International Baccalaureate Diploma Program May 2024

Extended Essay in Mathematics

Title: Fourier transforms and audio comparison

Research Question: *How can Fourier transforms be used to efficiently represent and accurately recognise audio signals?*

Word count: 3994

Table of Contents

1 Introduction **3**

2 Decomposing periodic functions into sinusoids **4**

2.1 Defining periodic functions 4

2.2 Periodicity of sinusoidal functions 4

2.3 Fourier Series 4

2.3.1 Definition of a Fourier series 4

2.3.2 Finding Fourier coefficients mathematically 5

2.3.3 Example calculation of Fourier series of a function 8

2.4 Expressing Fourier Series coefficients as complex numbers 11

3 The Fourier Transform **14**

3.1 Deriving the continuous Fourier transform 14

3.2 The discrete Fourier transform 15

3.2.1 Modifying the continuous Fourier transform 15

3.2.2 Optimising the DFT computation 16

3.2.3 Zero padding 19

3.2.4 Evaluating the efficiency of the FFT 19

4 Comparing signals using Fourier transforms **20**

4.1 Short-time Fourier transforms 20

4.2 Developing a recognition system 22

4.2.1 System overview 22

4.2.2 Setting STFT parameters 23

4.2.3 Methods used to evaluate recognition system 24

4.2.4 Results of testing 26

5 Conclusion **27**

Works cited **29**

A Appendices **31**

A.1 Proofs for vanishing integrals of sinusoids 31

A.2 Verification of result of complex Fourier series for square function 35

A.3 Additivity of DFT 36

A.4 Data tables 37

A.4.1 Comparison of DFT and FFT 37

A.4.2 List of audio tracks used 38

A.5 Code used to evaluate DFT and FFT 40

1 Introduction

In modern contexts, it is increasingly common for technology to be used to recognise audio signals from known sources, whether that is to identify copyrighted media posted online, or in mobile applications. To accomplish this, we require a means to efficiently represent and identify clips of recognisable audio from a database.

Audio is captured in the *time domain*, where the value of a signal is measured with respect to time[1]. However, time-domain values of real-world audio clips are unlikely to be identical to values of the original audio in databases, due to real-world factors such as uncontrolled external noise. Additionally, there is no way to compare the importance of individual time-domain values for identifying a signal, so accurate recognition of an audio signal would require an inefficiently large number of values to be compared.

It is therefore the aim of this essay to analyse mathematical methods by which discrete signals can be efficiently represented, such that even short portions of them can be accurately recognised, even when mixed with other audio, such as background noise.

To do this, I plan to investigate the use of Fourier transforms, a mathematical method to decompose signals into pure waves in the form of sine and cosine functions. This allows us to describe signals just in terms of their constituent frequencies, i.e., in the *frequency domain* instead of time domain. By giving us information about audio over a longer period of time, frequency-domain representations may enable accurate recognition of audio signals.

Therefore, the guiding research question of this essay is: **"How can Fourier transforms be used to efficiently represent and accurately recognise audio signals?"**

1. Smith, Steven W. *The Scientist and Engineer's guide to digitial signal processing.* California Technical Publishing. 1998.

2 Decomposing periodic functions into sinusoids

2.1 Defining periodic functions

A function $f(x)$ is said to be periodic in period T if, for all x in the domain of f,

$$f(x) = f(x + T)$$

The **fundamental period** of the function is the smallest positive value of T for which this equation holds. A periodic function with fundamental period T is called T-periodic. Periodicity holds for the sum of periodic functions, i.e., if $f(x)$ and $g(x)$ are T-periodic functions, $f(x + T) + g(x + T) = f(x) + g(x)$.

The **fundamental frequency** of a function is defined as the reciprocal of its fundamental period, i.e., $\frac{1}{T}$. A T-periodic function is then also periodic for all the **harmonics** of this frequency, i.e., the frequencies which are integral multiples of the fundamental frequency.

2.2 Periodicity of sinusoidal functions

Sinusoidal functions are those functions that are based on the trigonometric sine function $\sin(x)$. Since the cosine function $\cos(x)$ can also be written as a sine function. $\sin(x)$ and $\cos(x)$ are 2π-periodic. The more general form of sinusoids, $\sin(ax)$ and $\cos(ax)$, can be viewed as a compression by scale factor a, resulting in the fundamental period of these functions being rescaled from 2π to $\frac{2\pi}{a}$.

2.3 Fourier Series

2.3.1 Definition of a Fourier series

Similar to how a Taylor series may be used to decompose a function by writing it as a sum of polynomials, we can write an expansion of any periodic function $f(x)$ as the sum of harmonically related sinusoids. This is referred to as the **Fourier series** of a function[2].

Formally, **Fourier's theorem for periodic functions**[3] states that, for a piecewise continuous function $f(x)$ with period $2L$, there exist constants a_0, a_m, and b_m for which the the below infinite sum converges to $f(x)$ at each point of continuity.

$$f(x) = \frac{a_0}{2} + \sum_{m=1}^{\infty} \left[a_m \cos\left(\frac{m\pi x}{L}\right) + b_m \sin\left(\frac{m\pi x}{L}\right) \right] \tag{2.1}$$

The period of the sinusoids is equal to $\frac{2\pi L}{m\pi} = \frac{2L}{m}$. This means that the sinusoids for $m = 1$ have a period of 2L, i.e., their frequency $\frac{1}{2L}$ is equal to the fundamental frequency of $f(x)$. Sinusoids for $m \geq 2$ will have periods of $\frac{2L}{m}$, meaning that they will complete m complete periodic cycles over an interval of $2L$. Additionally, their frequencies will be $\frac{m}{2L}$, i.e., they will be harmonics of the fundamental frequency.

