# Analysing sound wave using fast Fourier transform

## I. SOUND WAVE

A sound wave consists of compression and rarefaction of air molecules, which travels through space with sound speed $c_s$. Fig. 1 depicts the propagation of a *sinusoidal* sound wave inside an infinitely long rectangular channel. The figure shows the snapshots of the air molecules (represented by the blue dots in the figure) at four different instantaneous time $t = 0\,\mathrm{s}$, $0.001\,\mathrm{s}$, $0.002\,\mathrm{s}$, and $0.003\,\mathrm{s}$. The regions of high density are called the compression regions (these look like fuzzy vertical blue bands in the figure) and the regions of low density are called the rarefaction regions. In the figure, these bands of compression and rarefaction regions travel along the positive $x$-direction with speed equals to the sound speed $c_s$. The wavelength of this sinusoidal sound wave is defined by the distance between two nearest compression bands, which is equal to $\lambda = 2\,\mathrm{m}$. As we can see from the figure, after time $t = 0.003\,\mathrm{s}$, the bands will have travelled a distance of $1\,\mathrm{m}$. Therefore the speed of sound in this example is equal to:

$$c_s = \frac{1\,\mathrm{m}}{0.003\,\mathrm{s}} \simeq 333\,\mathrm{ms}^{-1}. \tag{1}$$

In general, the speed of sound depends on various factors such as ambient pressure and temperature.

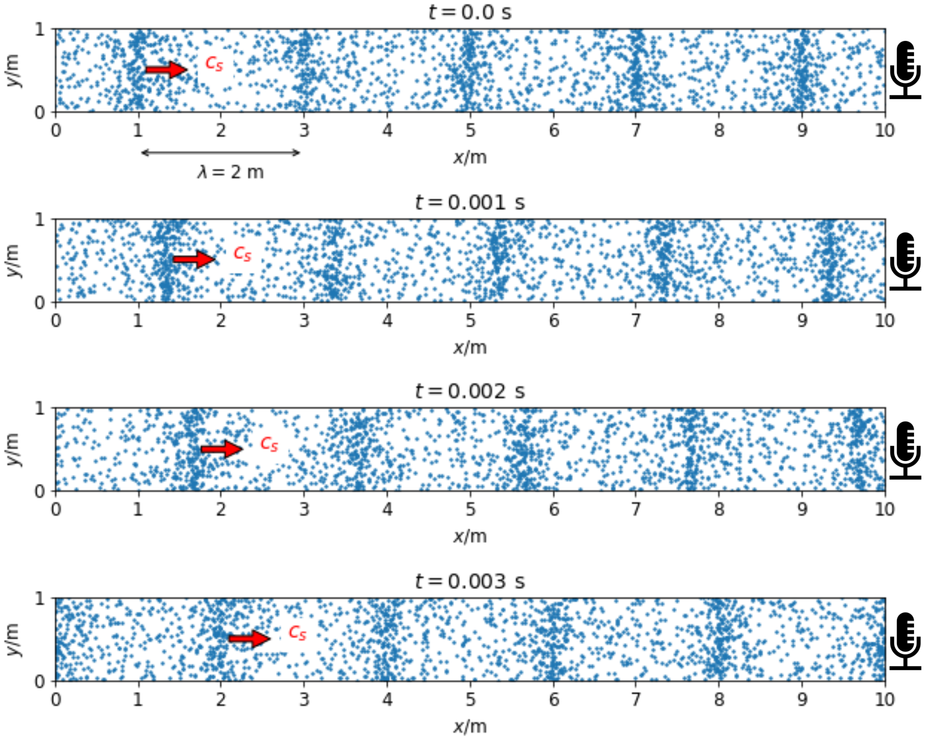

Figure 1. A sinusoidal sound wave travelling to the right with sound speed $c_s$ and wavelength $\lambda$. Top to bottom indicate the snapshots of the air molecules at four instantaneous times.

