
Truck at city speed at 20 m	0.2	80
Old vacuum cleaner at 1 m	0.1	70
Bird song at 1 m	0.02	60
Cricket chorus at 1 m	0.02	60
Human speech at 1 m	0.01	55
Buzzing mosquito	0.002	40
Human whisper at 1 m	0.001	30
Fluttering leaves	0.0002	20
Human breathing at 1 m	0.0001	10
Human hearing threshold at 1 kHz	0.00002	0

Table 4.4 Examples of sound pressure levels in water. All levels are broadband; the hearing thresholds are single-frequency. Nominal ranges from the source are given in

meters. Note that the different sources listed can have a range of levels and only one example is given

	Pa	dB re 1 μPa
Subsea earthquake	316,228	230
Seismic survey airgun at 1 m	10,000	200
Container ship at 1 m	5623	195
Humpback whale song at 1 m	1778	185
Zodiac at high speed at 1 m	178	165
Dolphin whistle at 1 m	32	150
Geotechnical drilling at 1 m	18	145
Jet ski	10	140
Toadfish at 1 m	10	140
Damsel fish at 1 m	1	120
Open ocean ambient noise at sea state 4	0.1	100
Open ocean ambient noise at sea state 0.5	0.01	80
California Sea lion hearing threshold at 10 kHz	0.001	60
Killer whale hearing threshold at 20 kHz	0.0001	40

4.2.5 Sound Exposure

Sound exposure (symbol: $E_{p,T}$; unit: Pa²s) is the integral over time of the squared pressure:

$$E_{p,T} = \int_{t_1}^{t_2} p^2(t) dt$$

Sound exposure increases with time. The longer the sound lasts, the greater the exposure. The

sound exposure level (abbreviation: SEL; symbol: $L_{E,p}$) is computed as:

$$L_{E,p} = 10 \log_{10} \left(\frac{E_{p,T}}{E_{p,0}} \right)$$

It is expressed in dB relative to $E_{p,0} = 400 \mu\text{Pa}^2\text{s}$ in air, and $E_{p,0} = 1 \mu\text{Pa}^2\text{s}$ in water. Sound exposure is proportional to the total energy of a sound wave.

4.2.6 When to Use SPL and SEL?

Sound pressure and sound exposure are closely related, and in fact, the sound exposure level can be computed from the sound pressure level as:

$$L_{E,p} = L_p + 10 \log_{10}(t_2 - t_1)$$

Conceptually, the difference is that the SPL is a time-average and therefore useful for sounds that don't change significantly over time, or that last for a long time, or that, for the assessments of noise impacts, can be considered continuous. Examples are workplace noise or ship noise. The SEL, however, increases with time and critically depends on the time window over which it is computed. It is therefore most useful for short-duration, transient sounds, such as pulses from explosions, pile driving, or seismic surveys. The SEL is then computed over the duration of the pulse.

It can be difficult to determine the actual pulse length as the exact start and end points are often not clearly visible, in particular in background noise. Therefore, in praxis, SEL is commonly computed over the 90% energy signal duration. This is the time during which 90% of the sound exposure occurs. Sound exposure is computed symmetrically about the 50% mark; i.e., from the 5% to the 95% points on the cumulative squared-pressure curve. SEL becomes (Fig. 4.5):

$$L_{E,p} = 10 \log_{10} \left(\frac{\int_{t_{5\%}}^{t_{95\%}} p^2(t) dt}{E_{p,0}} \right)$$

In the presence of significant background noise $p_n(t)$, the noise exposure needs to be subtracted from the overall sound exposure in order to yield the sound exposure due to the signal alone. In praxis, the noise exposure is computed over an equally long time window (from t_1 to t_2) preceding or succeeding the signal of interest:

$$L_{E,p} = 10 \log_{10} \left(\frac{\int_{t_{5\%}}^{t_{95\%}} p^2(t) dt - \int_{t_1}^{t_2} p_n^2(t) dt}{E_{p,0}} \right)$$

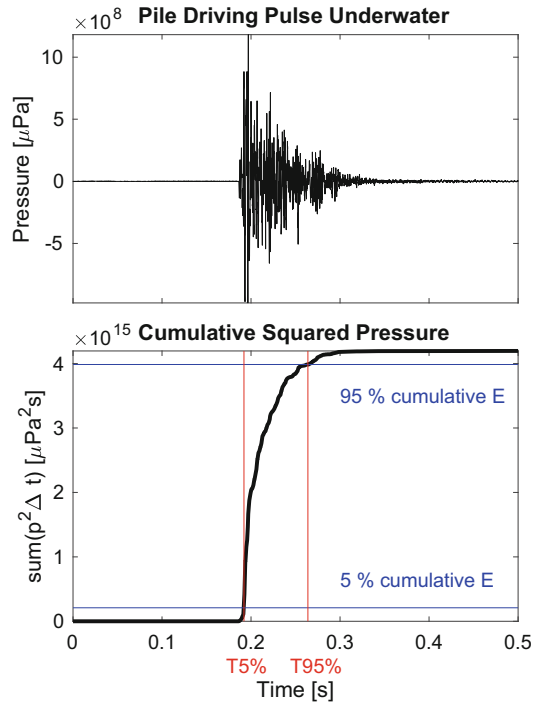


Fig. 4.5 Pressure pulse recorded from pile driving underwater (top) and cumulative squared-pressure curve (bottom). The horizontal lines indicate the 5% and 95% cumulative squared-pressure points on the y-axis. The vertical lines identify the corresponding times on the x-axis. The time between the 5% and 95% marks is the 90% energy signal duration. Recording from Erbe 2009

4.2.7 Acoustic Energy, Intensity, and Power

Apart from sound pressure and sound exposure, other physical quantities appear in the bioacoustics literature, but are often wrongly used. Acoustic energy refers to the total energy contained in an acoustic wave. This is the sum of kinetic energy (contained in the movement of the particles of the medium) and potential energy (i.e., work done by elastic forces in the medium). Acoustic energy E is proportional to squared pressure p and time interval Δt (i.e., to sound exposure) only in the case of a free plane wave or a spherical wave at a large distance from its source:

$$E = \frac{S}{Z} p^2 \Delta t$$

The proportionality constant is the ratio of surface area S through which the energy flows and acoustic impedance Z . Acoustic energy increases with time; i.e., the longer the sound lasts or the longer it is measured, the greater the transmitted energy. The unit of energy is joule [J] in honor of English physicist James Prescott Joule. In SI units:

$$1 \text{ J} = 1 \text{ kg m}^2/\text{s}^2$$

Acoustic power P is the amount of acoustic energy E radiated within a time interval Δt :

$$P = E/\Delta t$$

The unit of power is watt [W]. In SI units:

$$1 \text{ W} = 1 \text{ J/s} = 1 \text{ kg m}^2/\text{s}^3$$

Acoustic intensity I is the amount of acoustic energy E flowing through a surface area S perpendicular to the direction of propagation, per time Δt :

$$I = E/(S\Delta t) = P/S$$

For a free plane wave or a spherical wave at a large distance from its source, this becomes:

$$I = p^2/Z \quad (4.2)$$

The unit of intensity is W/m^2 . A conceptually different definition equates the instantaneous acoustic intensity with the product of sound pressure and particle velocity u :

$$I(t) = p(t) u(t)$$

The two concepts are mathematically equivalent for free plane and spherical waves and the unit of intensity is always W/m^2 .

The above quantities (energy, power, and intensity) are sometimes used interchangeably. That's wrong. They are not the same, but they are related. With E , P , I , S , and t denoting energy, power, intensity, surface area, and time, respectively:

$$P = E/\Delta t = I S$$

More information and definitions can be found in acoustic standards (including American National Standards Institute 2013; International Organization for Standardization 2017).

4.2.8 Particle Velocity

Particle velocity (symbol: u ; unit: m/s) refers to the oscillatory movement of the particles of the acoustic medium (i.e., molecules in air and water, and atoms in the ground) as a wave passes through. In the example of Fig. 4.1, the particle velocity is a sine wave, just like the acoustic pressure. Each particle oscillates about its equilibrium position. At this point, its displacement is zero, but its velocity is greatest (i.e., either maximally positive or maximally negative, depending on the direction in which the particle is moving). At the two turning points, the displacement from the equilibrium position is maximum and the velocity passes through zero, changing sign (i.e., direction) from positive to negative, or vice versa. Velocity is a vector, which means it has both magnitude and direction. Particle displacement (unit: m) and particle acceleration (unit: m/s^2) are also vector quantities. In fact, particle velocity is the first derivative of particle displacement with respect to time, and particle acceleration is the second derivative of particle displacement with respect to time. Measurements of particle displacement, velocity, and acceleration created by snorkeling are shown in Fig. 4.6.

Air molecules also move due to wind, and water molecules move due to waves and currents. But these types of movement are not due to sound. Wind velocity and current velocity are entirely different from the oscillatory particle velocity involved in the propagation of sound.

It is equally important to understand that the speed at which the particles move when a sound wave passes through is not equal to the speed of sound at which the sound wave travels through the medium. The latter is not an oscillatory quantity.

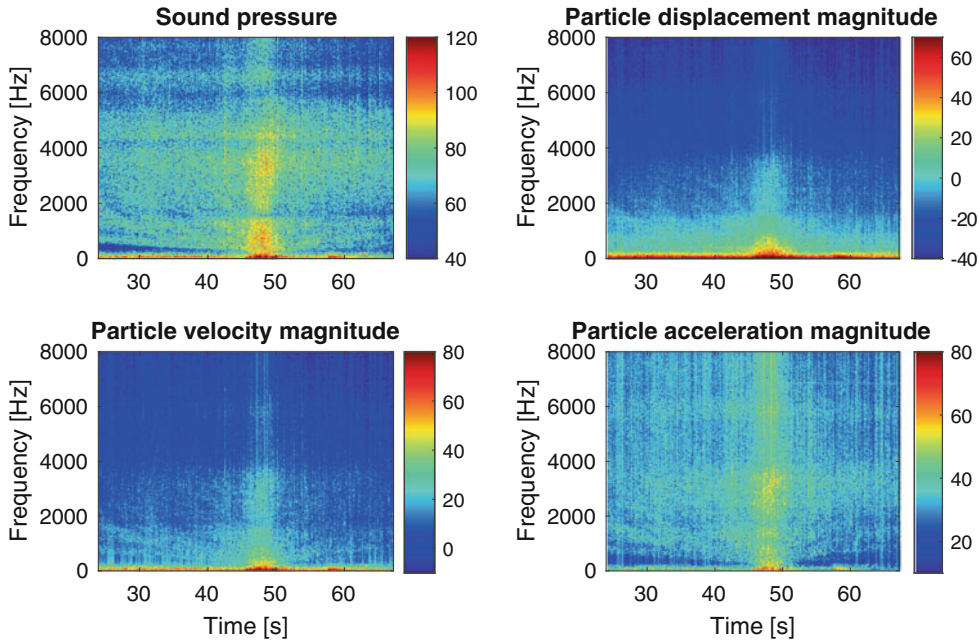


Fig. 4.6 Spectrograms of mean-square sound pressure spectral density [dB re $1 \mu\text{Pa}^2/\text{Hz}$], mean-square particle displacement spectral density [dB re $1 \text{pm}^2/\text{Hz}$], mean-square particle velocity spectral density [dB re $1 (\text{nm/s})^2/\text{Hz}$], and mean-square particle acceleration spectral density [dB re $1 (\mu\text{m/s}^2)^2/\text{Hz}$] recorded under water when a snorkeler swam above the recorder (Erbe et al. 2016b; Erbe et al. 2017a)

Hz], and mean-square particle acceleration spectral density [dB re $1 (\mu\text{m/s}^2)^2/\text{Hz}$] recorded under water when a snorkeler swam above the recorder (Erbe et al. 2016b; Erbe et al. 2017a)

4.2.9 Speed of Sound

The speed at which sound travels through an acoustic medium is called the speed of sound (symbol: c ; unit: m/s). It depends primarily on temperature and height above ground in air, and on temperature, salinity, and depth below the sea surface in water. The speed of sound is computed as the distance sound travels divided by time. It can also be computed from measurements of the waveform (i.e., wavelength, period, and frequency as in Fig. 4.1):

$$c = \lambda/\tau = \lambda f$$

In solid media, such as rock, two types of waves are supported, P- and S-waves (see Sect. 4.2.2), and the speeds (c_P and c_S) at which they travel differ. Table 4.5 gives examples for the speed of sound in air and water, and for P- and S-waves in some Earth materials. Example sound speed profiles (i.e., line graphs of sound speed

versus altitude or water depth) are given in Fig. 4.7.

4.2.10 Acoustic Impedance

Each acoustic medium has a characteristic impedance (symbol: Z). It is the product of the medium's density (symbol: ρ) and speed of sound: $Z = \rho c$. In air at 0°C with a density $\rho = 1.3 \text{ kg/m}^3$ and speed of sound $c = 330 \text{ m/s}$, the characteristic impedance is $Z = 429 \text{ kg}/(\text{m}^2\text{s})$. In freshwater at 5°C with a density of $\rho = 1000 \text{ kg/m}^3$ and a speed of sound $c = 1427 \text{ m/s}$, the characteristic impedance is $Z = 1,427,000 \text{ kg}/(\text{m}^2\text{s})$. In sea water at 20°C and 1 m depth with 3.4% salinity, a density of $\rho = 1035 \text{ kg/m}^3$, and a speed of sound of $c = 1520 \text{ m/s}$, the characteristic impedance is $Z = 1,573,200 \text{ kg}/(\text{m}^2\text{s})$. The characteristic impedance relates the sound pressure to particle velocity via $p = Z u$ for plane waves.

Table 4.5 P-wave and S-wave speeds of certain acoustic media

Medium	c_P [m/s]	c_S [m/s]
Air, 0 °C	330	
Air, 20 °C	343	
Freshwater, 5 °C	1427	
Freshwater, 20 °C	1481	
Salt water, 20 °C, salinity 3.4%, 1 m depth	1520	
Sand	800–2200	
Clay	1000–2500	
Sandstone	1400–4300	700–2800
Granite	5500–5900	2800–3000
Limestone	5500–6100	2800–3300

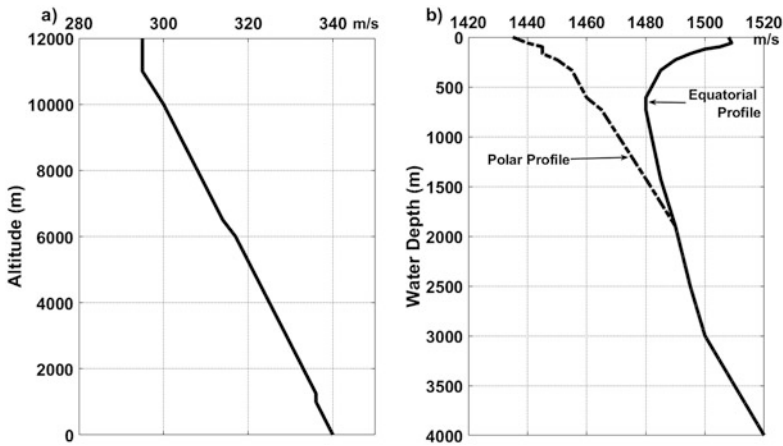


Fig. 4.7 Example profiles of the speed of sound in (a) air (data from The Engineering ToolBox; https://www.engineeringtoolbox.com/elevation-speed-sound-air-d_1534.html; accessed 16 April 2021) and (b) water in polar and equatorial regions (These data were collected and made freely available by the International Argo Program

and the national programs that contribute to it; <https://argo.ucsd.edu>, <https://www.ocean-ops.org>. The Argo Program is part of the Global Ocean Observing System. Argo float data and metadata from Global Data Assembly Centre (Argo GDAC); <https://doi.org/10.17882/42182>; accessed 16 April 2021). See Chaps. 5 and 6

4.2.11 The Decibel

Acousticians may deal with very-high-amplitude signals and very-low-amplitude signals; e.g., the sound pressure near an explosion might be 60,000 Pa, while the sound pressure from human breathing is only 0.0001 Pa. This means that the dynamic range of quantities in acoustics is large and, in fact, covers seven orders of magnitude (see Tables 4.3 and 4.4). Rather than handling multiple zeros and decimals, using a logarithmic scale compresses the dynamic range

into a manageable range of values. This is one of the reasons why the decibel is so popular in acoustics. Another reason is that human perception of the loudness of a sound is approximately proportional to the logarithm of its amplitude.