2.3.2 Finding Fourier coefficients mathematically

To compute the Fourier expansion of a function, the mathematical constants a_0, a_m, and b_m must be calculated. To accomplish this, we need a way to isolate each term containing the coefficients.

One property of sine and cosine functions which can be exploited for this purpose is the fact that they vanish when integrated over intervals in which they are periodic, i.e., from $-L$ to L for $2L$-periodic sinusoids. Mathematically, we can show that the following relations hold for all $n, m \in \mathbb{N}$, the complete proofs for which are given in Appendix A.1.

$$\int_{-L}^{L} \sin(\frac{n\pi x}{L}) = \int_{-L}^{L} \cos(\frac{n\pi x}{L})\, dx = 0 \text{ for all n} \tag{2.2}$$

$$\int_{-L}^{L} \sin(\frac{n\pi x}{L}) \cdot \cos(\frac{m\pi x}{L})\, dx = 0 \text{ for all n, m} \tag{2.3}$$

2. Osgood, Brad. *Lecture Notes for EE261: The Fourier Transform and its Applications.* Stanford University, 2007.

3. Mattuck, Arthur et al. "Fourier Series: Definitions and Coefficients - 18.03SC Differential Equations". 18.03SC Differential Equations. Massachusetts Institute of Technology.

$$\int_{-L}^{L} \sin(\frac{n\pi x}{L}) \cdot \sin(\frac{m\pi x}{L})\, dx = \begin{cases} 0 & \text{if } n = m = 0 \\ L & \text{if } n = m \neq 0 \\ 0 & \text{if } n \neq m \end{cases} \tag{2.4}$$

$$\int_{-L}^{L} \cos(\frac{n\pi x}{L}) \cdot \cos(\frac{m\pi x}{L})\, dx = \begin{cases} 2L & \text{if } n = m = 0 \\ L & \text{if } n = m \neq 0 \\ 0 & \text{if } n \neq m \end{cases} \tag{2.5}$$

Now, we reconsider the Fourier series from equation 2.1.

$$f(x) = \frac{a_0}{2} + \sum_{m=1}^{\infty} \left(a_m \cos\left(\frac{m\pi x}{L}\right) + b_m \sin\left(\frac{m\pi x}{L}\right)\right)$$

To find an expression for a_0, we can now exploit the vanishing property derived in equation 2.2, by taking the integral of both sides of the above expression over the period of $f(x)$.

$$\begin{aligned} \int_{-L}^{L} f(x)\, dx &= \int_{-L}^{L} \left[\frac{a_0}{2} + \sum_{m=1}^{\infty} a_m \cos\left(\frac{m\pi x}{L}\right) + \sum_{m=1}^{\infty} b_m \sin\left(\frac{m\pi x}{L}\right)\right] dx \\ &= a_0 L + \sum_{m=1}^{\infty} \left[\int_{-L}^{L} a_m \cos\left(\frac{m\pi x}{L}\right) dx + \int_{-L}^{L} b_m \sin\left(\frac{m\pi x}{L}\right) dx \right] \\ &= a_0 L + 0 \qquad \text{(using equation 2.2)} \end{aligned}$$

$$\therefore a_0 = \frac{1}{L} \int_{-L}^{L} f(x)\, dx \tag{2.6}$$

Similarly, we arrive at an expression for a_m by multiplying both sides by a cosine term $\cos\left(\frac{n\pi x}{L}\right)$, and taking their integrals:

$$\begin{aligned} \int_{-L}^{L} f(x) \cos\left(\frac{n\pi x}{L}\right) dx &= \int_{-L}^{L} \left[\frac{a_0}{2} + \sum_{m=1}^{\infty} a_m \cos\left(\frac{m\pi x}{L}\right) + \sum_{m=1}^{\infty} b_m \sin\left(\frac{m\pi x}{L}\right)\right] \cdot \cos\left(\frac{n\pi x}{L}\right) dx \\ &= \frac{a_0}{2} \int_{-L}^{L} \cos\left(\frac{n\pi x}{L}\right) + \sum_{m=1}^{\infty} a_m \int_{-L}^{L} \cos\left(\frac{m\pi x}{L}\right) \cdot \cos\left(\frac{n\pi x}{L}\right) + \end{aligned}$$

$$\sum_{m=1}^{\infty} b_m \int_{-L}^{L} \sin\left(\frac{m\pi x}{L}\right) \cdot \cos\left(\frac{n\pi x}{L}\right)$$

Every term in the expansion of the second term of the above expression will be 0, as per equation 2.5, except when $n = m$.

As per equations 2.2 and 2.4, the other terms will also evaluate to 0.