Now suppose that that we place a microphone at the end of the channel. The compression and rarefaction of air will cause the diaphragm inside the microphone to vibrate. This vibration is then converted into an electrical signal, which is shown in Fig. 2. The horizontal axis in the figure represents the time $t$ (in units of seconds). The vertical axis represents the voltage of the electrical signal produced by the microphone $V(t)$ (in some rescaled units, which we do not need to worry about). As we can see in this example, the audio signal in Fig. 2 has a sinusoidal form, which can be described by a trigonometric function:

$$V(t) = A\cos(\omega t + \phi), \tag{2}$$

where $V(t)$ is the audio signal (in some rescaled units), $t$ is time (in seconds), $A$ is the amplitude, $\omega$ is the angular frequency, and $\phi$ is the phase difference. The angular frequency $\omega$ is related to the frequency $f$ and period $T$ of the sound wave through this relation:

$$\omega = 2\pi f = \frac{2\pi}{T}. \tag{3}$$

From the plot below, we can measure the period to be $T = 0.006\,\mathrm{s}$, which translates to audio frequency of $f = \frac{1}{T} \simeq 167\,\mathrm{Hz}$. Hz (pronounced as Hertz) is the SI unit of frequency, defined to be $\mathrm{Hz} = \mathrm{s}^{-1}$.

In the equation above, $V$, $A$, $\omega$, $\phi$, and $t$ are all real. However, sometimes it might be useful to write the audio signal in a complex form (as we shall see later in Fourier series) as follows:

$$V(t) = Ce^{i\omega t} + C^* e^{-i\omega t}, \text{ where } C = \frac{A}{2}e^{i\phi} \text{ is the complex amplitude (the rest of the variables are real).} \tag{4}$$

The star $*$ above $C$ indicates complex conjugate operation.

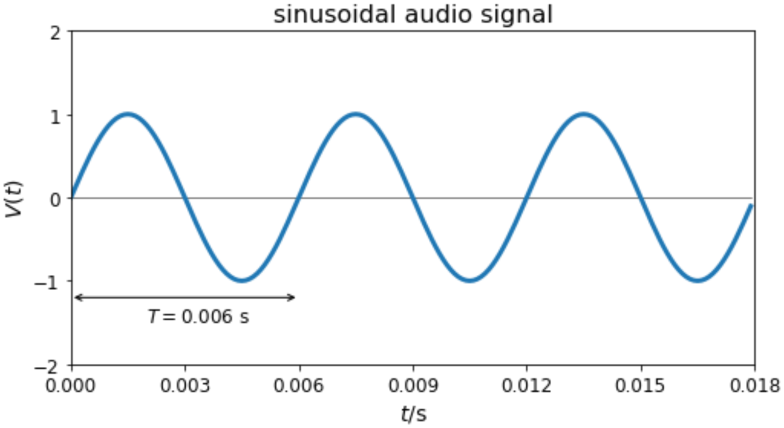

Figure 2. A sinusoidal audio signal with period $T = 0.006\,\mathrm{s}$ and frequency $f = \frac{1}{0.006\,\mathrm{s}} \simeq 167\,\mathrm{Hz}$.

## II. FOURIER SERIES

In Fig. 2, the audio signal can be described by a single trigonometric function. This is because we have assumed the sound wave to be sinusoidal. However this is not true in general. For example, the sound wave coming from a tuning fork is very close to a sinusoidal wave, on the other hand, the sound wave from a saxophone is far from being sinusoidal.

Let us consider another audio signal from an unknown musical instrument, which is depicted in Fig. 3 top. We can immediately tell that the signal is periodic with the same period $T = 0.006\,\mathrm{s}$ (or fundamental frequency $f \simeq 167\,\mathrm{Hz}$) as the one in Fig. 2. However, the shape of the audio signal in Fig. 3 is much more complicated than a sinusoidal wave and cannot be simply described by a single trigonometric function. Luckily, Fourier series allows us to decompose this periodic signal into a sum of trigonometric functions as follows:

$$V(t) = \sum_{p=-\infty}^{\infty} C_p e^{i\omega_p t}, \text{ where } \omega_p = \frac{2\pi p}{T} \text{ and } p \in \mathbb{Z}. \tag{5}$$

Each term in the Fourier series is a simple sinusoidal wave with angular frequency $\omega_p$ and complex amplitude $C_p$'s. The first non-constant term in the Fourier series corresponds to the *fundamental frequency* $\omega_1 = \frac{2\pi}{T}$ shown by first plot in the second row of Fig. 3. The next term in the Fourier series has double the fundamental frequency $\omega_2 = \frac{4\pi}{T}$ and is sometimes called the second harmonic (see second plot in the second row of Fig. 3). The next next term has triple the fundamental frequency $\omega_3 = \frac{6\pi}{T}$ and is sometimes called the third harmonic (see third plot in the second row of Fig. 3).