When quantities such as sound pressure or sound exposure are converted to logarithmic scale, the word “level” is added to the name. Sound pressure level and sound exposure level are much more commonly used than their linear counterparts, sound pressure and sound exposure.

By definition, the level L_Q of quantity Q is proportional to the logarithm of the ratio of Q and a reference value Q_0 , which has the same unit. In the case of a field quantity F , such as sound pressure or particle velocity, or an electrical quantity such as voltage or current, the level L_F is computed as

$$L_F = 20 \log_{10} \frac{F}{F_0}$$

In the case of a power quantity P , such as mean-square sound pressure or energy, the level L_P is computed as

$$L_P = 10 \log_{10} \frac{P}{P_0}$$

Both levels are expressed in decibels (dB). Note the different factors (20 versus 10) in the equations. It is critically important to always state the reference value F_0 or P_0 when discussing levels, because reference values differ between air and water.

4.2.11.1 Conversion from Decibel to Field or Power Quantities

The relationships for calculating field and power quantities from their levels are, respectively:

$$F = 10^{\frac{L_F}{20}} F_0, \text{ and } P = 10^{\frac{L_P}{10}} P_0 \quad (4.3)$$

The units of the calculated quantities correspond to the units of the reference quantity (F_0

or P_0). For example, an underwater tone at a level of 120 dB re 1 μPa rms has an rms pressure of 1 Pa. This is worked out as follows:

$$F = 10^{120/20} \times 1 \mu\text{Pa} = 10^6 \mu\text{Pa} = 1 \text{ Pa}$$

However, a tone of 120 dB re 20 μPa rms in air has an rms pressure of 20 Pa:

$$F = 10^{120/20} \times 20 \mu\text{Pa} = 10^6 \cdot 20 \mu\text{Pa} = 20 \text{ Pa}$$

4.2.11.2 Differences between Levels of like Quantities

A particular difference between two levels corresponds to particular ratios between their field and power quantities. The general relationships are:

$$L_{F1} - L_{F2} = 20 \log_{10} \frac{F_1}{F_2}$$

$$L_{P1} - L_{P2} = 10 \log_{10} \frac{P_1}{P_2}$$

$$\frac{F_1}{F_2} = 10^{\left(\frac{L_{F1} - L_{F2}}{20}\right)}$$

$$\frac{P_1}{P_2} = 10^{\left(\frac{L_{P1} - L_{P2}}{10}\right)}$$

Some common examples are given in Table 4.6. Note the inverse relationship between ratios for corresponding positive and negative level differences and also that each power

Table 4.6 Level differences and their corresponding field and power quantity ratios

Level difference ($L_{F1} - L_{F2}$ or $L_{P1} - L_{P2}$) in dB	Field quantity ratio (F_1/F_2); use for pressure, particle velocity, voltage, current, etc.	Power quantity ratio (P_1/P_2); use for power, intensity, energy, sound exposure, mean-square pressure, etc.
-40	1/100 = 0.01	1/10,000 = 0.0001
-20	1/10 = 0.1	1/100 = 0.01
-10	1/ $\sqrt{10}$ \approx 0.316	1/10 = 0.1
-6	1/2 = 0.5	1/4 = 0.25
-3	1/ $\sqrt{2}$ \approx 0.707	1/2 = 0.5
0	1	1
3	$\sqrt{2}$ \approx 1.41	2
6	2	4
10	$\sqrt{10}$ \approx 3.16	10
20	10	100
40	100	10,000

quantity ratio is the square of the corresponding field quantity ratio.

For example, a tone at a level of 120 dB re 1 μPa rms is 20 dB stronger than a tone at a level of 100 dB re 1 μPa rms, so from Table 4.6, the ratio of the two rms pressures is $p_1/p_2 = F_1/F_2 = 10$, and the ratio of their intensities is $I_1/I_2 = P_1/P_2 = 100$.

4.2.11.3 Amplification of Signals

The above formulae and Table 4.6 can also be used to calculate the effect of amplifying signals. For example, if an amplifier has a gain of 20 dB, then the rms voltage at the output of the amplifier will be 10 times the rms voltage at its input. Similarly, an amplifier with a 40 dB gain will increase the rms voltage by a factor of 100. If several amplifier stages are cascaded, then their combined gain is the sum of the gains of the individual stages (in dB).

When calibrating acoustic recordings (see Chap. 2), the gains of all components of the recording systems have to be summed. An underwater recording system (Fig. 4.8), for example, contains a hydrophone that converts received acoustic pressure to a time series of voltages at its output. The sensitivity of the hydrophone specifies this relationship. For example, a hydrophone with a sensitivity $N_S = -180$ dB re 1 V/ μPa produces $10^{-180/20} = 10^{-9}$ Volts output per 1 μPa input. A more sensitive hydrophone has a less negative sensitivity. The output voltage might be passed to an amplifier with $\Delta L_G = 20$ dB gain, after which it is digitized by a data acquisition board, such as a computer's soundcard. All

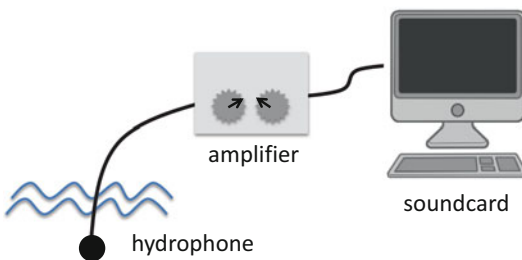


Fig. 4.8 Sketch of an example underwater recording setup. A terrestrial setup would have a microphone instead of a hydrophone

analog-to-digital converters have a digitization gain expressed in dB re FS/V, which specifies the input voltage that leads to full scale (FS). If the digitizer has a digitization gain $\Delta L_{DG} = 10$ dB re FS/V, then $10^{10/20}$ FS/V = $10^{1/2}$ FS/V is the relationship between FS and input voltage, meaning that FS is reached when the input is $1/10^{1/2}$ V = 0.32 V. The actual value of FS depends on the number of bits available. A 16-bit digitizer in bipolar mode (i.e., producing both positive and negative numbers) has a full-scale value of $2^{16-1} = 2^{15} = 32,768$. And so the digital values v representing the acoustic pressure will lie between $-32,768$ and $+32,767$ (with one of the possible numbers being 0). The final steps in relating these digital values to the recorded acoustic pressure entail dividing by FS, converting to dB, and subtracting all the gains:

$$\begin{aligned} L_p &= 20 \log_{10}(v/\text{FS}) - \Delta L_{DG} - \Delta L_G - N_S \\ &= 20 \log_{10}(v/\text{FS}) + 150 \text{ dB re } 1 \mu\text{Pa} \end{aligned}$$

4.2.11.4 Superposition of Field and Power Quantities

If two tones of the same frequency and level arrive in phase at a listener, then the amplitude is doubled and the combined level is therefore 6 dB above the level of each tone (see Table 4.6). If, on the other hand, there is a random phase difference between the two tones then, on average, the intensity of the two signals will sum. In this case (again from Table 4.6) the combined intensity is 3 dB higher than the level of each tone. For example, if each tone has a level of 120 dB re 1 μPa rms, then the two tones together have a level of 126 dB re 1 μPa rms if they are in phase. Their superposition has an average level of 123 dB re 1 μPa rms if they have a random phase difference. Summing signals that have the same phase, or a fixed phase difference, is known as coherent summation, whereas performing an “on average” summation of signals assuming a random phase is called incoherent summation.

The calculation is more complicated if the two tones have different levels. It is necessary to use Eq. (4.3) to convert both levels to corresponding

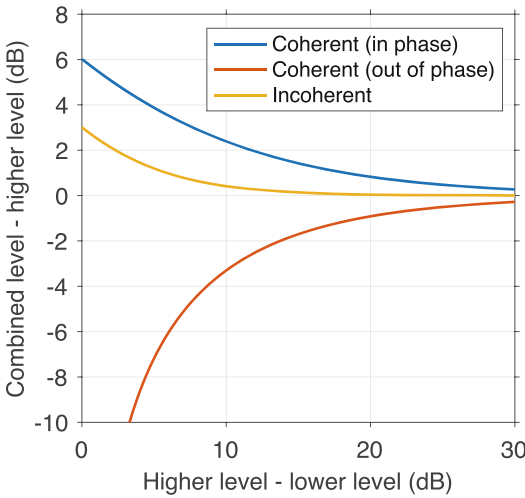


Fig. 4.9 Line graphs of the effect on the higher-level signal of combining two signals by coherent summation (assuming the signals are in phase or 180° out of phase) and incoherent summation

field (coherent summation) or power (incoherent summation) quantities, add these quantities, and then convert the result back to a level.

The outcome of this process is plotted in Fig. 4.9 in terms of the increase in the combined level from that of the higher-level signal as a function of the difference between the higher and lower levels. Note that this increase never exceeds 6 dB for a coherent summation or 3 dB for an incoherent summation. In the case of a coherent summation, proper account has to be taken of the relative phases of the two tones when adding the field quantities, and this can have a very large effect. Figure 4.9 shows the extreme cases: The upper limit occurs when the two signals are in phase, and the lower limit occurs when they have a phase difference of 180° (π radians). The latter case gives destructive interference and the combined level is lower than that of the highest individual signal. If the two individual signals have a 180° phase difference and the same amplitude, then the destructive interference is complete, the two signals cancel each other out, and the combined level is $-\infty$!

Another useful observation from Fig. 4.9 is that when the difference in level between the two individual signals is greater than 10 dB, the

incoherent summation is less than 0.5 dB higher than that of the higher of the two; and for many practical applications, the lower-level signal can be ignored.

4.2.11.5 Levels in Air Versus Water

Comparing sound levels in air and water is complicated and has caused much confusion in the past. For two sound sources of equal intensity I_a and I_w in air and water, respectively, the sound pressure level is 62 dB greater in water because of two factors: the greater acoustic impedance of water and the different reference pressures used in the two media.

The effect of the acoustic impedance can be seen as follows. Assuming $I_w = I_a$, then from (Eq. 4.2):

$$\frac{p_w^2}{Z_w} = \frac{p_a^2}{Z_a}, \text{ which is equivalent to } \frac{p_w}{p_a} = \frac{Z_w}{Z_a}.$$

This ratio of mean-square pressures in the two media can be expressed in terms of the density and speed of sound of the two media:

$$\frac{p_w}{p_a} = \frac{Z_w}{Z_a} = \frac{\rho_w c_w}{\rho_a c_a}.$$

Applying $10 \log_{10}()$ to these ratios, the difference between the mean-square sound pressure levels in water and air is:

$$\begin{aligned} L_{pw^2} - L_{pa^2} &= 10 \log_{10} \frac{p_w^2}{p_0^2} - 10 \log_{10} \frac{p_a^2}{p_0^2} \\ &= 10 \log_{10} \frac{p_w^2}{p_a^2} = 10 \log_{10} \frac{\rho_w c_w}{\rho_a c_a} \\ &= 36 \text{ dB} \end{aligned}$$

The difference between the sound pressure levels is, of course, also 36 dB:

$$\begin{aligned} L_{pw} - L_{pa} &= 20 \log_{10} \frac{p_w}{p_0} - 20 \log_{10} \frac{p_a}{p_0} \\ &= 20 \log_{10} \frac{p_w}{p_a} = 20 \log_{10} \sqrt{\frac{\rho_w c_w}{\rho_a c_a}} \\ &= 36 \text{ dB} \end{aligned}$$

In the above two equations, the same reference pressure p_0 is required. However, the convention

is to use $p_{a0}=20 \mu\text{Pa}$ in air and $p_{w0}=1 \mu\text{Pa}$ in water. The difference in reference pressures adds another 26 dB to the sound pressure level in water, because:

$$20 \log_{10} \frac{p_{a0}}{p_{w0}} = 20 \log_{10} \frac{20 \mu\text{Pa}}{1 \mu\text{Pa}} = 26 \text{ dB}$$

So, if two sound sources emit the same intensity in air and water, then the sound pressure level in water referenced to $1 \mu\text{Pa}$ is 62 dB (i.e., 36 dB + 26 dB) greater than the sound pressure level in air referenced to $20 \mu\text{Pa}$.

While this might be confusing, there would hardly be a sensible reason to compare levels in air and water. Such comparisons have been attempted in the past to give an analogy to levels with which humans have experience in air. For example, humans find 114 dB re $20 \mu\text{Pa}$ annoying and 140 dB re $20 \mu\text{Pa}$ painful, so what would be a similarly annoying level under water that might disturb animals?

But animals perceive sound differently from humans, hear sound at different frequencies and levels, and can have rather different auditory anatomy (see Chap. 10 on audiograms). As a result, a signal easily heard by a human could be barely audible to some animals or much louder to others. Even for divers, sound reception under water is quite a different process from sound reception in air, due to different acoustic impedance ratios of the acoustic medium and human tissues, and different sound propagation paths. Furthermore, the psychoacoustic effects (emotional impacts) of different types of noise on animals have not been examined thoroughly. Even in humans, for example, 110 dB re $20 \mu\text{Pa}$ of rock music does not provide the same experience as 110 dB re $20 \mu\text{Pa}$ of traffic noise.

4.2.12 Source Level

The source level (abbreviation: SL; symbol: L_S) is meant to be characteristic of the sound source and independent of both the environment in which the source operates and the method by which the source level is determined. In praxis, the determination of the source level has numerous problems.

Some sources are large in their physical dimensions and placing a recorder at short range (i.e., into the so-called near-field, see Sect. 4.2.13) will not result in a level that captures the full output of the source. Also, many sound sources do not operate in a free-field but rather near a boundary (e.g., air-ground, air-water, or water-seafloor). At such boundaries, reflection, scattering, absorption, and phase changes may occur, affecting the recorded level. In praxis, a sound source is recorded at some range in the far-field and an appropriate (and sometimes sophisticated) sound propagation model is utilized to account for the effects of the environment in order to compute a source level that is independent of the environment. Such source levels can then be applied to new situations and different environments in order to predict received levels elsewhere. Like other levels, the source level is expressed in dB relative to a reference value. It is further referenced to a nominal distance of 1 m from the source. The source level can be a sound pressure level or a sound exposure level, depending on the source and situation.

The radiated noise level (abbreviation: RNL; symbol L_{RN}) is more easily determined. It is the level of the product of the sound pressure and the range r at which the sound pressure is recorded, and it can be calculated as the received sound pressure level L_p plus a spherical propagation loss term:

$$L_{RN} = 20 \log_{10} \frac{p_{rms}(r)r}{p_0 r_0} = L_p + 20 \log_{10} \frac{r}{r_0}$$

It is expressed in dB relative to a reference value of $p_0 r_0 = 20 \mu\text{Pa m}$ in air and $p_0 r_0 = 1 \mu\text{Pa m}$ in water. The radiated noise level is dependent upon the environment and is therefore also called affected source level. Note that it is very common in the bioacoustic literature to report source levels and radiated noise levels as dB re $20 \mu\text{Pa} @ 1 \text{ m}$ in air and dB re $1 \mu\text{Pa} @ 1 \text{ m}$ in water. The ISO definition is mathematically different and the notation excludes “@ 1 m” (International Organization for Standardization 2017).

While the source level can be characteristic of the source, there are many factors that affect the source level. For example, larger ships typically

have a higher source level than smaller ships. Cars going fast have a higher source level than cars going slowly. Animals can vary the amplitude of the same sound depending on the context and their motivation. Different sound types can have different source levels. Territorial defense or aggressive sounds usually have the highest source level in a species' repertoire. Mother-offspring sounds often have the lowest source level in a species' repertoire, because mother and calf are typically close together and want to avoid detection by predators.

4.2.13 What Field? Free-Field, Far-Field, Near-Field

While this might read like the opening of a Dr. Seuss book, it is quite important to understand these concepts. The free-field, or free sound field, exists around a sound source placed in a homogeneous and isotropic medium that is free of boundaries. Homogeneous means that the medium is uniform in all of its parameters; isotropic means that the parameters do not depend on the direction of measurement. While the free-field assumption is commonly applied to estimates of particle velocity from pressure measurements or estimates of propagation loss, sound sources and receivers are rarely in a free-field. More often, sound sources and receivers are near a boundary. This is the case for sources such as trains or construction sites and for receivers such as humans, all of which are right at the air-ground boundary. This is also the case for sources such as ships at the water surface and for receivers such as fishes in shallow water, where they are near two boundaries: the air-water and the water-seafloor boundaries. At boundaries, some of the sound is transmitted into the other medium, some of it is reflected, some of it is scattered in various directions. For more detail on source-path-receiver models in air and water, see Chaps. 5 and 6.