$= 0 + L \cdot a_m + 0$

$$\therefore a_m = \frac{1}{L} \int_{-L}^{L} f(x) \cos\left(\frac{m\pi x}{L}\right) \, dx \tag{2.7}$$

Finally, the expression for b_m is found by multiplying both sides by a sine term $\sin\left(\frac{n\pi x}{L}\right)$, and taking their integrals:

$$\begin{aligned}
\int_{-L}^{L} f(x) \sin\left(\frac{n\pi x}{L}\right) \, dx &= \int_{-L}^{L} \left[\frac{a_0}{2} + \sum_{m=1}^{\infty} a_m \cos\left(\frac{m\pi x}{L}\right) + \sum_{m=1}^{\infty} b_m \sin\left(\frac{m\pi x}{L}\right)\right] \cdot \sin\left(\frac{n\pi x}{L}\right) \, dx \\
&= \frac{a_0}{2} \int_{-L}^{L} \sin\left(\frac{n\pi x}{L}\right) + \sum_{m=1}^{\infty} a_m \int_{-L}^{L} \cos\left(\frac{m\pi x}{L}\right) \cdot \sin\left(\frac{n\pi x}{L}\right) + \\
&\quad \sum_{m=1}^{\infty} b_m \int_{-L}^{L} \sin\left(\frac{m\pi x}{L}\right) \cdot \sin\left(\frac{n\pi x}{L}\right) \\
&= b_m L \qquad \text{(Using equations 2.2, 2.3, 2.4)}
\end{aligned}$$

$$\therefore b_m = \frac{1}{L} \int_{-L}^{L} f(x) \sin\left(\frac{m\pi x}{L}\right) \, dx \tag{2.8}$$

Thus, the coefficients a_0, a_m, and b_m of a Fourier series are given by:

$$\begin{aligned}
a_0 &= \frac{1}{L} \int_{-L}^{L} f(x) \, dx \\
a_m &= \frac{1}{L} \int_{-L}^{L} f(x) \cos\left(\frac{m\pi x}{L}\right) \, dx \\
b_m &= \frac{1}{L} \int_{-L}^{L} f(x) \sin\left(\frac{m\pi x}{L}\right) \, dx
\end{aligned}$$

We have therefore derived formulae by which we can calculate the Fourier coefficients of a $2L$-periodic function. $\frac{a_0}{2}$ represents the average value of $f(x)$ over its period, thus, it vertically translates the sum such that the sinusoids effectively oscillate around the average value of $f(x)$. a_m and b_m represent the magnitude present of each of the corresponding sinusoids with frequencies equal to the fundamental frequency of $f(x)$ and its harmonic frequencies.

2.3.3 Example calculation of Fourier series of a function

To better understand the behaviour of Fourier series, we can try to derive one for an example function of our choice. One function whose Fourier series illustrates the property of convergence, as well as the limitations of Fourier series in extreme cases, is the 2π-periodic odd square wave function, which is mathematically defined[4] below and plotted in Fig. 1.

$$f(x) = \begin{cases} +1 & \text{If } \sin(x) \text{ is positive} \\ -1 & \text{If } \sin(x) \text{ is negative} \end{cases} \tag{2.9}$$

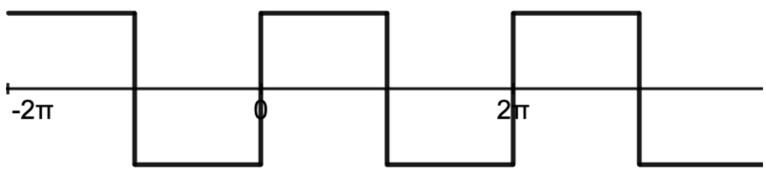


Figure 1: Odd square function with amplitude 1 and period 2π

Now, we try to find the Fourier series of $f(x)$ over the interval $(-\pi, \pi)$. Note that since $f(x)$ is an odd function, $f(-x) = -f(x)$. Hence, as for all odd functions, its integral over its period will be equal to 0:

$$\begin{aligned} &\int_{-\pi}^{\pi} f(x)\,dx \\ &= \int_{-\pi}^{0} f(x)\,dx + \int_{0}^{\pi} f(x)\,dx \\ &= 0 \end{aligned}$$

Hence, as per equation 2.6, we find that the constant term of the Fourier series of $f(x)$

4. Strang, Gilbert. *Computational Science and Engineering.* Cambridge University Press, 2007.

has a value of 0.

$$a_0 = \frac{1}{L}\int_{-\pi}^{\pi} f(x)\,dx = \frac{1}{L} \times 0 = 0$$

Next, we attempt to calculate the general form of a_m for this function.

$$\begin{aligned} a_m &= \frac{1}{L}\int_{-L}^{L} f(x)\cos\left(\frac{m\pi x}{L}\right) \\ &= \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos(mx) \end{aligned}$$

Notice that in addition to $f(x)$ being an odd function, $\cos(mx)$ is known to be an even function for all values of m, i.e., $\cos(-mx) = \cos(mx)$. An elementary property[5] of odd and even functions is that the multiple of odd and even functions is also odd. Hence, the above integral is effectively an integral of an odd function over its period. It thus also evaluates to 0, and therefore,

$$a_m = \frac{1}{L} \times 0 = 0$$

The Fourier series of $f(x)$ therefore has all $a_m = 0$, implying it has no effectively no cosine components. This makes sense considering that the square wave was defined based on a sine function. Now, finally, to find the coefficients of the sine components, we calculate the general form of b_m.

$$\begin{aligned} b_m &= \frac{2}{\pi}\int_{0}^{L} f(x)\sin(mx)\;dx \\ &= \frac{2}{\pi}\int_{0}^{\pi} f(x)\sin(mx)\;dx && (2L = 2\pi \text{ for a } 2\pi\text{-periodic function}) \\ &= \frac{2}{\pi}\left[\frac{-\cos(mx)}{m}\right]_0^{\pi} && (f(x) = 1 \text{ over the interval } (0,\pi)) \\ &= \frac{2}{\pi}\left[\frac{-\cos(m\pi)}{m} + 1\right] && (-\cos(0) = +1) \end{aligned}$$

5. "Numeracy, Maths and Statistics - Academic Skills Kit - Odd and Even Functions". www.ncl.ac.uk/webtemplate/ask-assets/external/maths-resources/core-mathematics/pure-maths/functions/odd-and-even-functions.html. Newcastle University, 2008. Accessed 20 June 2023.