To find the complex amplitudes $C_p$'s (or Fourier coefficients) we multiply the above equation by $e^{-i\omega_q t}$ and then integrate with respect to $t$ from $t = 0$ to $t = T$.

$$\int_0^T V(t) e^{-i\omega_q t}\, dt = \sum_{p=-\infty}^{\infty} C_p \int_0^T e^{i(\omega_p - \omega_q)t}\, dt, \tag{6}$$

where $p$ and $q$ are integers. We note that the integral:

$$\int_0^T e^{i(\omega_p-\omega_q)t}\,dt = \int_0^T e^{i\frac{2\pi}{T}(p-q)t}\,dt = \begin{cases} T & \text{if } p=q \\ 0 & \text{if } p\neq q \end{cases}. \tag{7}$$

More succintly, we can write,

$$\int_0^T e^{i(\omega_p-\omega_q)t}\,dt = T\delta_{pq}, \tag{8}$$

where the Kronecker delta $\delta_{pq}$ is defined to be equal to 1 if $p=q$ and 0 if $p\neq q$. Therefore,

$$\int_0^T V(t)e^{-i\omega_q t}\,dt = \sum_{p=-\infty}^{\infty} C_p T\delta_{pq} = TC_q \tag{9}$$

$$\Rightarrow C_q = \frac{1}{T}\int_0^T V(t)e^{-i\omega_q t}\,dt. \tag{10}$$

**Question 1.** (a) Show that if $C_{-p} = C_p^*$ for all $p \in \mathbb{Z}$, then $V(t)$ is real. (b) Show that if $V(t)$ is real, then $C_{-p} = C_p^*$ for all $p \in \mathbb{Z}$.

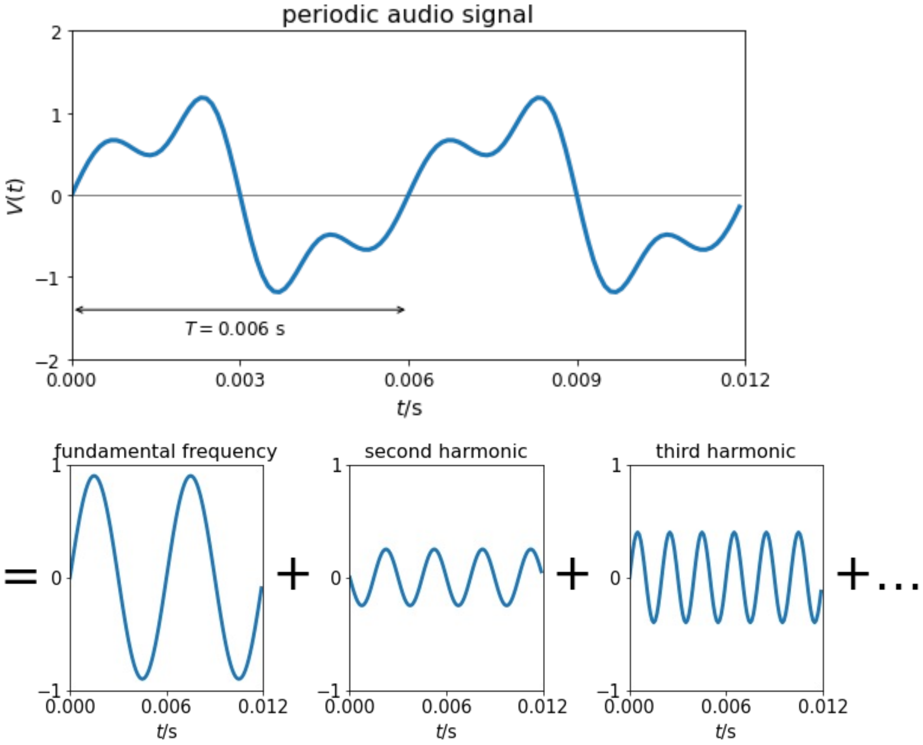

Figure 3. Any periodic function can be decomposed into an infinite sum of trigonometric function with frequency doubling in each subsequent term (Fourier series). The first term in the Fourier series has the frequency equals to the original function (fundamental frequency).