The far-field is the region that is far enough from the source so that the particle velocity and pressure are effectively in phase. The near-field is the region closer to the source where they become

out of phase either because sound from different parts of the source arrives at different times (This is the case of an extended source.) or because the curvature of the spherical wavefront from the source is too great to be ignored (This is the case of a source small enough to be considered a point source.). These two cases have different frequency dependence with the near-field to far-field transition distance increasing with increasing frequency for an extended source, and decreasing with increasing frequency for a small source. A single source may behave as a small source at low frequencies and as an extended source at high frequencies, which implies that there is some non-zero frequency at which it will have a minimum near-field to far-field transition distance. This has resulted in much confusion.

When is a sound source small versus extended? A sound source can be considered small when its physical dimensions are small compared to the acoustic wavelength. A fin whale (*Balaenoptera physalus*) with a head size of perhaps 6 m produces a characteristic 20-Hz signal that has a wavelength of about 70 m and so the whale can be considered small.

When studying the effects of noise on animals, however, the noise sources one deals with are mostly extended sources. In the near-field, the amplitudes of field and power quantities are affected by the physical dimension of the sound source. This is because the surface of an extended sound source can be considered an array of separate point sources. Each point source generates an acoustic wave. At any location, the instantaneous pressure (as an example of a field quantity) is the summation of the instantaneous pressures from all of the point sources. In the near-field, the various sound waves have traveled various distances and arrive at various phases. Therefore, the near-field consists of regions of destructive and constructive interference and the pressure amplitude depends greatly on where exactly in the near-field it is measured. There may be regions close to a sound source where the pressure amplitude is always zero. The interference pattern depends on the frequency of the sound, and the regions of destructive and constructive interference will be different depending on the

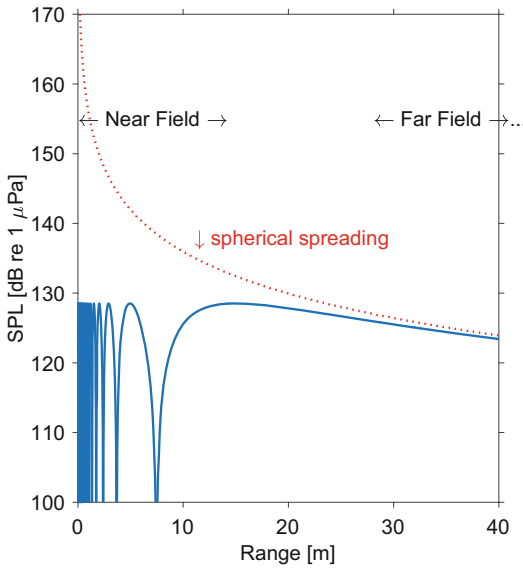


Fig. 4.10 Graph of sound pressure versus range, perpendicular from a circular piston such as a loudspeaker with radius 1 m, $f = 22$ kHz, under water

frequency of the sound. In the far-field of the extended source, the sound waves from the separate point sources have traveled nearly the same distance and arrive in phase. The pressure amplitude depends only on the range from the source and decreases monotonically with increasing range. The amplitudes of field quantities F and power quantities P decay with range r as:

$$F(r) \sim \frac{1}{r} \text{ and } P(r) \sim \frac{1}{r^2} \text{ in the far-field.}$$

The range at which the field transitions from near to far can be estimated as L^2/λ , where L is the largest dimension of the source and λ is the wavelength of interest. (Fig. 4.10).

All sound sources have near- and far-fields. The source level of a sound source is, in praxis, determined from measurements in the far-field by correcting for propagation loss. In the example of Fig. 4.10, the sound pressure level might be measured as 126 dB re 1 μPa at 30 m range from the source. A spherical propagation loss term ($20 \log_{10} \frac{r}{r_0} = 30$ dB; red dashed line in Fig. 4.10) is then applied to estimate the radiated noise level: 156 dB re 1 μPa m. This level is

higher than what would be measured with a receiver in the near-field (blue solid line in Fig. 4.10).

Radiated noise levels and source levels are useful to estimate the received level at some range in the far-field. They will always be higher than the levels that exist in the near-field. There has been a lot of confusion about this in the bioacoustics community, for example in the case of marine seismic surveys. A seismic airgun array (i.e., a number of separate seismic airguns arranged in a 2-dimensional array) might have physical dimensions of several tens of meters and a source level (in terms of sound exposure) of 220 dB re 1 $\mu\text{Pa}^2\text{s}$ m (e.g., Erbe and King 2009). However, *in situ* measurements near the array may never exceed 190 dB re 1 $\mu\text{Pa}^2\text{s}$, except in the immediate vicinity ($\ll 1$ m) of an individual airgun. This is because the highest level that may be recorded is close to an individual airgun in the array. The other airguns in the array are too far away to significantly add to the level of any particular airgun (see Fig. 4.9). At short range from the array, the sound waves from some airguns will add constructively and from others destructively, so that the measured pressure amplitude is always less than the amplitude from one airgun multiplied by the number of airguns in the array. Constructive superposition of sound waves from all airguns only happens in the far-field, where the pressure amplitude is reduced due to propagation loss.

4.2.14 Frequency Weighting

Frequency weightings are mathematical functions applied to sound measurements to compensate quantitatively for variations in the auditory sensitivity of humans and non-human animals (see Chap. 10 on audiometry). These functions “weight” the contributions of different frequencies to the overall sound level, de-emphasizing frequencies where the subject’s auditory sensitivity is less and emphasizing frequencies where it is greater. Frequency weighting essentially applies a band-pass filter to the sound. Weighting is applied before the

calculation of broadband SPLs or SELs. A number of weighting functions exist for different purposes: for example, A, B, C, D, Z, FLAT, and Linear frequency weightings to measure the effect of noise on humans. However, at present, only weightings A, C, and Z are standardized (International Electrotechnical Commission 2013).

4.2.14.1 A, C, and Z Frequency Weightings

A, C, and Z frequency weightings are derived from standardized equal-loudness contours. These are curves which demonstrate SPL variations over the frequency spectrum for which constant loudness is perceived (Suzuki and Takeshima 2004). Loudness is the human perception of sound pressure. Loudness levels are measured in units of phons, determined from referencing the equal-loudness contours. The number of phons n is equal in intensity to a 1-kHz tone with an SPL of n dB. The equal-loudness contours were developed from human loudness perception studies (Fletcher and Munson 1933; Robinson and Dadson 1956; Suzuki and Takeshima 2004) and are standardized (International Organization for Standardization 2003). Table 4.7 defines the A, C, and Z-weighting values at frequencies up to 16 kHz. Figure 4.11 displays the contours of the weightings.

A-weighting is the primary weighting function for environmental noise assessment. It covers a broad range of frequencies from 20 Hz to 20 kHz.

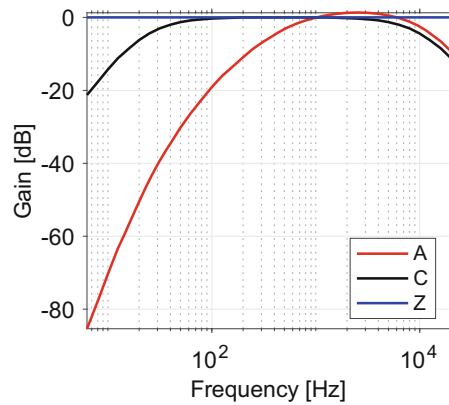


Fig. 4.11 Graph of A-, C-, and Z-weighting curves

The function is tailored to the perception of low-level sounds and represents an idealized human 40-phon equal-loudness contour. Measurements are noted as dB(A) or dBA.

The C-weighting function provides a better representation of human auditory sensitivity to high-level sounds. This weighting is useful for stipulating peak or impact noise levels and is used for the assessment of instrument and equipment noise.

The Z-weighting function (also known as the zero-weighting function) covers a range of frequencies from 8 Hz to 20 kHz (within ± 1.5 dB), replacing the “FLAT” and “Linear” weighting functions. It adds no “weight” to account for the auditory sensitivity of humans and is commonly used in octave-band analysis to analyze the sound source rather than its effect.

Table 4.7 A, C, and Z-weighting values

Frequency [Hz]	A-weighting [dB]	C-weighting [dB]	Z-weighting [dB]
63	-26.2	-0.8	0
125	-16.1	-0.2	0
250	-8.6	0	0
500	-3.2	0	0
1000	0	0	0
2000	1.2	-0.2	0
4000	1	-0.8	0
8000	-1.1	-3	0
16,000	-6.6	-8.5	0

4.2.14.2 Frequency Weightings for Non-human Animals

Equal-loudness contours for non-human animals are very challenging to develop as it is difficult to obtain the required data. Direct measurements of equal loudness in non-human animals have only been achieved for bottlenose dolphins (*Tursiops truncatus*; Finneran and Schlundt 2011); however, equal-response-latency curves have been generated from reaction-time studies and been used as proxies for equal-loudness contours (Kastelein et al. 2011). Several functions applicable to the assessment of noise impact on marine mammals have also been developed similar to the A-weighting function with adjustments for the hearing sensitivity of different marine mammal groups. Other weighting functions exist for other species.

4.2.14.3 M-Weighting

The M-weighting function was developed to account for the auditory sensitivity of five functional hearing groups of marine mammals (Southall et al. 2007). Development of this function was restricted by data availability and is limited in its capacity to capture all complexities of marine mammal auditory responses (Tougaard and Beedholm 2019). The function deemphasizes the frequencies near the upper and lower limits of the auditory sensitivities of each hearing group, emphasizing frequencies where exposure to high-amplitude noise is more likely to affect the focal species (Houser et al. 2017). M-weighted SEL is calculated through energy integration over all frequencies following the application of the M-weighting function to the noise spectrum. The M-weighting functions have continued to

evolve, reflecting the advancement in marine mammal auditory sensitivity and response research, with the most recent modifications proposed by Southall et al. (2019), including a redefinition of marine mammal hearing groups, function assumptions, and parameters. The updated functions are based on the following equation:

$$W(f) = C + 10 \log_{10} \frac{\left(\frac{f}{f_1}\right)^{2a}}{\left(\left[1 + \left(\frac{f}{f_1}\right)^2\right]^a \left[1 + \left(\frac{f}{f_2}\right)^2\right]^b\right)} \quad (4.4)$$

$W(f)$ is the weighting function amplitude [dB] at frequency f [kHz]; f_1 and f_2 are the low-frequency and high-frequency cut-off values [kHz], respectively. Constants a and b are the low-frequency and high-frequency exponent values, defining the rate of decline of the weighting amplitude at low and high frequencies, and C defines the vertical position of the curve (maximum weighting function amplitude is 0). Table 4.8 lists the function constants for each marine mammal hearing group and Fig. 4.12 plots the weighting curves.

4.2.15 Frequency Bands

Different sound sources emit sound at different frequencies and cover different frequency bands. The whistle of a bird is quite tonal, covering a narrow band of frequencies. An echosounder

Table 4.8 Constants of Eq. 4.4 for the six functional hearing groups of marine mammals (Southall et al. 2019)

Marine mammal hearing group	a	b	f_1 [kHz]	f_2 [kHz]	C [dB]
Low-frequency cetaceans (LF)	1	2	0.2	19	0.13
High-frequency cetaceans (HF)	1.6	2	8.8	110	1.20
Very-high-frequency cetaceans (VHF)	1.8	2	12	140	1.36
Sirenians (SI)	1.8	2	4.3	25	2.62
Phocid carnivores in water (PCW)	1	2	1.9	30	0.75
Phocid carnivores in air (PCA)	2	2	0.75	8.3	1.50
Other marine carnivores in water (OCW)	2	2	0.94	25	0.64
Other marine carnivores in air (OCA)	1.4	2	2	20	1.39

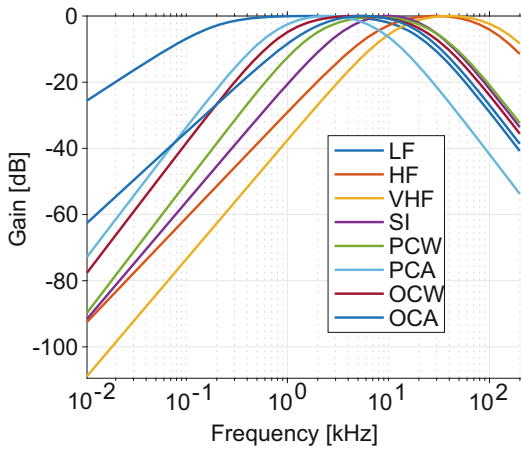


Fig. 4.12 Weighting curves calculated from the function $W(f)$ (Eq. 4.4) and constants (Table 4.8), for each marine mammal hearing group

emits a sharp tone, concentrating almost all acoustic energy in a narrow frequency band centered on one frequency. These are narrowband sources, while a ship propeller is a broadband source generating many octaves in frequency. The term frequency band refers to the band of frequencies of a sound. The bandwidth is the difference between the highest and the lowest frequency of a sound. The spectrum of a sound shows which frequencies are contained in the sound and the amplitude at each frequency.

Peak frequency and 3-dB bandwidth are often used to describe the spectral characteristics of a signal. Peak frequency is the frequency of maximum power of the spectrum. The 3-dB bandwidth is computed as the difference between the frequencies (on either side of the peak frequency), at which the spectrum has dropped 3 dB from its maximum (Fig. 4.13). Remember that a drop of 3 dB is equal to half power; and so the 3-dB bandwidth is the bandwidth at the half-power marks. Similarly, the 10-dB bandwidth is measured 10 dB down from the maximum power (i.e., where the power has dropped to one tenth of its peak).

For non-Gaussian spectra (e.g., bat or dolphin echolocation clicks), two other measures are useful: the center frequency f_c , which splits the power spectrum into two halves of equal

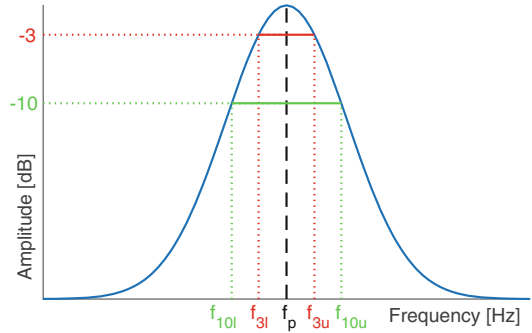


Fig. 4.13 Illustration of the 3-dB and 10-dB bandwidths of a signal; p: peak, l: lower, u: upper

power, and the rms bandwidth BW_{rms} , which measures the standard deviation about the center frequency. With $H(f)$ representing the Fourier transform, these quantities are computed as (Fig. 4.14):

$$f_c = \frac{\int_{-\infty}^{\infty} f |H(f)|^2 df}{\int_{-\infty}^{\infty} |H(f)|^2 df}$$

$$BW_{rms} = \sqrt{\frac{\int_{-\infty}^{\infty} (f - f_c)^2 |H(f)|^2 df}{\int_{-\infty}^{\infty} |H(f)|^2 df}}$$

Broadband sounds are commonly analyzed in specific frequency bands. In other words, the energy in a broadband sound can be split into a series of frequency bands. This splitting is done by a filter, which can be implemented in hardware or software. A low-pass filter lets low frequencies pass and reduces the amplitude of (i.e., attenuates) signals above its cut-off frequency. A high-pass filter lets high frequencies pass and reduces the amplitude of signals below its cut-off frequency. A band-pass filter passes signals within its characteristic pass-band (extending from a lower edge frequency to an upper edge frequency) and attenuates signals outside of this band. It is a common misconception that a filter removes all energy beyond its cut-off frequency. Instead, a filter progressively attenuates the

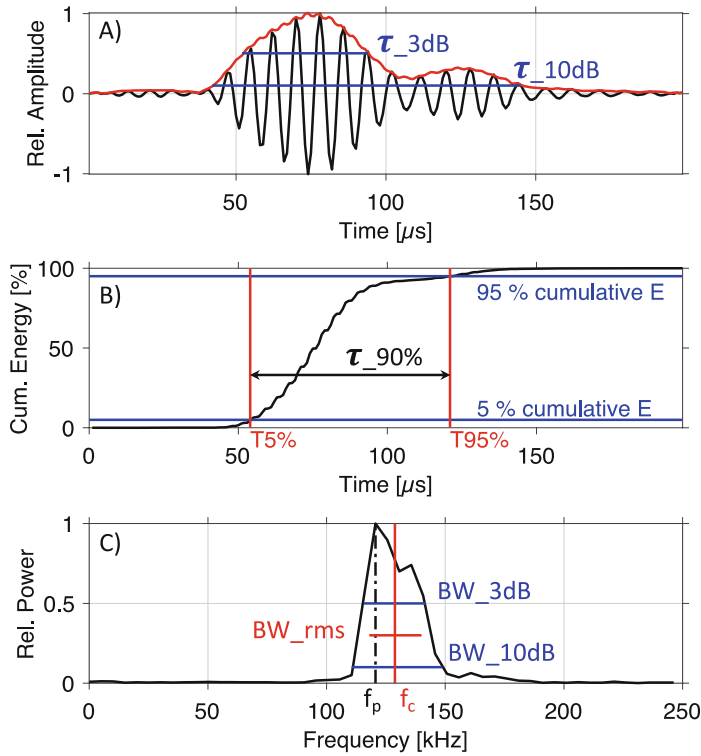


Fig. 4.14 Echolocation click from a harbor porpoise (*Phocoena phocoena*); (a) waveform and amplitude envelope (determined by Hilbert transform), (b) cumulative energy, and (c) spectrum. Three different duration parameters (τ) are shown. The 3-dB duration is the difference in time between the two points at half power (i.e., 3 dB down from the maximum of the signal envelope). The 10-dB duration is the time difference between the

points at one tenth of the peak power (i.e., 10 dB below the maximum). Computation of the 90% energy signal duration was explained in Sect. 4.2.6. Three bandwidth measures are shown. The 3-dB and 10-dB bandwidths are measured down from the maximum power, which occurs at the peak frequency f_p , and the rms bandwidth is measured about the center frequency f_c . Click recording courtesy of Whitlow Au

energy. At the cut-off frequency, the energy is typically reduced by 3 dB. Beyond the cut-off frequency, the attenuation increases; how rapidly depends on the order of the filter.