Finally, we can substitute $a_0 = 0$, $a_m = 0$, and the above formula for b_m to get the Fourier series of $f(x)$:

$$\begin{aligned} f(x) &= \sum_{m=1}^{\infty} b_m \sin\left(\frac{m\pi x}{L}\right) \\ &= \sum_{m=1}^{\infty} b_m \sin(mx) \quad \text{since } L = \pi \\ &= \sum_{m=1}^{\infty} \frac{2}{\pi}\left[\frac{-\cos(m\pi)}{m} + 1\right] \sin(mx) \end{aligned}$$

Testing some values of m gives a pattern for b_m, wherein only odd-numbered values $m = 2k+1$ give nonzero values of b_m.

$$\begin{aligned} &\frac{2}{\pi}\left[\frac{2}{1}, \frac{0}{2}, \frac{2}{3}, \frac{0}{4}, \frac{2}{5} \dots\right] \text{ for } m = 1, 2, 3, 4, 5 \dots \\ &= \frac{2}{\pi}(2k+1) \text{ for k} = 0 \text{ to } \infty \end{aligned}$$

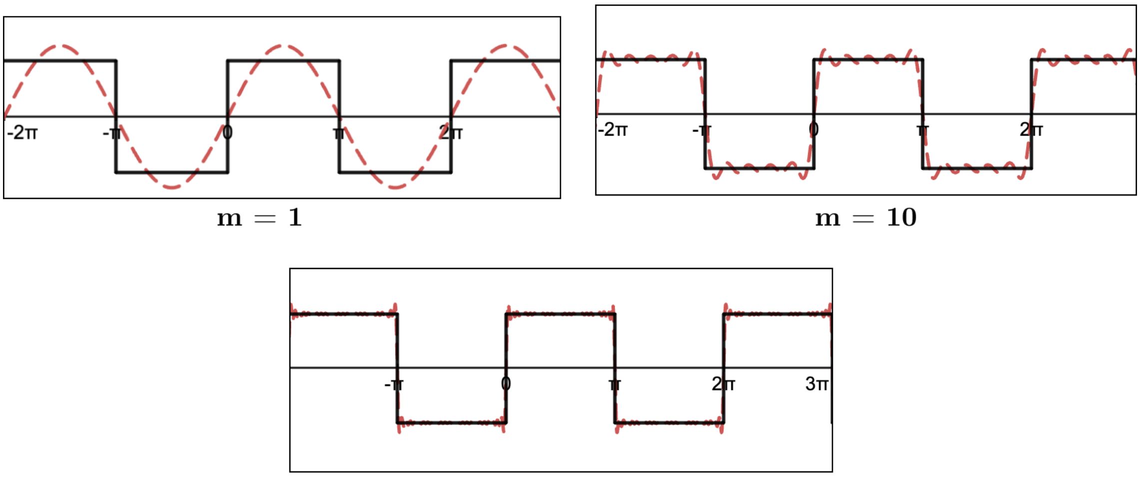
Hence, we can rewrite the Fourier series sum in terms of k.

$$= \frac{4}{\pi} \sum_{k=0}^{\infty} \frac{1}{2k+1} \sin[(2k+1)x] \tag{2.10}$$

Which can be expanded as

$$\frac{4}{\pi}\left[\frac{\sin x}{1} + \frac{\sin 3x}{3} + \frac{\sin 5x}{5} + \dots\right] \text{ for } k = 0, 1, 2, \dots$$

We plot the behaviour of this sum as m increases in Fig. 2.



m = 1 **m = 10**

m = 50

Figure 2: Comparison of Fourier series expansions of odd 2π-periodic square wave $f(x)$

We see by eye that as we arbitrarily expand the series by increasing m, i.e., the number of harmonics used, we more accurately approximate $f(x)$. This is consistent with the convergence of a Fourier series predicted by the theorem from Section 2.3.1.

2.4 Expressing Fourier Series coefficients as complex numbers

Consider a Fourier series in m as defined by equation 2.1. For convenience, we write $\frac{\pi}{L}$ as ω, which is a value proportional to the resolution of frequencies we resolve with a Fourier series. With this, the expression for a Fourier series from equation 2.1 can be written as:

$$\frac{a_0}{2} + \sum_{m=1}^{\infty} a_m \cos(m\omega x) + b_m \sin(m\omega x)$$

Now, we can try to write the two sinusoid terms as one term, by using Euler's identity.

$$e^{ix} = \cos(x) + i\sin(x)$$
$$\implies e^{-ix} = \cos(-x) + i\sin(-x)$$
$$\implies e^{-ix} = \cos(x) - i\sin(x)$$

$$\therefore e^{ix} + e^{-ix} = 2\cos(x) \qquad\qquad \therefore e^{ix} - e^{-ix} = 2i\sin(x)$$
$$\implies \cos(x) = \frac{e^{ix} + e^{-ix}}{2} \qquad\qquad \implies \sin(x) = \frac{e^{ix} - e^{-ix}}{2i}$$