## III. FOURIER TRANSFORM

In reality, the audio signal is not perfectly periodic since it might be contaminated by a background noise. Furthermore the oscillation of the sound wave itself decays gradually to zero, so a realistic representation of an audio signal recorded from a microphone might look something like Fig. 4(a).

To analyse this audio signal, we shall use Fourier transform, which is an extension of the Fourier series above by taking the limit $T \to \infty$ (loosely speaking, the function $V(t)$ is no longer periodic). In this case the angular frequency of each Fourier component becomes continuous $\omega_p \to \omega \in \mathbb{R}$. The Fourier transform of $V(t)$ is another function $\tilde{V}(\omega)$, which is a function of angular frequency $\omega$. $V(t)$ and $\tilde{V}(\omega)$ are related through the following relations:

$$V(t) = \int_{-\infty}^{\infty} \tilde{V}(\omega) e^{i\omega t} d\omega \quad \text{(inverse Fourier transform)} \tag{11}$$

$$\tilde{V}(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} V(t) e^{-i\omega t} dt \quad \text{(Fourier transform)}, \tag{12}$$

where $\omega \in \mathbb{R}$. The above relations can be derived by taking the limit $T \to \infty$ in the definition of Fourier series, in which case the summation over $\omega_p$ becomes an integral over $\omega$, and then relabelling $C_p \to \tilde{V}(\omega_p) \to \tilde{V}(\omega)$.

Now let us define the Dirac delta function $\delta(x-y)$, where $x, y \in \mathbb{R}$ such that

$$\delta(x-y) = \begin{cases} 0 & \text{if } x \neq y \\ \infty & \text{if } x = y \end{cases}. \tag{13}$$

The Dirac delta function has the following properties:

$$\int_{-\infty}^{\infty} \delta(x-y)\, dx = 1, \quad \text{and} \quad \int_{-\infty}^{\infty} \delta(x-y) f(x)\, dx = f(y). \tag{14}$$

In continous space, though it is ill-defined, its graph would take the shape of an infinitely thin curve that peaks at a single point, $x = y$, and is zero elsewhere. The area (integral) under this curve is equal to 1, see the second figure below.

**Question 2.** Using the definition of Fourier transform, show that

$$\int_{-\infty}^{\infty} e^{i(\alpha-\beta)t}\, dt = 2\pi\delta(\alpha-\beta), \tag{15}$$

for all $\alpha, \beta, t \in \mathbb{R}$.

**Question 3.** Show that if $V(t)$ is real then $\tilde{V}(-\omega) = \tilde{V}(\omega)^*$ for all $\omega \in \mathbb{R}$ and *vice versa*.

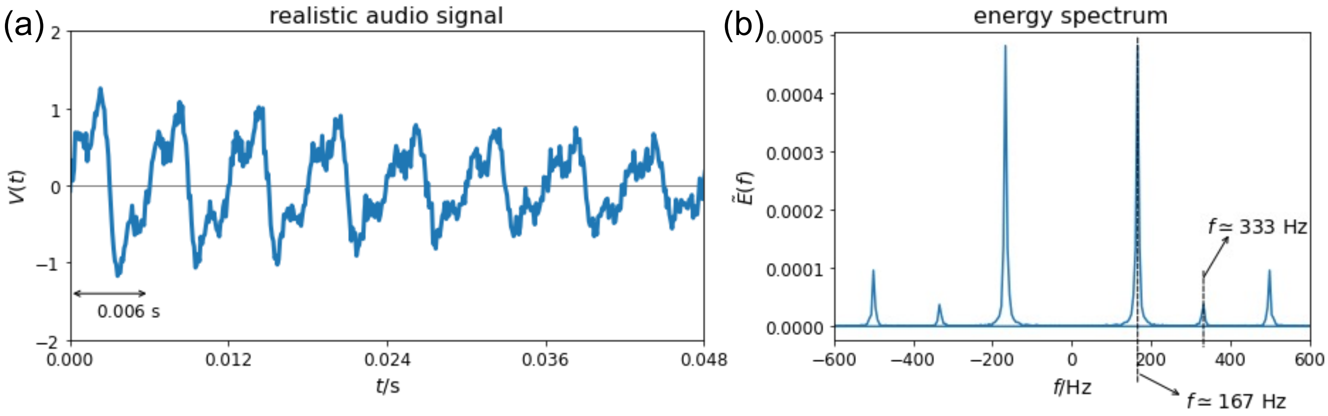

Figure 4. (a) shows a typical noisy audio signal. (b) shows the corresponding energy spectrum. The energy spectrum shows several sharp peaks. The first peak corresponds to the fundamental frequency $f = \frac{1}{0.006\,\mathrm{s}} \simeq 167\,\mathrm{Hz}$.