Band-pass filtering is very common in the study of broadband sounds, in particular broadband noise such as aircraft or ship noise. A number of band-pass filters are used that have adjacent pass-bands such that the sound spectrum is split into adjacent frequency bands. If these bands all have the same width, then the filters are said to have constant bandwidth. In contrast, proportional bandwidth filters split sound into adjacent bands that have a constant ratio of upper to lower frequency. These bands become wider with increasing frequency (e.g., octave bands).

Octave bands are exactly one octave wide, with an octave corresponding to a doubling of frequency. The upper edge frequency of an octave band is twice the lower edge frequency of the band: $f_{up} = 2 f_{low}$. Fractional octave bands are a fraction of an octave wide. One-third octave bands are common. The center frequencies f_c of adjacent 1/3 octave bands are calculated as $f_c(n) = 2^{n/3}$, where n counts the 1/3 octave bands. The lower and upper frequencies of band n are calculated as:

$$f_{low}(n) = 2^{-1/6} f_c(n) \text{ and } f_{up}(n) = 2^{1/6} f_c(n)$$

Another example for proportional bands are decadedades. Their center frequencies f_c are

Table 4.9 Center frequencies of adjacent 1/3 octave bands [Hz]. The table can be extended to lower and higher frequencies by division and multiplication by 10, respectively

10	12.5	16	20	25	31.5	40	50	63	80
100	125	160	200	250	315	400	500	630	800
1000	1250	1600	2000	2500	3150	4000	5000	6300	8000
10,000	12,500	16,000	20,000	25,000	31,500	40,000	50,000	63,000	80,000

calculated as $f_c(n) = 10^{n/10}$, where n counts the dec decades. The lower and upper frequencies of band n are calculated as:

$$f_{low}(n) = 10^{-1/20} f_c(n)$$

$$f_{up}(n) = 10^{1/20} f_c(n)$$

Dec decades are a little narrower than 1/3 octaves by about 0.08%. Dec decades are often erroneously called 1/3 octaves in the literature. Given this confusion and inconsistencies in rounding, preferred center frequencies have been published (Table 4.9).

4.2.16 Power Spectral Density

The spectral density of a power quantity is the average of that quantity within a specified frequency band, divided by the bandwidth of that band. Spectral densities are typically computed for mean-square sound pressure or sound exposure. Furthermore, spectral densities are most commonly computed in a series of adjacent constant-bandwidth bands, where each band is exactly 1 Hz wide. The spectral density then describes how the power quantity of a sound is distributed with frequency. The mean-square sound pressure spectral density level is expressed in dB:

$$L_{p,f} = 10 \log_{10} \left(\frac{\overline{p_f^2}}{p_{f_0}^2} \right)$$

The reference value $p_{f_0}^2$ is $1 \mu\text{Pa}^2/\text{Hz}$ in water. In air, it is more common to take the square root and report spectral density in dB re $20 \mu\text{Pa}/\sqrt{\text{Hz}}$.

4.2.17 Band Levels

Band levels are computed over a specified frequency band. Band levels can be computed from spectral densities by integrating over frequency before converting to dB.

Consider the sketched mean-square sound pressure spectral density as a function of frequency (Fig. 4.15). The band level L_p in the band from f_{low} to f_{up} is the total mean-square sound pressure in this band:

$$L_p = 10 \log_{10} \left(\frac{\int_{f_{low}}^{f_{up}} p_f^2 df}{p_{f_0}^2 f_0} \right)$$

$$= 10 \log_{10} \left(\frac{\overline{p_f^2} (f_{up} - f_{low})}{p_{f_0}^2 f_0} \right)$$

$$= 10 \log_{10} \left(\frac{\overline{p_f^2}}{p_{f_0}^2} \right)$$

$$+ 10 \log_{10} \left(\frac{f_{up} - f_{low}}{f_0} \right)$$

where the reference frequency f_0 is 1 Hz. The band level of mean-square sound pressure is thus equal to the level of the average mean-square sound pressure spectral density plus $10 \log_{10}$ of the bandwidth. The band level is expressed in dB re $1 \mu\text{Pa}^2$ in water. In the in-air literature, it is more common to take the square root and report band levels in dB re $20 \mu\text{Pa}$. The frequency band should always be reported as well.

The wider the bands, the higher the band levels, as illustrated for 1/12, 1/3, and 1 octave bands in Fig. 4.16.

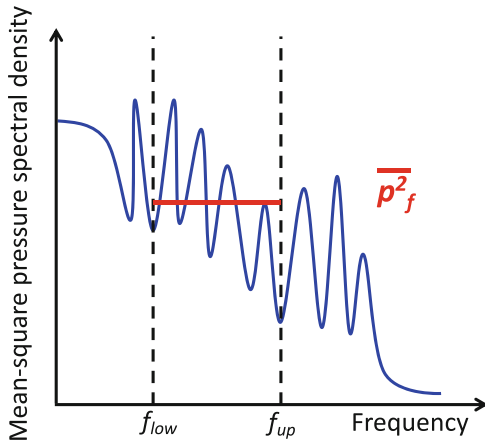


Fig. 4.15 Graph of mean-square pressure spectral density (blue) and its average $\overline{p_f^2}$ (red) in the frequency band from f_{low} to f_{up}

4.3 Acoustic Signal Processing

4.3.1 Displays of Sounds

A signal can be represented in the time domain and displayed as a waveform, or in the frequency domain and displayed as a spectrum. Waveform plots typically have time on the x-axis and amplitude on the y-axis. Waveform plots are useful for analysis of short pulses or clicks. Before the common use of desktop computers, acoustic waveforms were commonly displayed by oscilloscopes (or oscillographs). The display of the waveform was called an oscillogram. Power spectra are typically displayed with frequency on the x-axis and amplitude on the y-axis.

A few examples of waveforms and their spectra are shown in Fig. 4.17.² A constant-wave sinusoid (a) has a spectrum consisting of a single spike at the signal's fundamental frequency, in this case 1 kHz. The signal shown in (b) has the same fundamental frequency of 1 kHz, but its spectrum shows additional overtones at integer multiples of the fundamental that are due to its more complicated shape. A pulse (c) has a quite

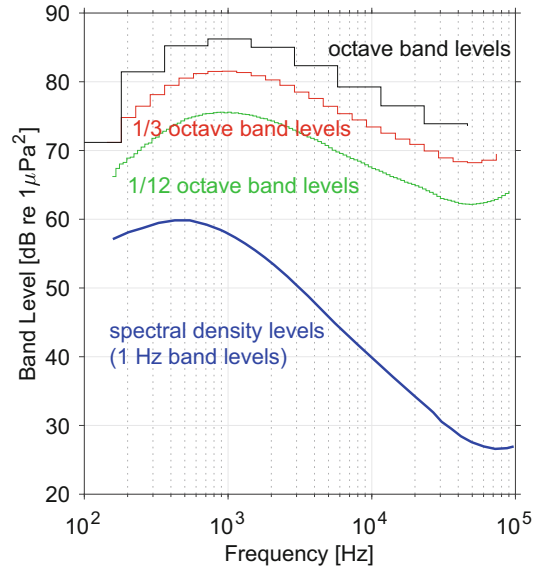


Fig. 4.16 Illustration of band levels versus spectral density levels, for the example of wind-driven noise under water at Sea State 2. Band levels are at least as high as the underlying spectral density levels. There are twelve 1/12-octave bands in each octave, and three 1/3-octave bands. The wider the band, the higher the level, because more power gets integrated

different spectrum to the previous repetitive signals, with a maximum at zero frequency and decaying in a series of ripples (known as sidelobes) that decrease in amplitude as frequency increases. It turns out that the shorter the pulse is, the wider is the initial spectral peak. Also, the faster the rise and fall times are, the more pronounced the sidelobes are and the slower they decay. Panel (d) shows the waveform and spectrum of a 1-kHz sinusoidal signal that has been amplitude-modulated by the pulse shown in (c). The effect of this is to shift the spectrum of the pulse so that what was at zero frequency is now at the fundamental frequency of the sinusoid, and to mirror it around that frequency. Another way of thinking about this is that the effect of truncating the sinusoid is to broaden its spectrum from the spike shown in (a). The effect of changing the frequency during the burst can be seen in (e). In this case, the frequency has been swept from 500 Hz to 1500 Hz over the 10-ms burst duration. This has the effect of broadening the spectrum and smoothing out the sidelobes that were

² Dan Russell's animations of the Fourier compositions of different waveforms: <https://www.acs.psu.edu/drussell/Demos/Fourier/Fourier.html>; accessed 12 October 2020.

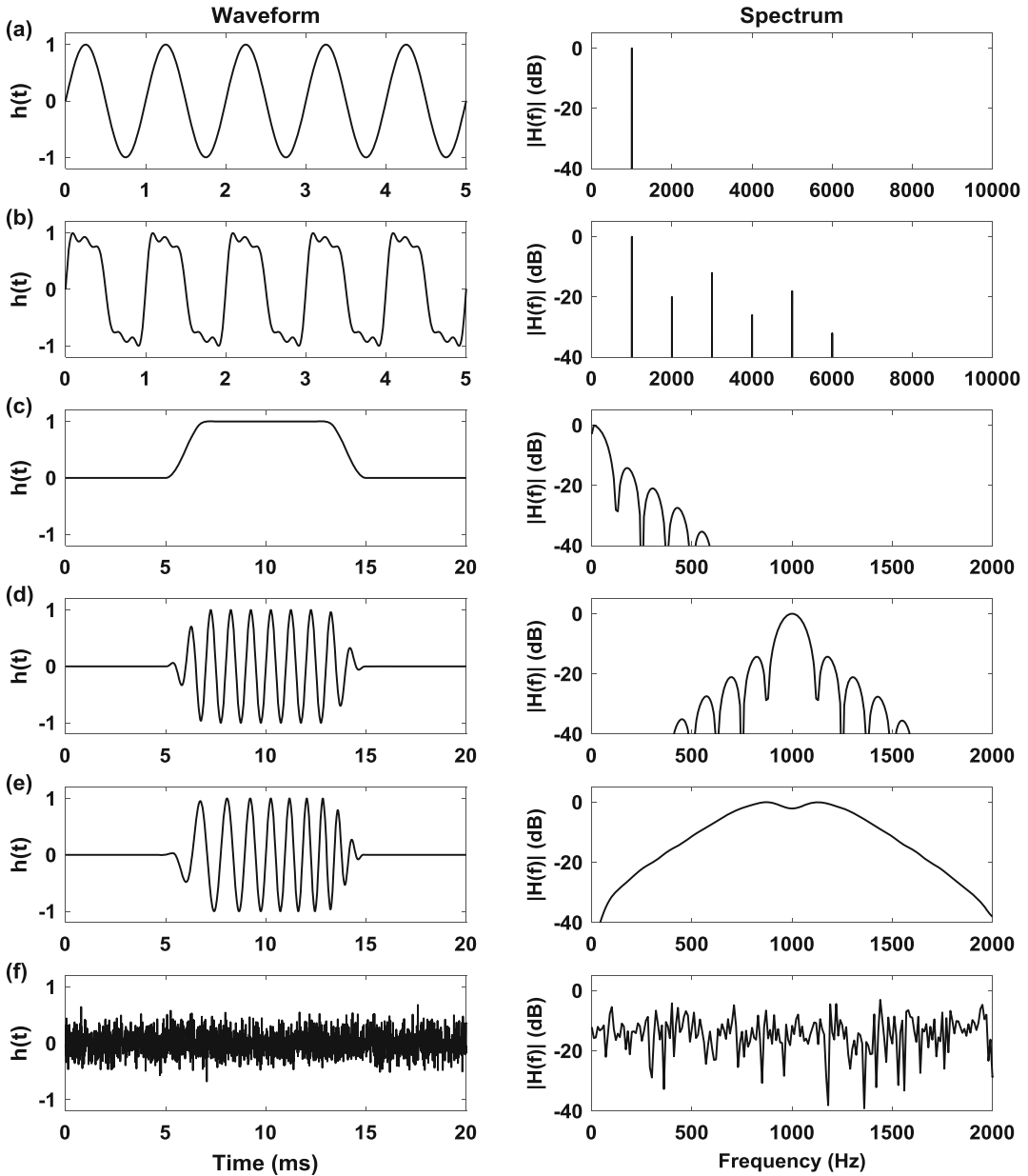


Fig. 4.17 Examples of signal waveforms (left) and their spectra (right). (a) A sine wave with a frequency of 1000 Hz; (b) a signal consisting of a sine wave with a fundamental frequency of 1000 Hz and five overtones; (c) a 10-ms long pulse with 2-ms rise and fall times; (d) a

10-ms long tone burst with a center frequency of 1000 Hz and 2-ms rise and fall times; (e) a 10-ms long FM sweep from 500 Hz to 1500 Hz with 2-ms rise and fall times; and (f) uncorrelated (white) random noise

apparent in (d). Finally, (f) shows a waveform consisting of uncorrelated noise and its spectrum. In this context “uncorrelated” means that knowledge of the noise at one time instant gives no

information about what it will be at any other time instant. This type of noise is often called white noise because it has a flat spectrum (like white light), but as can be seen in this example,

the spectrum of any particular white noise signal is itself quite noisy and it is only flat if one averages the spectra of many similar signals, or alternatively the spectra of many segments of the same signal.

A spectrogram is a plot with, most commonly, time on the x-axis and frequency on the y-axis. A quantity proportional to acoustic power is displayed by different colors or gray levels. If properly calibrated, a spectrogram will show mean-square sound pressure spectral density. A spectrogram is computed as a succession of Fourier transforms. A window is applied in the time domain containing a fixed number of samples of the digital time series. The Fourier transform is computed over these samples. Amplitudes are squared to yield power. The power spectrum is then plotted as a vertical column with frequency on the y-axis. The window in the time domain is then moved forward in time and the next samples of the digital time series are taken and Fourier-transformed. This second spectrum is then plotted next to the first spectrum, as the second vertical column in the spectrogram. The window in the time domain is moved again, the third Fourier transform is computed and plotted as the third column of the spectrogram, and so forth (see examples in Fig. 4.2). The spectrogram, therefore, shows how the spectrum of a sound changes over time. With modern signal processing software, researchers are able to listen to the sounds in real-time while viewing the spectral patterns.