We can therefore split the Fourier series into separate sums of cosines and sines and then make the above substitutions for $\cos(m\omega x)$ and $\sin(m\omega x)$.

$$\begin{aligned}
&= \frac{a_0}{2} + \sum_{m=1}^{\infty}\left[\frac{a_m}{2}e^{\omega imx} + \frac{a_m}{2}e^{-\omega imx}\right] + \sum_{m=1}^{\infty}\left[\frac{-ib_m}{2}e^{\omega imx} + \frac{ib_m}{2}e^{-\omega imx}\right] \\
&= \frac{a_0}{2} + \sum_{m=1}^{\infty}\frac{a_m - ib_m}{2}e^{\omega imx} + \sum_{m=1}^{\infty}\frac{a_m + ib_m}{2}e^{-\omega imx} \\
&= \frac{a_0}{2} + \sum_{m=1}^{\infty}\frac{a_m - ib_m}{2}e^{\omega imx} + \sum_{m=-1}^{-\infty}\frac{a_{-m} + ib_{-m}}{2}e^{\omega imx}
\end{aligned}$$

Note that since $\cos(-m) = \cos(m)$, as per equation 2.7, $a_{-m} = a_m$. Similarly, since $\sin(-m) = -\sin(m)$, as per equation 2.8, $b_{-m} = -b_m$. Hence, the sum simplifies to:

$$\begin{aligned}
&= \frac{a_0}{2} + \sum_{m=1}^{\infty}\frac{a_m - ib_m}{2}e^{\omega imx} + \sum_{m=-1}^{-\infty}\frac{a_m - ib_m}{2}e^{\omega imx} \\
&= \sum_{m=1}^{\infty}\frac{a_m - ib_m}{2}e^{\omega imx} + \frac{a_0}{2} + \sum_{m=-1}^{-\infty}\frac{a_m - ib_m}{2}e^{\omega imx}
\end{aligned}$$

To be able to simplify this expression of three terms into one summation, we need to prove, when $m = 0$:

$$\frac{a_m - ib_m}{2}e^{\omega imx} = \frac{a_0}{2}$$

$$\because b_0 = \frac{1}{L}\int_{-L}^{L} f(x)\sin\left(\frac{0\pi x}{L}\right)\,dx = 0 \text{ and } a_0 = \frac{1}{L}\int_{-L}^{L} f(x)\cos\left(\frac{0\pi x}{L}\right)\,dx \neq 0$$
$$\therefore \frac{a_0}{2} \text{ satisfies the expression for } m = 0$$

Hence, we can write the entire expression as one term, factoring out the fraction term as a complex number coefficient c_m:

$$\sum_{m=-\infty}^{\infty} c_m e^{\omega imx} \qquad \text{where } c_m = \frac{a_m - ib_m}{2} \tag{2.11}$$

We can prove that, for a real function $f(x)$, the sum is real. First, we establish that, per the definition of complex numbers, $c_{-m} = \overline{c_m}$, where n is positive. For each n, we can group $c_m e^{\omega imx}$ with $c_{-m} e^{-\omega imx}$, therefore:

$$c_m e^{\omega imx} + c_{-m} e^{-\omega imx} = c_m e^{-\omega imx} + \overline{c_m} e^{\overline{\omega imx}} = 2\text{Re}(c_m e^{\omega imx})$$

By substituting in known equations 2.7 and 2.8 into the expression for c_m, we can find an expression for c_m:

$$c_m = \frac{1}{2L}\int_{-L}^{L} f(x)\cos(mx)\,dx - \frac{1}{2L}\int_{-L}^{L} f(x)\sin(mx)\,dx$$
$$\implies c_m = \frac{1}{2L}\int_{-L}^{L} f(x)\left[\cos(mx) - \sin(mx)\right]$$

$$\because e^{-\omega imx} = \cos(mx) - i\sin(mx$$
$$\therefore c_m = \frac{1}{2L}\int_{-L}^{L} e^{-\omega imx} f(x)\,dx \tag{2.12}$$

The zeroeth Fourier coefficient is equal to the average value of the function over its period:

$$c_0 = \frac{1}{2L}\int_{-L}^{L} f(x)\,dx \tag{2.13}$$

Hence, we have arrived at formulae to calculate the coefficients c_0 and c_m of the complex Fourier series of a function $f(x)$. These coefficients represent the magnitude present in $f(x)$ of each of the frequencies corresponding to e^{wimx}.

In Appendix A.2, we confirm that using the complex coefficient form of the Fourier series gives the result for the square function calculated in Section A.2.

3 The Fourier Transform

3.1 Deriving the continuous Fourier transform

So far, we have defined Fourier series as expansions of periodic functions, and found expressions in equation 2.13 and equation 2.12 for the coefficients of this series. The goal of the **Fourier transform** is to express any function, including a non-periodic one, in terms of its Fourier series coefficients.