## IV. ENERGY SPECTRUM

From electromagnetism, we learnt that the rate of power dissipation is given by:

$$P(t) = \frac{V(t)^2}{R}, \quad \text{where } R \text{ is the electrical resistance.} \tag{16}$$

The total energy dissipation is then given by the integral

$$E = \frac{1}{R}\int_{-\infty}^{\infty} V(t)^2\, dt \tag{17}$$

Substituting the definition for Fourier transform to the above, we get:

$$\begin{aligned} E &= \frac{1}{R}\int dt \int d\omega \int d\omega'\, \tilde{V}(\omega)\tilde{V}(\omega')e^{i(\omega+\omega')t} \qquad (18)\\ &= \frac{2\pi}{R}\int d\omega \int d\omega'\, \tilde{V}(\omega)\tilde{V}(\omega')\delta(\omega+\omega') \qquad (19)\\ &= \frac{2\pi}{R}\int d\omega\, \tilde{V}(\omega)\tilde{V}(-\omega) \qquad (20)\end{aligned}$$

Since $V(t)$ is real, we have $\tilde{V}(-\omega) = \tilde{V}(\omega)^*$ from the previous exercise and thus the total energy dissipation can be written as:

$$E = \frac{2\pi}{R}\int_{-\infty}^{\infty} d\omega\, |\tilde{V}(\omega)|^2 \tag{21}$$

Now we can define the energy spectrum to be $\tilde{E}(\omega) = |\tilde{V}(\omega)|^2$. Physically, Fourier transform allows us to decompose an electrical signal into an infinite sum of sinusoidal oscillations with different angular frequencies $\omega$'s. The energy spectrum $\tilde{E}(\omega)$ gives the energy contribution from a single oscillation with particular angular frequency $\omega$.

Fig. 4(b) shows the energy spectrum $\tilde{E}(f)$ as a function of frequency $f$, which corresponds to the noisy audio signal $V(t)$ in Fig. 4(a). Note that the frequency $f$ is related to the angular frequency by a factor of $2\pi$, *i.e.* $\omega = 2\pi f$. As we can see in the figure below, the energy spectrum spectrum is symmetric with respect to $f \to -f$ (and for this reason, $\tilde{E}(f)$ is usually plotted on the positive $x$-axis only). Furthermore we also observe several sharp peaks in the energy spectrum. The first peak $f \simeq 167\,\text{Hz}$ corresponds to the fundamental frequency of the signal. (Although the signal is no longer periodic, it still retains some underlying periodic characteristics.) The second peak $f \simeq 333\,\text{Hz}$ (which is double the fundamental frequency) corresponds to the second harmonic and so on.