4.3.2 Fourier Transform

It turns out that any signal can be broken down into a sum of sine waves with different amplitudes, frequencies, and phases. This is done by the Fourier transform, named after French mathematician and physicist Joseph Fourier. While the original signal can be represented as a time series $h(t)$ (e.g., sound pressure $p(t)$) in the time domain, the Fourier transform transforms the signal into the frequency domain, where it is represented as a spectrum $H(f)$. The magnitude of H is the amount of that

frequency in the original signal. $H(f)$ is a complex function and the argument contains the phase of that frequency. The inverse Fourier transform recreates the original signal from its Fourier components. For a continuous function with t representing time and f representing frequency, the Fourier transform is (i is the imaginary unit):

$$H(f) = \int_{-\infty}^{\infty} h(t)e^{-2\pi ift} dt$$

and the inverse Fourier transform is:

$$h(t) = \int_{-\infty}^{\infty} H(f)e^{2\pi ift} df$$

While a sound wave might be continuous, during digital recording or digitization of an analogue recording, its instantaneous pressure is sampled at equally spaced times over a finite window in time. This results in a finite and discrete time series. The equations for the discrete Fourier transform are similar to the above, where the integrals are replaced by summations. The fast Fourier transform (FFT) is the most common mathematical algorithm for computing the discrete Fourier transform. In animal bioacoustics, the FFT is the most commonly used algorithm to compute the frequency spectrum of a sound. The most common display of the frequency spectrum is as a power spectrum. Here, the amplitudes $H(f)$ are squared and in this process, the phase information is lost and, therefore, the original time series cannot be recreated. If sufficient care is taken to properly preserve the phase information, it is not only possible, but often very convenient, to transform a signal into the frequency domain using the FFT, carry out processing (such as filtering) in this domain, and then use an inverse FFT to resynthesize the processed signal in the time domain.

4.3.3 Recording and FFT Settings

Sounds in the various displays can look rather different depending on the recording and analysis parameters. There is no set of parameters that will produce the best display for all sounds. Rather,

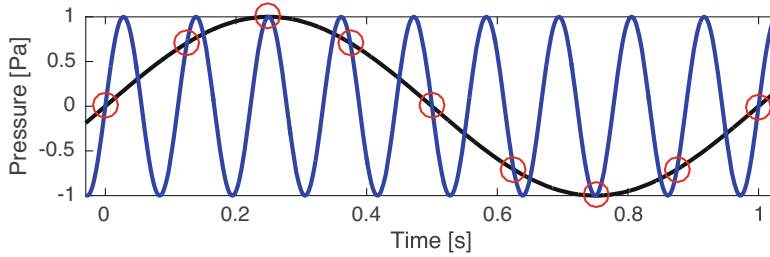


Fig. 4.18 Waveforms of a 1-Hz sine wave (black) and a 9-Hz sine wave (blue), both sampled 8 times per second (i.e., $f_s = 8$ Hz) as indicated by the red circles. Note that the

red samples fit either sine wave. In fact, there is an infinite number of signals that fit these samples

the ideal parameters depend on the question being asked, and it is important to have a thorough understanding of each of the parameters or selectable settings, and how they interact.

4.3.3.1 Sampling Rate

Microphones and hydrophones produce continuous voltages in response to sounds. The voltage outputs are termed analogue in that they are direct analogues of the acoustic signal. Analogue-to-digital converters sample the voltages of the signal and the level is expressed as a number (a digit) for each of the samples. The sampling rate is the number of samples per second and its unit is 1/s. The inverse is called the sampling frequency (symbol: f_s ; unit: Hz). Music on commercial CDs is digitized at 44.1 kHz (i.e., there are 44,100 samples stored every second). At high sampling rates, the digital sound file becomes very large for long-duration sound. The rate at which sounds are sampled by a digital recorder is typically stored in the header of the sound file. This file is a list of numbers with each number being the sound pressure at that sample point. Digital sound files are an incomplete record of the original signal; the intervals in the original signal between samples are lost during digitizing. The result is that there is a maximum frequency (related to the sampling rate) that can be resolved during Fourier analysis. Imagine a low-frequency sine wave. Only a few samples are needed to determine its frequency and amplitude and to recreate the full sine wave (by interpolation) from its samples. Those few samples might not be enough if the frequency is higher.

4.3.3.2 Aliasing

Aliasing is a phenomenon that occurs due to sampling. A continuous acoustic wave is digitally recorded by sampling at a sampling frequency f_s and storing the data as a time series $p(t)$. It turns out that different signals can produce the identical time series $p(t)$ and are therefore called aliases of each other. In Fig. 4.18, $p_{black}(t)$ has a frequency $f_{black} = 1$ Hz, while $p_{blue}(t)$ has a frequency $f_{blue} = 9$ Hz. A recorder that samples at $f_s = 8$ Hz would measure the pressure as indicated by the red circles from either the red or the blue time series. Based on the samples only, it is impossible to tell which was the original time series. In fact, there is an infinite number of signals that fit these samples. If f_0 is the lowest frequency that fits these samples, then the frequency of the n^{th} alias is $f_a(n)$, with n being an integer number:

$$\frac{f_a(n)}{f_s} = \frac{f_0}{f_s} + n$$

The most common problem of aliasing in animal bioacoustics occurs if a high-frequency animal sound is recorded at too low a sampling frequency. After FFT, the spectrum or spectrogram displays a sound at an erroneously low frequency. The Nyquist frequency (named after Harry Nyquist, a Swedish-born electronic engineer) is the maximum frequency that can be determined and is equal to half the sampling frequency. This requires some *a priori* information of the sounds to be recorded before a recording system is put together. The

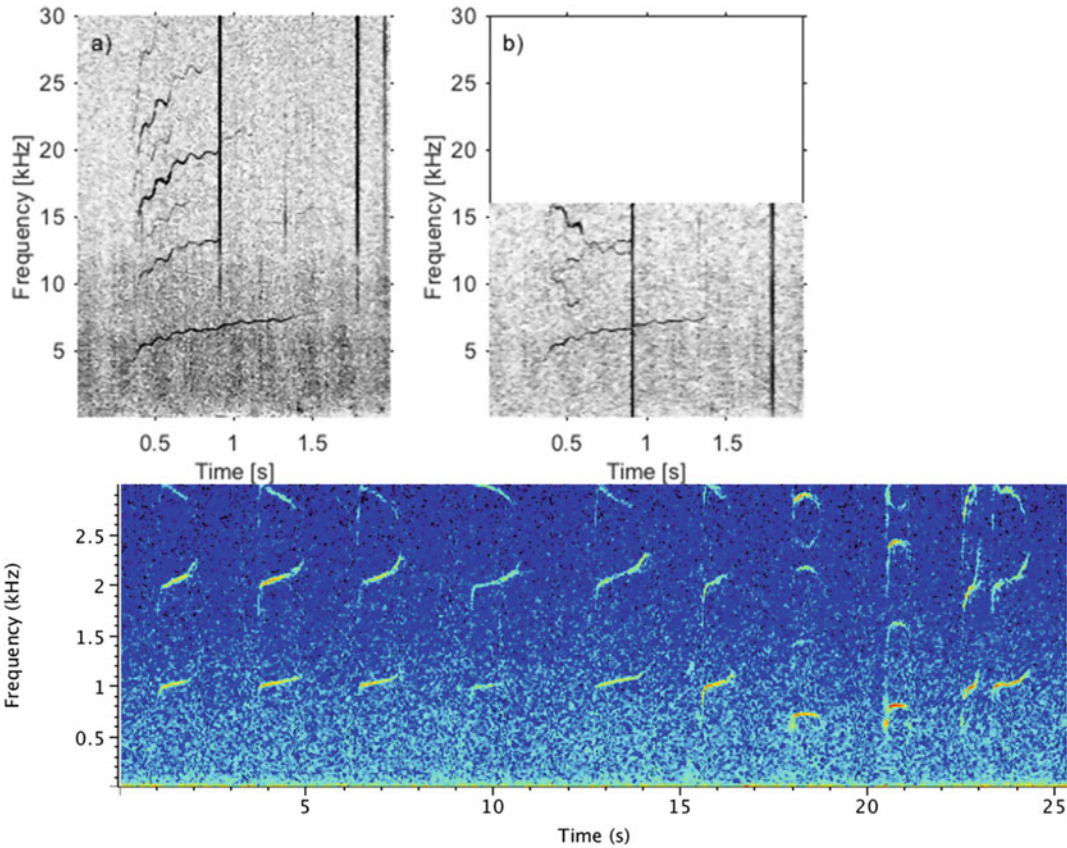


Fig. 4.19 Examples of folding (aliasing). Top: A killer whale sound sampled at 96 kHz (a) and at 32 kHz (b) (Wellard et al. 2015). If no anti-aliasing filter is applied, frequencies above the Nyquist frequency (i.e., 16 kHz in the right panel) will appear reflected downwards;

upsweeps greater than the Nyquist frequency appear as downsweeps. Bottom: Humpback whale (*Megaptera novaeangliae*) notes recorded with a sampling frequency of 6 kHz, but without an anti-aliasing filter. Contours above 3 kHz appear mirrored about the 3-kHz edge

higher the sampling frequency is, the higher the maximum frequency that can be accurately digitized.

In praxis, in order to avoid higher frequencies of animal sounds being erroneously displayed and interpreted as lower frequencies, an anti-aliasing filter is employed in the recording system. This is a low-pass filter with a cut-off frequency below the Nyquist frequency. Frequencies higher than the Nyquist frequency are thus attenuated, so that the effect of aliasing is diminished.

An example of aliasing is given in Fig. 4.19. Spectrograms of the same killer whale (*Orcinus orca*) call are shown sampled at 96 kHz and at

32 kHz. Without an anti-aliasing filter, energy is mirror-inverted or reflected about the Nyquist frequency of 16 kHz in the second case. Conceptually, energy is folded down about the Nyquist frequency by as much as it was above the Nyquist frequency.

4.3.3.3 Bit Depth

When a digitizer samples a sound wave (or the voltage at the end of a microphone), it stores the pressure measures with a limited accuracy. Bit depth is the number of bits of information in each sample. The more bits, the greater the resolution of that measure (i.e., the more accurate the pressure measure). Inexpensive sound digitizers

use 12 bits per sample. Commercially available CDs store each sample with 16 bits of storage, which allows greater accuracy in records of pressure. Blue-ray discs typically use 24 bits per sample. The more bits per sample, the larger the sound file to be stored, but the larger the dynamic range (ratio of loudest to quietest) of sounds that can be captured.

4.3.3.4 Audio Coding

Audio coding is used to compress large audio files to reduce storage needs. A common format is MP3, which can achieve 75–95% file reduction compared to the original time series stored on a CD or computer hard drive. Most audio coding algorithms aim to reduce the file size while retaining reasonable quality for human listeners. The MP3 compression algorithm is based on perceptual coding, optimized for human perception, ignoring features of sound that are beyond normal human auditory capabilities. Playing MP3 files back to animals might result in quite different perception compared to the playback of the original time series. Unfortunately, this is very often ignored in animal bioacoustic experiments. Lossless compression does exist (e.g., Free Lossless Audio Codec, FLAC; see Chap. 2 on recording equipment). For animal bioacoustics research, it is best to use lossless compression or none at all.

4.3.3.5 FFT Window Size (NFFT)

During Fourier analysis of a digitized sound recording, a fixed number of samples of the original time series is read and the FFT is computed on this window of samples. The number of samples is a parameter passed to the FFT algorithm and is typically represented by the variable NFFT. If NFFT samples are read from the original time series, then the Fourier transform will produce amplitude and phase measures at NFFT frequencies. However, the FFT algorithm produces a two-sided spectrum that is symmetrical about 0 Hz and contains NFFT/2 positive frequencies and NFFT/2–1 negative frequencies. To compute the power spectrum, after FFT, the amplitudes of all frequencies (positive and negative) are squared and summed. In the usual case of

a time series consisting of real (i.e., not complex) numbers, the same result is obtained by doubling the squared amplitudes of the positive frequencies and discarding the negative frequencies. This means that NFFT samples in the time domain yield NFFT/2 measures in the frequency domain. The FFT values, and therefore the power spectrum calculated from them, are output at a frequency spacing:

$$\Delta f = \frac{f_s}{\text{NFFT}}$$

For example, if a sound recording was sampled at 44.1 kHz and the FFT was computed over NFFT = 1024 samples, then the frequency spacing would be 43.07 Hz and the power spectrum would contain 512 frequencies: 43.07 Hz, 86.14 Hz, . . . , 22,050 Hz. A different way of looking at this is that the FFT produces spectrum levels in frequency bands of constant bandwidth. And the center frequencies in this example are 43.07 Hz, 86.14 Hz, . . . , 22,050 Hz. If there were two tones at 30 Hz and 50 Hz, then the combination of recording settings ($f_s = 44.1$ kHz) and analysis settings (NFFT = 1024) would be unable to separate these tones. Their power would be added and reported as the single level in the frequency band centered on 43.07 Hz. To separate these two tones, a frequency spacing of no more than 20 Hz is required. This is achieved by increasing NFFT. To yield a 1-Hz frequency spacing, 1 s of recording needs to be read into the FFT; i.e., NFFT = $f_s \times 1$ s.

As the NFFT increases, the frequency spacing decreases, but at the cost of the temporal resolution. This is because an increase in NFFT means that more samples from the original time series are read in order to compute one spectrum. More samples implies that the time window over which the spectrum is computed increases. In the above example, with $f_s = 44.1$ kHz, NFFT = 1024 samples correspond to a time window Δt of 0.023 s:

$$\Delta t = \frac{\text{NFFT}}{f_s} = \frac{1}{\Delta f}$$

While 44,100 samples last 1 s, 1024 samples only last 0.023 s. The spectrum is computed over

a time window of 0.023 s length. If the recording contained dolphin clicks of 100 μs duration, then the spectrum would be averaging over multiple clicks and ambient noise. To compute the spectrum of one click, a time window of 100 μs is desired and corresponds to $\text{NFFT} = f_s \times 100 \mu\text{s} = 4$. This is a very short window. The resulting frequency spacing would be impractically coarse:

$$\Delta f = \frac{f_s}{\text{NFFT}} = \frac{44,100 \text{ Hz}}{4} = 11,025 \text{ Hz}$$

There is a trade-off between frequency spacing and time resolution in Fourier spectrum analysis. This is often referred to as the Uncertainty Principle (e.g., Beecher 1988): $\Delta f \times \Delta t = 1$. In spectrograms, using a large NFFT will result in sounds looking stretched out in time, while a small NFFT will result in sounds looking smudged in frequency. The combination of recording settings (f_s) and analysis settings

(NFFT) should be optimized for the sounds of interest.

4.3.3.6 FFT Window Function

The computation of a discrete Fourier transform over a finite window of samples produces spectral leakage, where some power appears at frequencies (called sidelobes) that are not part of the original time series but rather due to the length and shape of the window. If a window of samples is read off the time series and passed straight into the FFT, then the window is said to have rectangular shape. The rectangular window function has values of 1 over the length of the window and values of 0 outside (i.e., before and after). The window function is multiplied sample by sample with the original time series so that NFFT values of unaltered amplitude are passed to the FFT algorithm. A rectangular window produces a large number of sidelobes (Fig. 4.20).

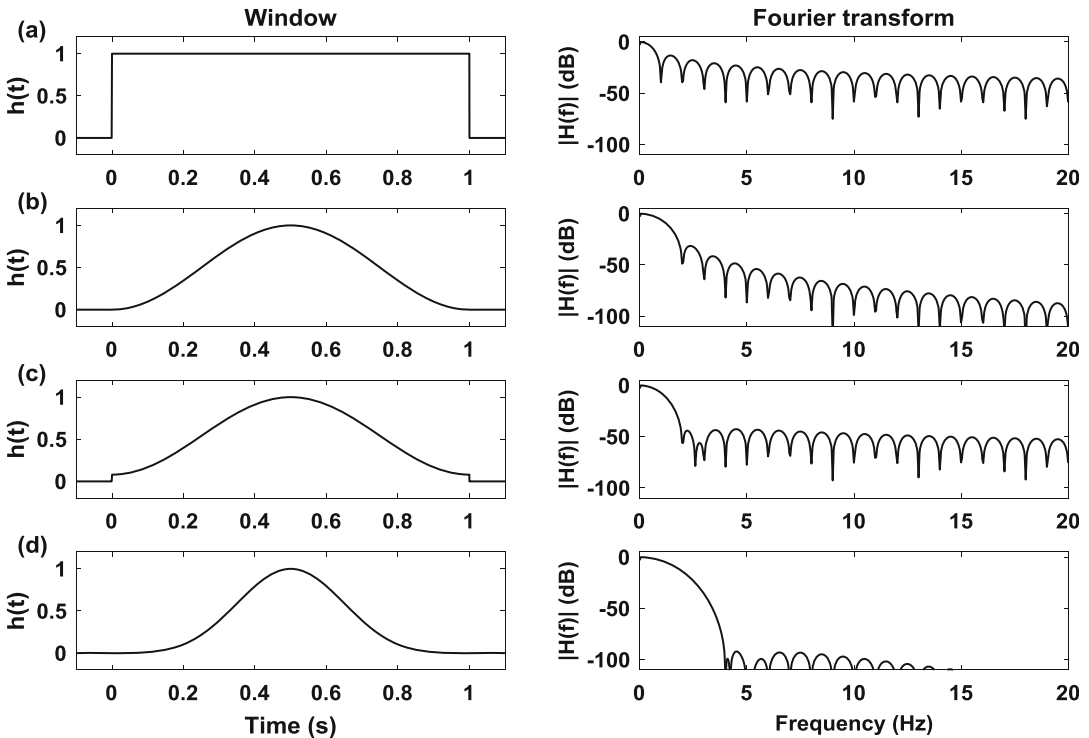


Fig. 4.20 Comparison of some window functions (left) and their Fourier transforms (right) for (a) rectangular, (b) Hann, (c) Hamming, and (d) Blackman-Harris windows

Spectral leakage can be reduced by using non-rectangular windows such as Hann, Hamming, or Blackman-Harris windows. These have values of 1 in the center of the window, but then taper off toward the edges to values of 0. The amplitude of the original time series is thus weighted. The benefits are fewer and weaker sidelobes, which result in less spectral leakage.