To do this, we can view a non-periodic function $f(x)$ as the limiting case of a periodic function as its period $2L$ approaches infinity. Alternatively, this can be expressed as the limit as ω approaches 0:

$$\because \lim_{L\to\infty} \omega = \lim_{L\to\infty} \frac{\pi}{L} = 0$$
$$\therefore \lim_{L\to\infty} = \lim_{\omega\to 0}$$

So we can write the limit of the Fourier series sum as the period $2L$ approaches infinity as:

$$f(x) = \lim_{\omega\to 0} \sum_{m=-\infty}^{\infty} c_m e^{\omega i m x}$$

Now, we can substitute in the expression for c_m from equation 2.12.

$$f(x) = \lim_{\omega\to 0} \sum_{m=-\infty}^{\infty} \left[\left(e^{\omega i m x}\right) \times \left(\frac{1}{2L} \int_{-L}^{L} e^{-\omega i m x} f(x)\, dx \right) \right]$$

Since $\omega = \frac{\pi}{L}$, $L = \frac{\pi}{\omega}$. The limits $\pm L$ can thus be expressed as $\pm\frac{\pi}{\omega}$; and the fraction $\frac{1}{2L}$ can be written as $\frac{\omega}{2\pi}$. Making these substitutions in the above expression, we get:

$$f(x) = \lim_{\omega\to 0} \sum_{m=-\infty}^{\infty} \omega \frac{1}{2\pi} \left[\left(e^{\omega i m x}\right) \times \left(\int_{-\pi/\omega}^{\pi/\omega} e^{-\omega i m x} f(x)\, dx \right) \right]$$

In the limit, this sum becomes a Riemann integral with respect to ω.

$$= \int_{-\infty}^{\infty} \frac{1}{2\pi} \left[e^{\omega imx} \int_{-\pi/\omega}^{\pi/\omega} e^{-\omega imx} f(x)\, dx \right] d\omega$$
$$= \int_{-\infty}^{\infty} \frac{1}{2\pi} \left[\int_{-\infty}^{\infty} f(x) e^{-\omega imx}\, dx \right] e^{\omega imx}\, dw$$
$$\text{(The limits of integration } \pm\tfrac{\pi}{\omega} \to \pm\infty \text{ as } \omega \to 0\text{)}$$

Now, the expression inside the square brackets can be seen as a function of m, which, when evaluated for different values of m, gives the complex Fourier series coefficients of $f(x)$. Thus, it can be thought as the continuous Fourier transform[6] $\hat{f}(m)$ of $f(x)$, formally defined by equation 3.1.

$$\hat{f}(m) = \int_{-\infty}^{\infty} f(x) e^{-\omega imx}\, dx \tag{3.1}$$

Each of the values of $\hat{f}(m)$ represent the magnitude of a sinusoid present in $f(x)$ with a frequency equal to either the fundamental frequency of $f(x)$ or its harmonics.

3.2 The discrete Fourier transform

3.2.1 Modifying the continuous Fourier transform

So far, we have dealt with cases where we are trying to approximate a continuous periodic function in terms of a Fourier series. However, with a real-world *discrete* audio signal, we only have access to a finite sequence of samples separated by a fix interval. This can be thought of as a sequence of N individual samples f_0, f_1, $\ldots f_k$, $\ldots f_{N-1}$ being drawn at intervals of time Δt from a continuous source function $f(t)$.

If we wish to represent the source function in terms of a Fourier series, equation 3.1 implies that the Fourier coefficients $\hat{f}_0$, $\hat{f}_1$, $\ldots \hat{f}_k$,$\ldots \hat{f}_{N-1}$, would be given by evaluating $\int_{-\infty}^{\infty} f(t) e^{-\omega imt}\, dt$. However, if we are to calculate these values for this discrete sampled signal, we must modify this expression. Since the signal only exists at the sampled points, i.e., between $t = 0$ and $t = (N-1)\Delta t$, these values must be the limits of the integral.

6. Brunton, Steve. *Data Driven Science and Engineering.* Cambridge University Press, 2019.

Additionally, each discrete sample of the signal can be viewed as an impulse[7] with an area equal to the magnitude of f_k. Hence, we arrive at the following expression for extracting Fourier coefficients to represent the signal:

$$\begin{aligned}\hat{f}(m) &= \int_0^{(N-1)\Delta t} f(t)e^{-\omega imt}\,dt \\ &= f_0 e^{-\omega im0} + f_1 e^{-\omega im\Delta t} + \cdots + f_k e^{-\omega imk\Delta t} + \ldots f_{N-1}e^{-\omega im(N-1)\Delta t} \\ &= \sum_{k=0}^{N-1} f_k e^{-\omega imk\Delta t}\end{aligned}$$

ω was originally defined as the constant $\frac{\pi}{L}$, where $2L$ was the fundamental period of the original function. Now, when dealing with discrete samples f_k, we can assume that they collectively comprise one period from the unknown original continuous function $f(t)$, i.e., $2L = N\Delta t$ for for $f(t)$. Hence, $\omega = \frac{2\pi}{N\Delta t}$. Then, the m-th Fourier coefficient $\hat{f}_m$ is equal to

$$\hat{f}_m = \sum_{k=0}^{N-1} f_k e^{-i\frac{2\pi}{N}mk} \tag{3.2}$$

Since we wish to extract a finite number of frequencies, we must decide the range of values of m to use. The value of the exponential term will repeat for $m > N$, since, for $m = N + l$,

$$e^{\frac{-2\pi i(N+l)k}{N}} = e^{\frac{-2\pi iNk}{N}} \times e^{\frac{-2\pi ilk}{N}} = 1 \times e^{\frac{-2\pi ilk}{N}}$$

Therefore, the sum for $m > N$ will return the same value as for a previous corresponding value of m. We thus only use values of m from 0 to $N - 1$.