**Question 4.** Show that the energy spectrum is symmetric, *i.e.* $\tilde{E}(\omega) = \tilde{E}(-\omega)$.

## V. DISCRETE FOURIER TRANSFORM

TO BE ADDED LATER

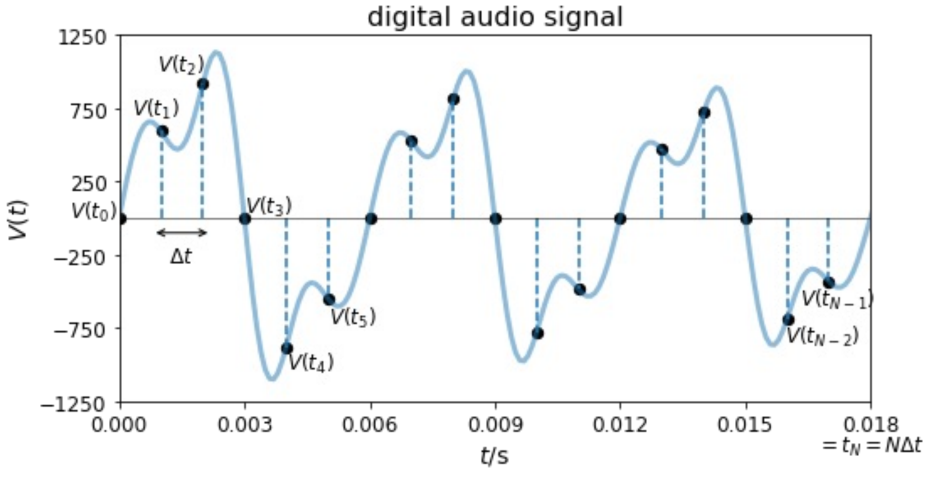

Figure 5. In computer, the audio signal $V(t)$ is discretized into a finite set of points $\{V(t_0), V(t_1), \ldots V(t_{N-1})\}$.

## VI. STORING AUDIO WAVE INTO A COMPUTER

Let's say we record a sound on our microphone for the total duration of $t_N$ (also called the sampling time). How is this audio signal $V(t)$ stored inside our computer? Obviously a computer cannot store an infinite amount of information, so the computer has to divide the signal $V(t)$ into discrete values $V(t_0), V(t_1), V(t_2), \ldots, V(t_{N-1})$ at discrete times $t_n$'s, as we can see in Fig. 5. This also means that the time is discretized into: $t \to t_n = n\Delta t$, where $n = 0, 1, 2, \ldots, N-1$ and $\Delta t$ is the timestep. In Fig. 5, the total time (or sampling time) is $t_N = 0.018\,\text{s}$, the timestep is $\Delta t = 0.001\,\text{s}$, and the total number of points is $N = 18$. We also define the framerate to be the total number of points $N$ per unit time. In Fig. 5, we can calculate the framerate to be:

$$\text{framerate} = \frac{N}{t_N} = \frac{18}{0.018\,\text{s}} = 1000\,\text{s}^{-1}. \tag{22}$$

Usually, when we record a sound using recording software such as Audacity, we need to specify this framerate. Higher framerate will give a better sound quality but the file size will also be bigger! For a 16-bit digital audio, the values of $V$ ranges from $-32768$ to $32767$ in integer steps (note that $V(t)$ is in some rescaled units). Therefore the vertical $V$-axis in Fig. 5 is also discrete.

## VII. IMPORTING .WAV FILE

There are different audio formats in computer such as .mp3 and .wav. The difference is that the former is a compressed file while the later is an uncompressed file. In this Tutorial, we will only consider '.wav' files.

Let us now put everything we have learnt above into practice. Inside the folder "./samples/" we have various .wav files from different notes from different musical instruments. First let us have a look at the file called "./samples/piano-C4.wav". This is a recording of the note C4 in a piano. This note has a fundamental frequency of 261.63 Hz. Let us now analyse this wave file below. First we need to import the essential libraries such as numpy, matplotlib, sys and wave into Python, as you can see in the first few lines of the code below. The library sys is used to access the filesystem inside our computer and the library wave is used to import and export .wav files. Next we will open the file "./samples/piano-C4.wav" and store it into an object, called inputfile, using the method wave.open("./samples/piano-C4.wav", "r"). The letter "r" indicates that we are reading the file (while the letter "w" indicates writing into a file). Before we proceed further, we need some information about this audio signal, which we just imported. For example, we need the framerate, which we can obtain using the method .getframerate(). In this case the framerate is $11025\,\text{s}^{-1}$, which we can confirm by printing the value into the screen. We also need the total number of points, *i.e.* $N$, which we can obtain using the method .getnframes(). In this case we get $N = 29750$. Hence we can calculate the total recording time (or sampling time) $t_N$:

$$t_N = \frac{N}{\text{framerate}} \simeq 2.70\,\text{s}. \tag{23}$$

We also need to compute the timestep $\Delta t$:

$$\Delta t = \frac{t_N}{N} = \frac{1}{\text{framerate}} \simeq 0.0000907\,\text{s}. \tag{24}$$

To read the actual audio signal itself from the .wav file, we use the method .readframes(-1). If you try to print the output you will get something like this: b'\xf6\xf1\xff\xf1\xff...\xf2\xff'. The strange combination of three characters, separated by backslash, actually represents a binary number. We need to convert these binary numbers, called bytes objects, into decimals using the numpy method np.frombuffer(V, dtype=np.int16). The int16 data type indicates that it is a 16-bit audio, so that the values of $V$ ranges from $-32768$ to $32767$ in integer steps.