The smallest difference in frequency between two tones that can be separated in the spectrum is called the frequency resolution and is determined by the width of the main lobe of the window function. There is therefore a trade-off between the reduction in sidelobes and a wider main lobe, which results in poorer frequency resolution.

In order to not miss a strong signal or strong amplitude at the edges of the window where the amplitude is weighted by values close to 0, overlapping windows are used. Rather than reading samples in adjacent windows, windows commonly have 50% overlap. A spectrogram that was computed with 50% overlapping windows will have twice the number of spectrum columns and appear to have finer time resolution. Each spectrum column still has the same Δt as for a spectrogram without overlapping windows, but there will be twice as many spectrum columns making the spectrogram appear finer in time.

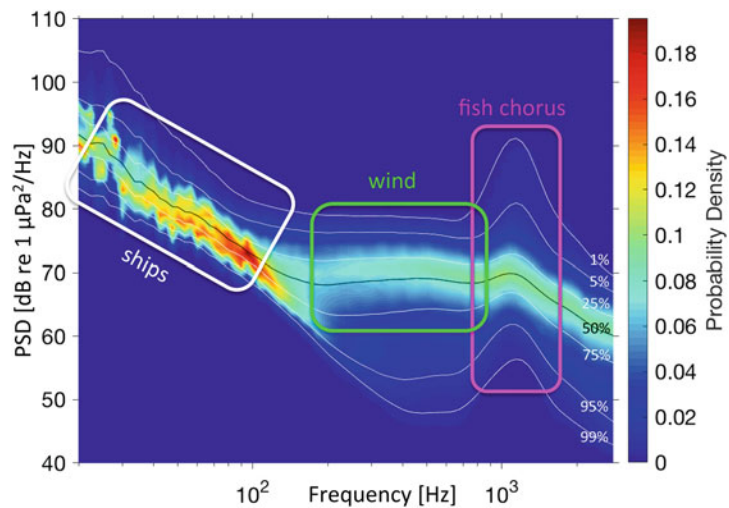
Zeros can be appended to each signal block (after windowing) to increase NFFT and therefore

reduce the frequency spacing Δf . This so-called zero-padding produces a smoother spectrum but does not improve the frequency resolution, which is still determined by the shape of the window and the duration of the signal to which the window was applied.

4.3.4 Power Spectral Density Percentiles and Probability Density

When recording soundscapes on land or under water, sounds fade in and out, from a diversity of sources and locations. A soundscape is dynamic, changing on short to long time scales (see Chap. 7). The variability in sound levels can be expressed as power spectral density (PSD) percentiles. The n^{th} percentile gives the level that is exceeded $n\%$ of the time (note: in engineering, the definition is commonly reversed). The 50th percentile corresponds to the median level. An example from the ocean off southern Australia is shown in Fig. 4.21. The median ambient noise level is represented by the thin black line and goes from about 90 dB re $1 \mu\text{Pa}^2/\text{Hz}$ at 20 Hz to 60 dB re $1 \mu\text{Pa}^2/\text{Hz}$ at 30 kHz. The lowest thin gray line corresponds to the 99th percentile. It gets quieter than this only 1% of the time. Levels at low frequencies (20–50 Hz) never drop below

Fig. 4.21 Percentiles of ambient noise power spectral densities measured off southern Australia over a year. Lines from top to bottom correspond to the following percentiles: 1, 5, 25, 50 (black), 75, 95, and 99



75 dB re 1 $\mu\text{Pa}^2/\text{Hz}$ because of the persistent noise from distant shipping.

These plots not only give the statistical level distribution over time, but can also identify the dominant sources in a soundscape based on the shapes of the percentile curves. The hump from 100 Hz to lower frequencies is characteristic of distant shipping. The more leveled curves at mid-frequencies (200–800 Hz) are characteristic of wind noise recorded under water. The median level of about 68 dB re 1 $\mu\text{Pa}^2/\text{Hz}$ corresponds to a Sea State of 4. The hump at 1.2 kHz is characteristic of chorusing fishes. While there are likely other sounds in this soundscape at certain times (e.g., nearby boats or marine mammals), they do not occur often enough or at a high enough level, to stand out in PSD percentile plots.

Probability density of PSD identifies the most common levels. In Fig. 4.21, at 100 Hz, the most common (probable) level was 75 dB re 1 $\mu\text{Pa}^2/\text{Hz}$. This was equal to the median level at this frequency. The red colors indicate that the median levels were also the most probable levels. At mid-to-high frequencies, the levels were more evenly distributed (i.e., only shades of blue and no red colors). The most probable levels are not necessarily equal to the median levels. A case where the most probable level (again from distant shipping) was below the median (due to strong pygmy blue whale, *Balaenoptera musculus breviceauda*, calling) is shown in Fig. 4.6, and a case where two different levels were equally likely (due to two seismic surveys at different ranges) is shown in Fig. 4.8, both of Erbe et al. 2016a.³ PSD percentile and probability density plots (as well as other graphs) can be created for both terrestrial and aquatic environments with the freely available software suite by Merchant et al. 2015.

4.4 Localization and Tracking

There are a few simple ways to gain information about the rough location and movement of a

sound source. By listening in air with two ears, we can tell the direction to the sound source and whether it remains at a fixed location or approaches or departs. From recordings made over a period of time, the closest point of approach (CPA) is often taken as the point in time when mean-square pressure (or some other acoustic quantity like particle displacement, velocity, or acceleration) peaked (Fig. 4.22).

Whether a sound source is approaching or departing can also be told from the Doppler shift. As a car or a fire engine drives past and as an airplane flies overhead, the pitch drops. In fact, as each approaches, the frequency received by a listener or a recorder is higher than the emitted frequency, and as each departs, the received frequency is lower than the emitted frequency.⁴ At CPA, the received frequency equals the emitted frequency. The time of CPA can be identified in spectrograms as the point in time when the steepest slope in the decreasing frequency occurred as the sound source passed or as the point in time when the frequency had decreased half-way (Fig. 4.23). The Doppler shift Δf can easily be quantified as

$$\Delta f = \frac{v}{c} f_0$$

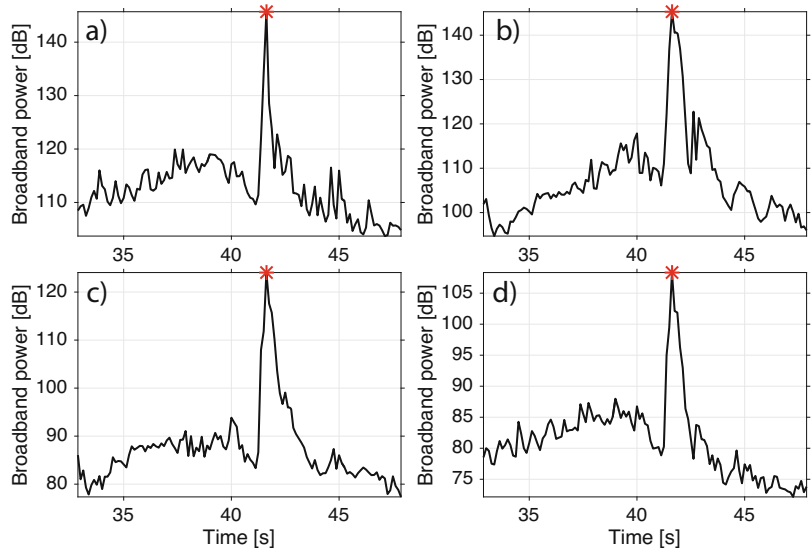
where v is the speed of the source relative to a fixed receiver, c is the speed of sound, and f_0 is the frequency emitted by the source (i.e., half-way between the approaching and the departing frequencies). From a spectrogram, not only the CPA, but also the speed of the sound source can be determined.

In the example of Fig. 4.23, one of the engine harmonics dropped from 96 Hz to 64 Hz. So the emitted frequency was 80 Hz and the Doppler shift was 16 Hz. With a speed of sound in air of 343 m/s, the airplane flew at 70 m/s = 250 km/h. The interesting part of this example is that the recorder was actually resting on the riverbed, in 1 m of water, and hence in a different acoustic medium to the source. How this affects the results

³ https://www.acoustics.asn.au/conference_proceedings/AASNZ2016/papers/p14.pdf; accessed 13 October 2020.

⁴ Doppler shift animations by Dan Russell: <https://www.acs.psu.edu/drussell/Demos/doppler/doppler.html>; accessed 13 October 2020.

Fig. 4.22 Graphs of (a) square pressure [dB re $1 \mu\text{Pa}^2$], (b) square particle displacement [dB re 1pm^2], (c) square particle velocity [dB re $1 (\text{nm/s})^2$], and (d) square particle acceleration [dB re $1 (\mu\text{m/s}^2)^2$] as a swimmer swims over a hydrophone. The closest point of approach is identified as the time of peak levels (i.e., at 42 s) (Erbe et al. 2017a)

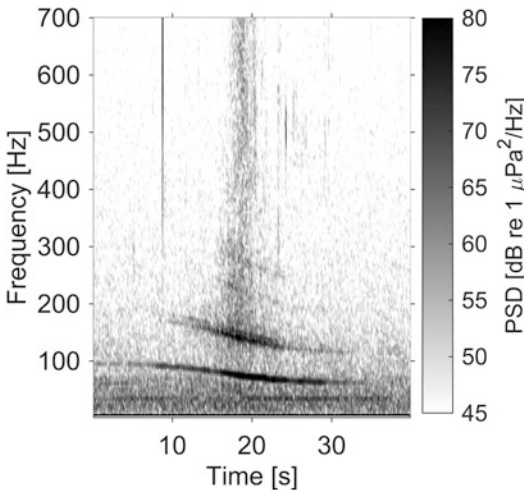


depends on the depth of the hydrophone relative to the acoustic wavelength. In this particular instance, the hydrophone was a small fraction of an acoustic wavelength below the water surface and the signal reached it via the evanescent wave (see Chap. 6 on sound propagation). The evanescent wave traveled horizontally at the in-air sound

speed, so it was the in-air sound speed that determined the Doppler shift. If the measurement had been carried out in deeper water with a deeper hydrophone, the signal would have been dominated by the air-to-water refracted wave, and the Doppler shift would have been determined by the in-water sound speed.

To accurately locate a sound source in space, signals from multiple simultaneous acoustic receivers need to be analyzed. These receivers are placed in specific configurations, known as arrays. Methods of localization are dependent on the configuration of the receiver array, the acoustic environment, spectral characteristics of the sound, and behavior of the sound source. There are three broad classes of these methods: time difference of arrival, beamforming, and parametric array processing methods. The following sections provide a condensed overview of the three methods. For a comprehensive treatise, please refer to the following: Schmidt 1986; Van Veen and Buckley 1988; Krim and Viberg 1996; Au and Hastings 2008; Zimmer 2011; Chiariotti et al. 2019.

Fig. 4.23 Spectrogram of an airplane flying over the Swan River, Perth, Australia, into Perth Airport. Recordings were made in the river, under water. The closest point of approach occurred at about 18 s, when the frequencies of the engine tone and its overtones dropped fastest (Erbe et al. 2018)



Tracking is a form of passive acoustic monitoring (PAM), where an estimation of the behavior of an active sound source is maintained over time. Passive acoustic tracking has many demonstrated applications in the underwater and terrestrial domains.

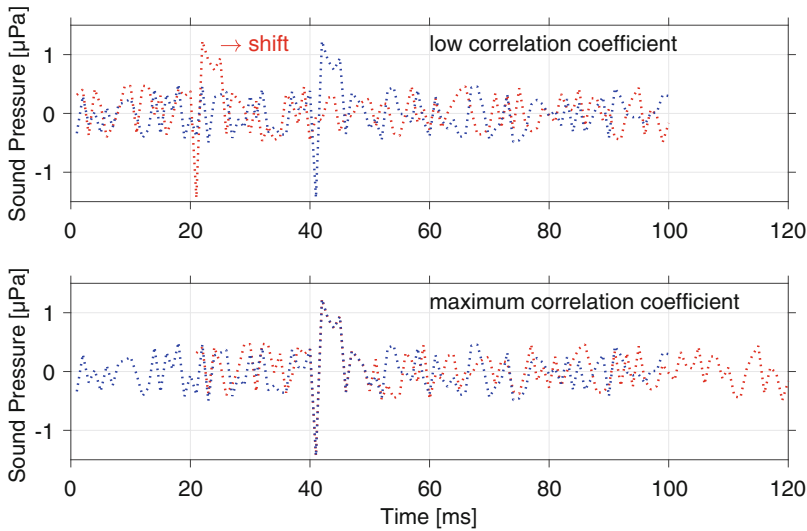


Fig. 4.24 Determining TDOA by cross-correlation. Top: Two 100-ms time series were recorded by two spatially separated receivers. A signal of interest arrived 20 ms into the recording at receiver 1 (red) and 40 ms into the recording at receiver 2 (blue). The dot product (i.e., correlation

coefficient) is low. Bottom: The red time series is shifted sample by sample against the blue time series and the dot product computed over the overlapping samples. When the signals line up, the correlation coefficient is maximum. In this example, the TDOA was 20 ms

4.4.1 Time Difference of Arrival

Localization by Time Difference Of Arrival (TDOA) is a two-step process. The first step is to measure the difference in time between the arrivals of the same sound at any pair of acoustic receivers. The second step is to apply appropriate geometrical calculations to locate the sound source. TDOA methods work best for signals that contain a wide range of frequencies (i.e., have a wide bandwidth), which includes short pulses, FM sweeps, and noise-like signals.

4.4.1.1 Generalized Cross-Correlation

TDOAs are commonly determined by cross-correlation. The time series of recorded sound pressure by two spatially separated receivers are cross-correlated as a sliding dot product. This means that each sample from receiver 1 is multiplied with a corresponding sample from receiver 2, and the products are summed over the full length of the overlapping time series. This yields the first cross-correlation coefficient. Next, the time series from receiver 1 (red in Fig. 4.24) is shifted by 1 sample against the time

series from receiver 2 (blue), and the dot product is computed again (over the overlapping samples), yielding the second cross-correlation coefficient. By sliding the two time series against each other (sample by sample) and computing the dot product, a time series of cross-correlation coefficients forms. A peak in cross-correlation occurs when the time series have been shifted such that the signal recorded by receiver 1 lines up with the signal recorded by receiver 2. The number of samples by which the time series were shifted, divided by the sampling frequency of the two receivers, is the TDOA.

Generalized cross-correlation is a common way of determining TDOA. It is suitable for localization in air and water in environments with high noise and reverberation and can be computed in either the time or frequency domains (Padois 2018).