3.2.2 Optimising the DFT computation

The periodic nature of the complex exponential $e^{\frac{-2\pi i}{N}}mk$ implies that its value will repeat for some combinations of m and k. We can thus try to search for specific patterns in its values to make the DFT calculation more efficient. For simplicity, we define $W_N =$

7. Delgutte, Bertrand and Julie Greenberg. "Chapter 4 - The Discrete Fourier Transform". HST.582J Biomedical Signal and Image Processing. *Harvard-MIT Division of Health Sciences and Technology.* 1999.

$e^{\frac{-2\pi i}{N}}$, such that the sum from equation 3.2 becomes

$$\sum_{k=0}^{N-1} f_k W_N^{mk}$$

We can begin by investigating even powers of W_N, starting with its square. W_N^2 can be written as $e^{\frac{-4\pi i}{N}} = e^{\frac{-2\pi i}{N/2}} = W_{N/2}$. Therefore, any power n of $W_{N/2}$ is equal to an even power $2n$ of W_N, i.e., $W_{N/2}^n = W_N^{2n}$. In general, $W_N^{2nl} = W_{N/2}^{nl}$.

Similar attempts can be made to derive patterns for odd powers of W_N, which are of the form W_N^{2n+1}. This can be rewritten as $W_N^{2n} \times W_N = W_{N/2}^n \times W_N$. In general, $W_N^{(2n+1)l} = W_{N/2}^{nl} \times W_N^l$.

The radix-2 Cooley-Tukey algorithm[8] makes use of these properties to make the process of calculating the DFT more efficient. Assuming that the number of samples is divisible by two, i.e., N is even, its first step involves the splitting of the DFT, where the single sum defining $\hat{f}_m$ is split into two sums over the even indices $0, 2, 4, \ldots$ and odd indices $1, 3, 5 \ldots$.

$$\begin{aligned}\hat{f}_m &= \sum_{k=0}^{N-1} f_k W_N^{mk} \\ &= \sum_{k=0}^{\frac{N}{2}-1} f_{2k} W_N^{(2k)m} + \sum_{k=0}^{\frac{N}{2}-1} f_{2k+1} W_N^{(2k+1)m}\end{aligned}$$

Using the above-derived properties of W, this becomes

$$\begin{aligned}&\sum_{k=0}^{\frac{N}{2}-1} f_{2k} W_{N/2}^{km} + \sum_{k=0}^{\frac{N}{2}-1} f_{2k} W_{N/2}^{km} \times W_N^m \\ &= \sum_{k=0}^{\frac{N}{2}-1} f_{2k} W_{N/2}^{km} + W_N^m \sum_{k=0}^{\frac{N}{2}-1} f_{2k} W_{N/2}^{km}\end{aligned}$$

The sums above are equivalent to two DFTs of order $N/2$, one on even-indexed samples and one on odd-indexed samples. Let E_m represent the DFT of the even-indexed inputs f_{2k} and O_m the DFT of the odd-indexed inputs f_{2k+1}. Then, it appears as though $\hat{f}_m = E_m + W_N^m O_m$, i.e., the coefficients in a DFT of length N can be calculated using

8. Cooley, James W, and John W Tukey. "An algorithm for the machine calculation of complex Fourier series". *Mathematics of computation*, vol. 19, no. 90, 1965, pp. 297–30.

the corresponding coefficients in the two smaller DFTs. This can be done for $m = 0$ till $N/2 - 1$. However, for $m = N/2$ to $m = N$, there are no corresponding values in E_M and O_m, since they come from DFTs of data of order $N/2$, which are thus themselves of order $N/2$.

To calculate the second half of the coefficients for $m = N/2$ to $m = N$, we can investigate how the values of the E_m and O_m behave beyond $N/2$. We can do so by checking if the spectrum of coefficients returned by the DFT, as defined in equation 3.2, is periodic in N:

$$\begin{aligned}
\hat{f}_{m+N} &= \sum_{k=0}^{N-1} f_k e^{-i\frac{2\pi}{N}(m+N)k} \\
&= (\sum_{k=0}^{N-1} f_k e^{-i\frac{2\pi}{N}mk})e^{-2\pi ik} \\
&= (\hat{f}_m)e^{-2\pi ik} \\
&= \hat{f}_m \times 1 = \hat{f}_m
\end{aligned}$$

Thus, the DFT spectrum of order N is periodic in N. The values of the spectrum of the two smaller DFTs of order $N/2$ for $m = N/2$ to $m = N$ will be equal to the corresponding value at each $m - N/2$. Hence, for $m = N/2$ to $m = N$, $\hat{f}_m = E_{m+N/2} + W_N^m O_{m+N/2}$. We can therefore compute a DFT of a sequence f_n of order N with the below steps[9].

1. Separate f_n into two sequences of its even-indexed samples and odd-indexed samples
2. Compute the DFTs E_m and O_m of these sequences separately
3. The DFT $\hat{f}_m$ of the original sequence f_n is then obtained as:

$$\hat{f}_m = \begin{cases} E_m + W_N^m O_m & 0 \leq m < \frac{N}{2} \\ E_{m+N/2} + W_N^m O_{m+N/2} & \frac{N}{2} \leq m < N \end{cases}$$

However, we can also divide the DFTs in step 2 by the above process. Thus, a DFT of order N is recursively broken up by the radix-2 algorithm into DFTs of order $N/2$, $N/4$, $N/8$... all the way to DFTs of length 2; under the assumption that N is a power of two.