To plot the the audio signal $V(t)$ as a function of time $t$ we use the library matplotlib.pyplot. First we need to define an array $t$ to represent the time $\{0, \Delta t, 2\Delta t, \ldots, (N-1)\Delta t\}$. To do this we use the method t = np.arange(0, tN, dt). Finally, to make the plot, we use the method plt.plot(t, V).

```
import numpy as np
import matplotlib.pyplot as plt
import wave  # this library is to import and export .wav files
import sys  # this library is to access the filesystem
```

```
# read a .wav file and store it into an object
inputfile = wave.open('./samples/piano-C4.wav', 'r')

# get the framerate of the audio file (in units of seconds^-1) and print into the screen
framerate = inputfile.getframerate()
print(f'framerate = {framerate} s^-1')

# get the total number of points
N = inputfile.getnframes()
print(f'N = {N}')

# calculate the total time
tN = N/framerate
print(f't_N = {N/framerate} s')

# calculate the timestep
dt = 1/framerate
print(f'dt = {dt} s')

V = inputfile.readframes(-1)  # read audio signal from the input .wav file
V = np.frombuffer(V, dtype=np.int16)  # convert the binary format into integers

t = np.arange(0, tN, dt)

fig, ax = plt.subplots(figsize=(8, 4))

ax.set_title("C4 note from a piano")
ax.set_xlabel('$t/$s', fontsize=14)
ax.set_ylabel('$V(t)$', fontsize=14)
ax.set_xlim(0, 2.7)
ax.set_ylim(-3000, 3000)
ax.tick_params(axis='both', which='major', labelsize=12)

plt.plot(t, V)
plt.show()
```

If we run the code above, it will print the framerate, $N$, $t_N$, and $\Delta t$ of the "piano-C4.wav" file into the screen, and plot the audio signal $V(t)$ as a function of $t$, as shown in Fig. 6(a).

**Question 5.** Add few lines of code to the above to plot the audio signal $V(t)$ over a narrower time interval $t \in [0.1\,\mathrm{s}, 0.14\,\mathrm{s}]$.

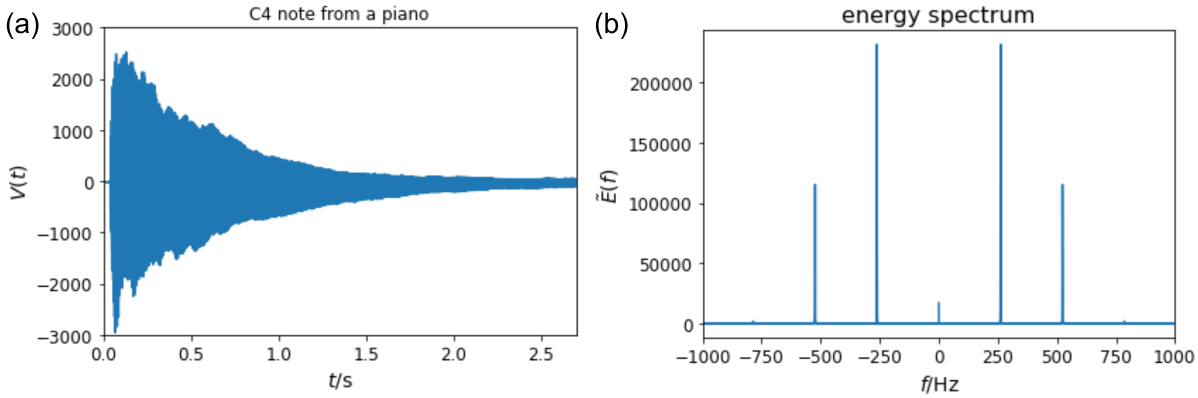

Figure 6. (a) shows a recording of the note C4 from a piano. (Recording duration $\simeq 2.7\,\mathrm{s}$.) (b) shows its corresponding energy spectrum.

## VIII. ANALYSING SOUND WAVE USING FAST FOURIER TRANSFORM

TO BE ADDED LATER
**Question XX.** Write a code which tells the note of a piano key.
**Question XX.** Noise reduction?