4.4.1.2 TDOA Hyperbolas

TDOAs are always computed between two receivers (from a pair of receivers). Figure 4.25 sketches the arrangement of an animal A (at point A) and two receivers (R_1 and R_2) in space. The

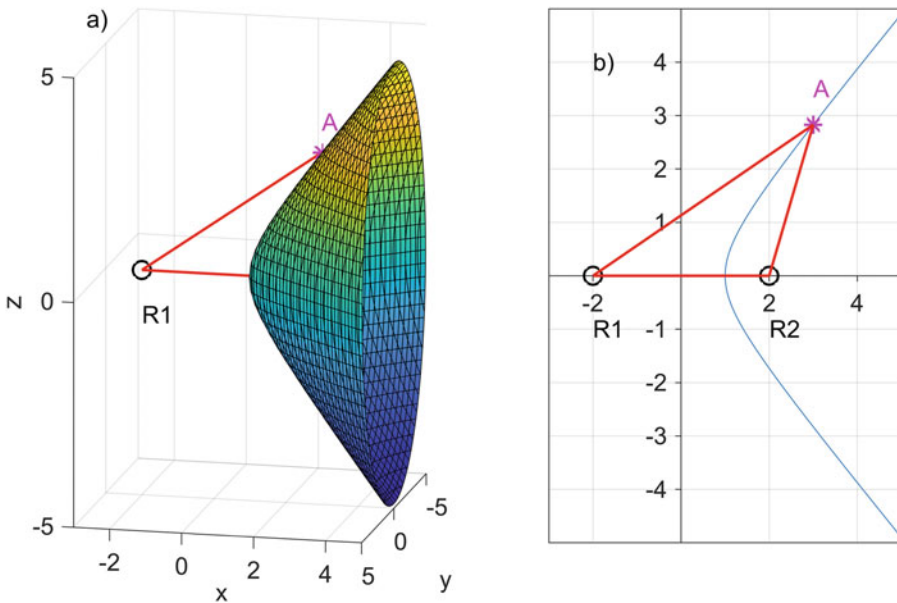


Fig. 4.25 Graphs of localization hyperbolas with two receivers; (a) 3D hyperboloid and (b) 2D hyperbola (i.e., cross-section) in the x-z plane. A marks the animal’s

position; R₁ and R₂ mark the receiver positions. R₂ is hidden inside the hyperboloid in the 3D image

distances $A-R_1$ (mathematically noted as a line connecting points A and R₁ and then taking the magnitude of it: $|\overline{AR_1}|$), $A-R_2$, and R_1-R_2 are shown as red lines. If A produces a sound that is recorded by both R₁ and R₂, then the arrival time at point R₁ is equal to the distance $A-R_1$, divided by the speed of sound c , and the arrival time at R₂ is equal to the distance $A-R_2$, divided by the speed of sound c . The TDOA is simply the difference between the two arrival times:

$$TDOA = \frac{|\overline{AR_1}| - |\overline{AR_2}|}{c}$$

It turns out mathematically that the animal can be anywhere on the hyperboloid and the TDOA will be the same. In other words, the TDOA defines a surface (in the shape of a hyperboloid) on which the animal may be located. With two receivers in the free-field, the animal’s position cannot be specified further. If there are boundaries near the animal and/or receivers (e.g., if a bird is tracked with receivers on the ground), then the possible location of the animal can be easily limited (i.e., the bird cannot fly underground, eliminating half of the space).

Reflections off boundaries can also be used to refine the location estimate. Finally, if one deploys more than two receivers, TDOAs can be computed between all possible pairs of receivers, yielding multiple hyperboloids that will intersect at the location of the animal.

4.4.1.3 TDOA Localization in 2 Dimensions

Localization in 2D space is, of course, simpler than in 3D, though it might seem a little contrived. In Fig. 4.26, the airport arrival flight path goes straight over a home. TDOA is used to locate (and perhaps track) each airplane. Two receivers on the ground will yield the upper half of the hyperbola in Fig. 4.25b as possible airplane locations. We know the airplane cannot be underground, but in terms of its altitude and range, two receivers are unable to resolve these. A third receiver in line with R₁ and R₂ is needed. With three receivers in a line array, three TDOAs can be computed and three hyperbolas can be drawn. Any two of these hyperbolas will intersect at two points: one above and one below the x-axis (i.e., above and below ground). Knowing that the

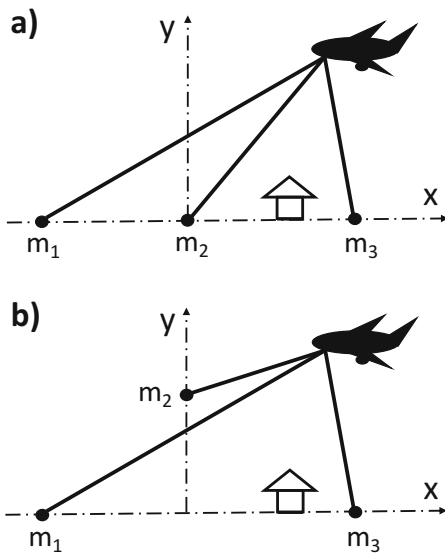


Fig. 4.26 Sketches of a three-microphone line array (a) and a triangular array (b)

airplane is above ground allows its position to be uniquely determined. If there were no boundary (i.e., ground in this case), an up-down ambiguity would remain; the plane could be at either of the two intersection points. Using more than three receivers in a line array (and thus adding more TDOAs and hyperbolas) will not improve the localization capability as all hyperbolas will intersect in the same two points: one above and one below the array. The up-down ambiguity can be resolved by using a 2D rather than 1D (i.e., line) arrangement. If one microphone is moved away from the line (as in Fig. 4.26b), the TDOA hyperbolas will intersect in just one point: the exact location of the airplane.

4.4.1.4 TDOA Localization in 3 Dimensions

The more common problem is to localize sound sources in 3-d space; i.e., when the sound source and the receivers are not in the same plane. Here, a line array of at least three receivers will result in hyperboloids that intersect in a circle. No matter how many receivers are in the line array, all TDOA hyperboloids will intersect in the same circle. There is up-down and left-right, in fact, circular ambiguity about the line of receivers.

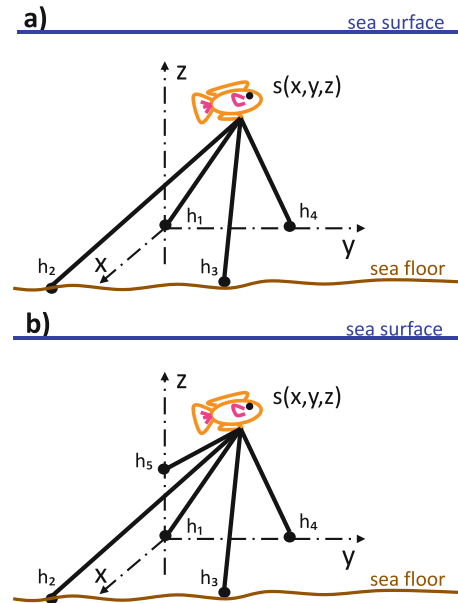


Fig. 4.27 Sketches of seafloor-mounted arrays with 4 (a) and 5 (b) hydrophones

This is a common situation with line arrays towed behind a ship in search of marine fauna.

In order to improve localization, a fourth receiver is needed that is not in line with the others. With four receivers, three hyperboloids can be computed that will intersect in two points: one above the plane of receivers and one below, yielding another up-down ambiguity. If the receiver sits on the ground or seafloor, then one of the points can be eliminated and the sound source uniquely localized. Otherwise, a fifth hydrophone is needed that is not in the same plane as the other four, allowing general localization in 3D space (Fig. 4.27).

The dimensions of an acoustic array used for TDOA localization are determined by the expected distance to the sound source and the likely uncertainty in the TDOA measurements, which is inversely proportional to the bandwidth of the sounds being correlated. A rough estimate of the TDOA uncertainty, δ_t (s), is $\delta_t \approx 1/BW$ where BW is the signal bandwidth (Hz). The corresponding uncertainty in the difference in distances from the two hydrophones to the source is then $\delta_d = c\delta_t$, where c is the sound speed (m/s).

When a sound source is far away from an array of receivers, the TDOAs can still be used to determine the direction of the sound source but any estimate of its distance will become inaccurate.

4.4.2 Beamforming

TDOA methods give poor results for sources that emit narrow-bandwidth signals such as continuous tones (e.g., some sub-species of blue whale) and can also be confounded in situations where there are many sources of similar signals in different directions from the array (e.g., a fish chorus). However, a properly designed array can be used to determine the direction of narrowband sources and can also determine the directional distribution of sound produced by multiple, simultaneously emitting sources using a processing method called beamforming. If two or more spatially separated arrays can be deployed, then the directional information they produce can be combined to obtain a spatial localization of the source. Alternatively, if the source is known to be stationary, or moving sufficiently slowly, localization can be achieved by moving a single array, for example by towing it behind a ship.

For the convenient, and hence commonly used case of an array consisting of a line of equally spaced hydrophones, beamforming requires the hydrophone spacing to be less than half the acoustic wavelength of the sound being emitted by the source. Also, the accuracy of the bearing estimates improves as the length of the array increases. These two factors combined mean that a useful array for beamforming is likely to require at least eight hydrophones, and even that would give only modest bearing accuracy. Consequently, 16-element or even 24-element arrays are commonly deployed in practice. A straight-line array used for beamforming suffers from the same ambiguity as a TDOA array in which all the hydrophones are in a straight line. As in the TDOA case, this ambiguity can be countered by offsetting some of the hydrophones from the straight line, however beamforming requires the

relative positions of all the hydrophones to be accurately known, so this is not always easy to achieve in practice.

Beamforming itself is relatively simple conceptually, but there are many subtleties (for details, see Van Veen and Buckley 1988; Krim and Viberg 1996). As for TDOA methods, the starting point is that when sound from a distant source arrives at an array of hydrophones, it will arrive at each hydrophone at a slightly different time, with the time differences depending on the direction of the sound source. The simplest type of beamformer is the delay and sum beamformer in which the array is “steered” in a particular direction by calculating the arrival time differences corresponding to that direction, delaying the received signals by amounts that cancel out those time differences, and then adding them together. This has the effect of reinforcing signals coming from the desired direction, while signals from other directions tend to cancel out. This isn’t a perfect process and the array will still give some output for signals coming from other directions. The relative sensitivity of the beamformer output to signals coming from different directions can be calculated and gives the beam pattern of the array. The beam pattern of a line array depends on the steering direction, with the narrowest beams occurring when the array is steered at right-angles to the axis of the array (broadside), and the broadest beams when steered in the axial direction (end-fire). There are a number of other beamforming algorithms that can give improved performance in particular circumstances; see the above references for details.

4.4.3 Parametric Array Processing

The array requirements for parametric array processing methods are similar to those for beamforming, but these methods attempt to circumvent the direct dependence of the angular accuracy on the length of the array (in acoustic wavelengths) that is inherent to beamforming. A summary of these methods can be found in Krim and Viberg (1996). One of the earliest and best

known parametric methods is the multiple signal classification (MUSIC) algorithm proposed by Schmidt 1986. These methods can give more accurate localization than beamforming in situations where there is a high signal-to-noise ratio and a limited number of sources, however they are significantly more complicated to implement and more time-consuming to compute. They also rely on more assumptions and are more sensitive to errors in hydrophone positions than beamforming.

4.4.4 Examples of Sound Localization in Air and Water

Passive acoustic localization in air poses logistical challenges with sound attenuating more rapidly in air than in water. This is an issue when localizing sound sources in open environments, as suitable recordings can only be collected if the microphone array is positioned closely around the source with localization error increasing with distance.

Sound source localization in the terrestrial domain is generally undertaken using one of three methods. Firstly, TDOA is perhaps most commonly applied to wildlife monitoring, including birds (McGregor et al. 1997) and bats (e.g., Surlykke et al. 2009; Koblitz 2018). Secondly, beamforming is more often utilized in environmental noise measurement and management (e.g., Huang et al. 2012; Prime et al. 2014; Amaral et al. 2018). Thirdly, the perhaps less common MUSIC approach has been utilized in bird monitoring and localization in noisy environments (Chen et al. 2006).

Under water, both fixed and towed hydrophone arrays are common. TDOA is the most common approach in the case of localizing cetaceans (Watkins and Schevill 1972; Janik et al. 2000) and fishes (Parsons et al. 2009; Putland et al. 2018). Under specific conditions, one or two hydrophones may suffice to localize a sound source by TDOA.

Multi-path propagation in shallow water may allow localization with just one hydrophone. TDOAs are computed between the surface-

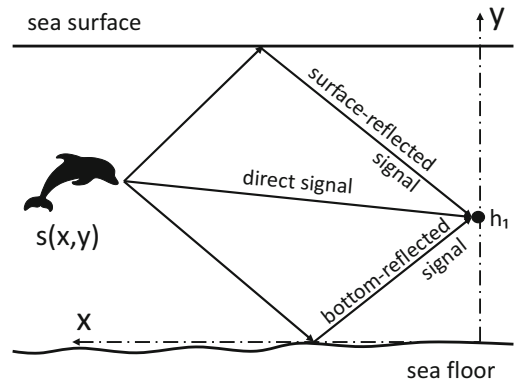


Fig. 4.28 Sketch of localization in shallow water using a single hydrophone (Cato 1998)

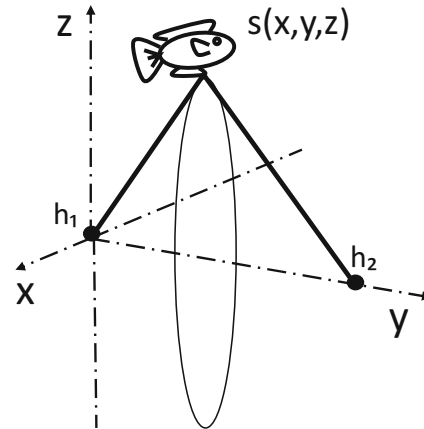


Fig. 4.29 Sketch of two hydrophones localizing a fish in 3D space with circular ambiguity using TDOA and intensity differences (Cato 1998)

reflected, seafloor-reflected, and direct sound propagation paths yielding both range and depth of the animal (Fig. 4.28), while not being able to resolve circular symmetry (Cato 1998; Mouy et al. 2012).

Using TDOAs in addition to differences in received intensity (when the source is located much closer to one of two receivers) may allow localization in free space to a circle between the two receivers and perpendicular to the line of two receivers (Cato 1998), see Fig. 4.29.

Beamforming is an established method for localizing soniferous marine animals (Miller and Tyack 1998) and anthropogenic sound sources

such as vessels (Zhu et al. 2018). A MUSIC approach to localization also has applications in the underwater domain, having previously been used for recovering acoustically-tagged artifacts by autonomous underwater vehicles (AUVs) (Vivek and Vadakkepat 2015).

Finally, target motion analysis involves marking the bearing to a sound source (from directional sensors or a narrow-aperture array) successively over time. If the animal calls frequently and moves slowly compared to the observation platform, successive bearings will intersect at the animal location (e.g., Norris et al. 2017).

4.4.5 Passive Acoustic Tracking

Passive acoustic tracking is the sequential localization of an acoustic source, useful for monitoring its behavior. Such behavior includes kinetic elements (e.g., swim path and speed) and acoustic elements (such as vocalization rate and type). In praxis, the process is a bit more complicated than just connecting TDOA locations over time. Animals will be arriving and departing; there

may be more than one animal vocalizing; any one animal will have quiet times between vocalizations. So, TDOA locations need to be joined into tracks; tracks need to be continued; old tracks need to be terminated; new tracks need to be initiated; tracks may need to be merged or split. Different algorithms have been developed to aid this process, with Kalman filtering being common (Zimmer 2011; Zarchan and Musoff 2013).

While radio telemetry has historically been the primary approach to terrestrial animal tracking, passive acoustic telemetry has grown in popularity as more animals can be monitored non-invasively (e.g., McGregor et al. 1997; Matsuo et al. 2014). Passive acoustic tracking in water is a well-established method of monitoring the behavior of aquatic fauna, including their responses to environmental and anthropogenic stimuli (e.g., Thode 2005; Stanistreet et al. 2013). Both towed and moored arrays are used, with towed arrays providing greater spatial coverage in the form of line-transect surveys.

4.5 Symbols and Abbreviations (Table 4.10)

Table 4.10 Most common quantities and abbreviations in this chapter

Quantity	Abbreviation	Symbol	Unit
Frequency		f	Hz
Sampling frequency		f_s	Hz
Wavelength		λ	m
Speed of sound		c	m/s
Particle velocity		u	m/s
Period of oscillation		τ	s
Time variable		t	s
Sound pressure		$p(t)$	Pa
Peak sound pressure		P_{pk}	Pa
Peak-to-peak sound pressure		P_{pk-pk}	Pa
Root-mean-square sound pressure		p_{rms}	Pa
Sound pressure level	SPL	L_p	dB re 1 μ Pa or 20 μ Pa
Peak sound pressure level	SPL _{pk}	$L_{p,pk}$	dB re 1 μ Pa or 20 μ Pa
Radiated noise level	RNL	L_{RN}	dB re 1 μ Pa m or 20 μ Pa m
Sound exposure level	SEL	$L_{E,p}$	dB re 1 μ Pa ² s or 400 μ Pa ² s
Source level	SL	L_S	dB re 1 or 20 μ Pa m
Number of Fourier components	NFFT		
Power spectral density level	PSD	$L_{p,f}$	dB re 1 μ Pa ² /Hz or 400 μ Pa ² /Hz
Time difference of arrival	TDOA		s

4.6 Summary

This chapter presented an introduction to acoustics and explained the basic quantities and concepts relevant to terrestrial and aquatic animal bioacoustics. Specific terminology that was introduced includes sound pressure, sound exposure, particle velocity, sound speed, longitudinal and transverse waves, frequency modulation, amplitude modulation, decibel, source level, near-field, far-field, frequency weighting, power spectral density, and one-third octave band level, amongst others. The chapter further introduced basic signal sampling and processing concepts such as sampling frequency, Nyquist frequency, aliasing, windowing, and Fourier transform. The chapter concluded with an introductory treatise of sound localization and tracking, including time difference of arrival and beamforming.

References

- Acton WI (1974) The effects of industrial airborne ultrasound on humans. *Ultrasonics* 12(3):124–128. [https://doi.org/10.1016/0041-624X\(74\)90069-9](https://doi.org/10.1016/0041-624X(74)90069-9)
- Amaral FR, Serrano Rico JC, Medeiros MAF (2018) Design of microphone phased arrays for acoustic beamforming. *J Braz Soc Mech Sci Eng* 40(7):354. <https://doi.org/10.1007/s40430-018-1275-5>
- American National Standards Institute (2013) *Acoustical Terminology (ANSI/ASA S1.1-2013)*. Acoustical Society of America, Melville, NY, USA
- Au WWL, Hastings M (2008) *Principles of marine bioacoustics*. Springer Verlag, New York
- Baumann-Pickering S, McDonald MA, Simonis AE, Solsona Berga A, Merckens KPB, Oleson EM, Roch MA, Wiggins SM, Rankin S, Yack TM, Hildebrand JA (2013) Species-specific beaked whale echolocation signals. *J Acoust Soc Am* 134(3):2293–2301. <https://doi.org/10.1121/1.4817832>
- Beecher MD (1988) Spectrographic analysis of animal vocalizations: Implications of the “Uncertainty Principle”. *Bioacoustics* 1(2–3):187–208. <https://doi.org/10.1080/09524622.1988.9753091>
- Caldwell MC, Caldwell DK (1965) Individualized whistle contours in bottlenosed dolphins (*Tursiops truncatus*). *Nature* 207(4995):434–435. <https://doi.org/10.1038/207434a0>
- Cato DH (1998) Simple methods of estimating source levels and locations of marine animal sounds. *J Acoust Soc Am* 104(3):1667–1678. <https://doi.org/10.1121/1.424379>
- Chen C-E, Ali AM, Wang H (2006) Design and testing of robust acoustic arrays for localization and enhancement of several bird sources. *Proceedings of the 5th International Conference on Information Processing in Sensor Networks:268–275*
- Chiariotti P, Martarelli M, Castellini P (2019) Acoustic beamforming for noise source localization – Reviews, methodology and applications. *Mech Syst Signal Process* 120:422–448. <https://doi.org/10.1016/j.ymssp.2018.09.019>
- Dziak RP, Bohnenstiehl DR, Matsumoto H, Fox CG, Smith DK, Tolstoy M, Lau T-K, Haxel JH, Fowler MJ (2004) P- and T-wave detection thresholds, Pn velocity estimate, and detection of lower mantle and core P-waves on ocean sound-channel hydrophones at the Mid-Atlantic Ridge. *Bull Seismol Soc Am* 94(2):665–677. <https://doi.org/10.1785/0120030156>
- Erbe C (2009) Underwater noise from pile driving in Moreton Bay, Qld. *Acoust Aust* 37(3):87–92
- Erbe C (2013) Underwater noise of small personal watercraft (jet skis). *J Acoust Soc Am* 133(4):EL326–EL330. <https://doi.org/10.1121/1.4795220>
- Erbe C, King AR (2009) Modelling cumulative sound exposure around marine seismic surveys. *J Acoust Soc Am* 125(4):2443–2451. <https://doi.org/10.1121/1.3089588>
- Erbe C, McCauley R, Gavrilov A, Madhusudhana S, Verma A (2016a). The underwater soundscape around Australia. *Proceedings of Acoustics 2016*, 9–11 November 2016, Brisbane, Australia.
- Erbe C, Parsons M, Duncan AJ, Allen K (2016b) Underwater acoustic signatures of recreational swimmers, divers, surfers and kayakers. *Acoust Aust* 44(2):333–341. <https://doi.org/10.1007/s40857-016-0062-7>
- Erbe C, Parsons M, Duncan AJ, Lucke K, Gavrilov A, Allen K (2017a) Underwater particle motion (acceleration, velocity and displacement) from recreational swimmers, divers, surfers and kayakers. *Acoust Aust* 45:293–299. <https://doi.org/10.1007/s40857-017-0107-6>
- Erbe C, Parsons M, Duncan AJ, Osterrieder S, Allen K (2017b) Aerial and underwater sound of unmanned aerial vehicles (UAV, drones). *J Unmanned Veh Syst* 5(3):92–101. <https://doi.org/10.1139/jjuvs-2016-0018>
- Erbe C, Williams R, Parsons M, Parsons SK, Hendrawan IG, Dewantama IMI (2018) Underwater noise from airplanes: An overlooked source of ocean noise. *Mar Pollut Bull* 137:656–661. <https://doi.org/10.1016/j.marpolbul.2018.10.064>
- Finneran J, Schlundt C (2011) Subjective loudness level measurements and equal loudness contours in a bottlenose dolphin (*Tursiops truncatus*). *J Acoust Soc Am* 130(5):3124–3136. <https://doi.org/10.1121/1.3641449>
- Fletcher H, Munson WA (1933) Loudness, its definition, measurement and calculation. *J Acoust Soc Am* 5(2):82–108. <https://doi.org/10.1121/1.1915637>
- Holland RA, Waters DA, Rayner JMV (2004) Echolocation signal structure in the Megachiropteran bat

- Rousettus aegyptiacus* Geoffroy 1810. *J Exp Biol* 207(25):4361. <https://doi.org/10.1242/jeb.01288>
- Houser DS, Yost W, Burkard R, Finneran JJ, Reichmuth C, Mulsow J (2017) A review of the history, development and application of auditory weighting functions in humans and marine mammals. *J Acoust Soc Am* 141(3):1371–1413. <https://doi.org/10.1121/1.4976086>
- Huang X, Bai L, Vinogradov I, Peers E (2012) Adaptive beamforming for array signal processing in aeroacoustic measurements. *J Acoust Soc Am* 131(3):2152–2161. <https://doi.org/10.1121/1.3682041>
- International Electrotechnical Commission (2013) *Electroacoustics - Sound level meters - Part 1: Specifications (IEC 61672-1 Ed. 2.0 b:2013)*. New York.
- International Organization for Standardization (2003) *Acoustics—Normal equal-loudness-level contours (ISO 226:2003)*.
- International Organization for Standardization (2007) *Acoustics — Definitions of basic quantities and terms (ISO/TR 25417)*. Geneva, Switzerland.
- International Organization for Standardization (2017) *Underwater acoustics—Terminology (ISO 18405)*. Geneva, Switzerland.
- Janik VM, Van Parijs SM, Thompson PM (2000) A two-dimensional acoustic localization system for marine mammals (Note). *Mar Mamm Sci* 16(2):437–447. <https://doi.org/10.1111/j.1748-7692.2000.tb00935.x>
- Kamminga C, Beitsma GR (1990) Investigations on cetacean sonar IX, remarks on dominant sonar frequencies from *Tursiops truncatus*. *Aquat Mamm* 16(1):14–20
- Kastelein RA, Wensveen PJ, Terhune JM, de Jong CAF (2011) Near-threshold equal-loudness contours for harbour seals (*Phoca vitulina*) derived from reaction times during underwater audiometry: a preliminary study. *J Acoust Soc Am* 129(1):488–495. <https://doi.org/10.1121/1.3518779>
- Koblitz JC (2018) Arrayvolution: using microphone arrays to study bats in the field. *Can J Zool* 96(9):933–938. <https://doi.org/10.1139/cjz-2017-0187>
- Krim H, Viberg M (1996) Two decades of array signal processing research: the parametric approach. *IEEE Signal Process Mag* 13(4):67–94. <https://doi.org/10.1109/79.526899>
- Kuehne LM, Erbe C, Ashe E, Bogaard LT, Collins MS, Williams R (2020) Above and below: Military aircraft noise in air and under water at Whidbey Island, Washington. *J Mar Sci Eng* 8(11):923. <https://doi.org/10.3390/jmse8110923>
- Leighton TG (2018) *Ultrasound in air—Guidelines, applications, public exposures, and claims of attacks in Cuba and China*. *J Acoust Soc Am* 144(4):2473–2489. <https://doi.org/10.1121/1.5063351>
- Marley SA, Erbe C, Salgado Kent CP (2017) Underwater recordings of the whistles of bottlenose dolphins in Fremantle Inner Harbour, Western Australia. *Sci Data* 4(170):126. <https://doi.org/10.1038/sdata.2017.126>
- Matsuo I, Wheeler A, Kloeppe L, Gaudette J, Simmons JA (2014) Acoustic tracking of bats in clutter environments using microphone arrays. *J Acoust Soc Am* 135(4):2207–2207. <https://doi.org/10.1121/1.4877207>
- McGregor PK, Dabelsteen T, Clark CW, Bower JL, Holland J (1997) Accuracy of a passive acoustic location system: empirical studies in terrestrial habitats. *Ethol Ecol Evol* 9(3):269–286. <https://doi.org/10.1080/08927014.1997.9522887>
- Merchant ND, Frstrup KM, Johnson MP, Tyack PL, Witt MJ, Blondel P, Parks SE (2015) Measuring acoustic habitats. *Methods Ecol Evol* 6(3):257–265. <https://doi.org/10.1111/2041-210X.12330>
- Miller PJ, Tyack PL (1998) A small towed beamforming array to identify vocalizing resident killer whales (*Orcinus orca*) concurrent with focal behavioral observations. *Deep Sea Res Part II Top Stud Oceanogr* 45(7):1389–1405. [https://doi.org/10.1016/S0967-0645\(98\)00028-9](https://doi.org/10.1016/S0967-0645(98)00028-9)
- Mouy X, Hannay D, Zykov M, Martin B (2012) Tracking of Pacific walrus in the Chukchi Sea using a single hydrophone. *J Acoust Soc Am* 131(2):1349–1358. <https://doi.org/10.1121/1.3675008>
- Norris TF, Dunleavy KJ, Yack TM, Ferguson EL (2017) Estimation of minke whale abundance from an acoustic line transect survey of the Mariana Islands. *Mar Mamm Sci*. <https://doi.org/10.1111/mms.12397>
- Padois T (2018) Acoustic source localization based on the generalized cross-correlation and the generalized mean with few microphones. *J Acoust Soc Am* 143(5):EL393–EL398. <https://doi.org/10.1121/1.5039416>
- Parrack HO (1966) Effect of Air-Borne Ultrasound on Humans. *Int J Audiol* 5(3):294–308. <https://doi.org/10.3109/05384916609074198>
- Parsons MJ, McCauley RD, Mackie MC, Siwabessy P, Duncan AJ (2009) Localization of individual mulloway (*Argyrosomus japonicus*) within a spawning aggregation and their behaviour throughout a diel spawning period. *ICES J Mar Sci* 66(6):1007–1014. <https://doi.org/10.1093/icesjms/fsp016>
- Prime Z, Doolan C, Zajamsek B (2014) Beamforming array optimisation and phase averaged sound source mapping on a model wind turbine. *Inter-Noise and Noise-Con Congress and Conference Proceedings* 249(7):1078–1086
- Putland RL, Mackiewicz AG, Mensinger AF (2018) Localizing individual soniferous fish using passive acoustic monitoring. *Ecol Inform* 48:60–68. <https://doi.org/10.1016/j.ecoinf.2018.08.004>
- Robinson DW, Dadson RS (1956) A re-determination of the equal-loudness relations for pure tones. *Br J Appl Phys* 7(5):166–181. <https://doi.org/10.1088/0508-3443/7/5/302>
- Schmidt R (1986) Multiple emitter location and signal parameter estimation. *IEEE Trans Antennas Propag* 34(3):276–280. <https://doi.org/10.1109/TAP.1986.1143830>
- Southall BL, Bowles AE, Ellison WT, Finneran JJ, Gentry RL, Greene CRJ, Kastak D, Ketten DR, Miller JH, Nachtigall PE, Richardson WJ, Thomas JA, Tyack

- PL (2007) Marine mammal noise exposure criteria: Initial scientific recommendations. *Aquat Mamm* 33(4):411–521. <https://doi.org/10.1080/09524622.2008.9753846>
- Southall BL, Finneran JJ, Reichmuth C, Nachtigall PE, Ketten DR, Bowles AE, Ellison WT, Nowacek DP, Tyack PL (2019) Marine mammal noise exposure criteria: Updated scientific recommendations for residual hearing effects. *Aquat Mamm* 45(2):125–232. <https://doi.org/10.1578/AM.45.2.2019.125>
- Stanistreet JE, Risch D, Van Parijs SM (2013) Passive acoustic tracking of singing humpback whales (*Megaptera novaeangliae*) on a Northwest Atlantic feeding ground. *PLoS One* 8(4):e61263. <https://doi.org/10.1371/journal.pone.0061263>
- Surlykke A, Pedersen SB, Jakobsen L (2009) Echolocating bats emit a highly directional sonar sound beam in the field. *Proc R Soc B Biol Sci* 276(1658):853–860. <https://doi.org/10.1098/rspb.2008.1505>
- Suzuki Y, Takeshima H (2004) Equal-loudness-level contours for pure tones. *J Acoust Soc Am* 116(2): 918–933. <https://doi.org/10.1121/1.1763601>
- Taylor B, Thompson A (eds) (2008) *The International System of Units (SI)*. National Institute of Standards and Technology, Gaithersburg, MD
- Thode A (2005) Three-dimensional passive acoustic tracking of sperm whales (*Physeter macrocephalus*) in ray-refracting environments. *J Acoust Soc Am* 118(6):3575–3584. <https://doi.org/10.1121/1.2049068>
- Tonin R (2018) A review of wind turbine-generated infrasound: Source, measurement and effect on health. *Acoust Aust* 46(1):69–86. <https://doi.org/10.1007/s40857-017-0098-3>
- Tougaard J, Beedholm K (2019) Practical implementation of auditory time and frequency weighting in marine bioacoustics. *Appl Acoust* 145:137–143. <https://doi.org/10.1016/j.apacoust.2018.09.022>
- Van Veen BD, Buckley KM (1988) Beamforming: a versatile approach to spatial filtering. *IEEE ASSP Mag* 5(2):4–24. <https://doi.org/10.1109/53.665>
- Vivek R, Vadakkepat P (2015) Multiple signal classification (MUSIC) based underwater acoustic localization module (UALM) for AUV. Paper presented at the 2015 IEEE Underwater Technology (UT) Conference. <https://doi.org/10.1109/UT.2015.7108288>
- Ward R, Parnum I, Erbe C, Salgado-Kent CP (2016) Whistle characteristics of Indo-Pacific bottlenose dolphins (*Tursiops aduncus*) in the Fremantle Inner Harbour, Western Australia. *Acoust Aust* 44(1): 159–169. <https://doi.org/10.1007/s40857-015-0041-4>
- Watkins WA, Schevill WE (1972) Sound source location by arrival-times on a non-rigid three-dimensional hydrophone array. *Deep-Sea Res Oceanogr Abstr* 19(10):691–706. [https://doi.org/10.1016/0011-7471\(72\)90061-7](https://doi.org/10.1016/0011-7471(72)90061-7)
- Wellard R, Erbe C, Fouda L, Blewitt M (2015) Vocalisations of killer whales (*Orcinus orca*) in the Bremer Canyon, Western Australia. *PLoS One* 10(9): e0136535. <https://doi.org/10.1371/journal.pone.0136535>
- Zarchan P, Musoff H (2013) *Fundamentals of Kalman filtering: a practical approach*, 4th edn. American Institute of Aeronautics and Astronautics, Inc., Reston, VA
- Zhu C, Garcia H, Kaplan A, Schinault M, Handegard NO, Godø OR, Huang W, Ratilal P (2018) Detection, localization and classification of multiple mechanized ocean vessels over continental-shelf scale regions with passive ocean acoustic waveguide remote sensing. *Remote Sens* 10(11):1699
- Zimmer WMX (2011) *Passive acoustic monitoring of cetaceans*. Cambridge University Press, Cambridge

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