9. Demaine, Erik. "Divide & Conquer: FFT - MIT 6.046J Design and Analysis of Algorithms", uploaded by MIT OpenCourseWare, 5 Mar 2016, youtu.be/iTMn0Kt18tg. Accessed 11 November 2023.

With this approach, the DFT can be divided into $\log_2 N$ stages, each of which contain $\frac{N}{2} + \frac{N}{2}$ computations, therefore reducing the number of calculations required to $N \log_2 N$, compared to the N^2 of the naive approach in Section 3.2.1[10]. This technique is referred to as the **Fast Fourier Transform** (FFT) algorithm for the rest of the essay, and all further DFT calculations are assumed to be done using the FFT algorithm.

3.2.3 Zero padding

One way to use the FFT algorithm on signals for which the length N is not a power of two would be to perform zero padding[11]: appending zero-valued samples to the end of the signal until its new length M is equal to a power of two. Let the original signal be f_k. Then, g_k is defined as:

$$g_k = \begin{cases} f_k & 0 \leq k \leq N-1 \\ 0 & N \leq k \leq M-1 \end{cases}$$

The FFT of g_k is then taken in place of f_k.

3.2.4 Evaluating the efficiency of the FFT

We can attempt to evaluate the extent to which the theoretical efficiency gain predicted in Section 3.2.2 is true in practice by comparing the time taken for a computer implementation of the naive DFT for segments of an audio signal with the time taken using the FFT, applying zero padding where necessary. 24 samples of length $500 + 500l$ for $l = 0$ to $l = 23$ from a test audio signal are tested. The results of this experiment are plotted in Fig. 3, and fully listed in Appendix A.4.1. The code used is recorded in Appendix A.5. The quadratic function $y = ax^2 + bx + c$ is fitted to the times taken with the naive DFT approach, and the 'linearithmic'[12] function $y = ax\log_2 x + b$ is fitted to the times taken with the FFT approach.

10. Brunton, Steven. "ME565 Lecture 16: Discrete Fourier Transforms (DFT)", 28 Apr 2016, youtu.be/KTj1YgeN2sY. Accessed 11 November 2024.
11. Rowell, Derek. 2.161 "Signal Processing: Continuous and Discrete - Lecture 11". 2008, Massachusetts Institute of Technology.
12. Segdewick, Robert, et al. *Introduction to Programming in Python*. Pearson Education, Inc. 2015.

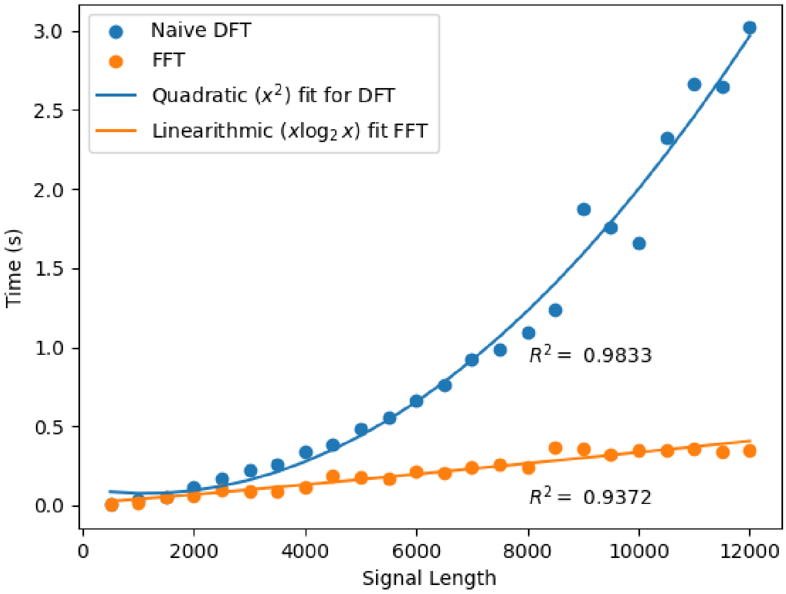


Figure 3: Comparison of time taken to execute DFT and FFT for 19 signals of increasing lengths.

High values of the coefficient of determination R^2, both greater than 0.9, indicate a very strong fit[13] for both sets of data. Thus, the theoretical efficiency gain of the FFT holds true in practice, even with the use of zero padding.

4 Comparing signals using Fourier transforms

4.1 Short-time Fourier transforms

The spectrum returned by the DFT provides information about the frequencies present in an entire signal. However, it does not reflect changes in frequency content over time. One way to obtain localised frequency information would be to repeatedly take the DFTs of short **frames** of samples from the signal. This is referred to as the **Short-Time Fourier Transform**, or STFT[14].

Two parameters controlling the STFT are the frame length N_F, which is defined as the number of samples present in each frame; and the hop length N_H, which controls the gap

13. Pardoe, Iain et al. "2.7 - Coefficient of Determination and Correlation Examples". *STAT 462: Applied Regression Analysis*. Pennsylvania State University Eberly College of Science. 2018. online.stat.psu.edu/stat462/node/97. Accessed 10 Jan 2024.
14. Mcfee, Brian. Digital Signals Theory. New York University, 2020.

between the beginning of one frame and the next[15]. The number of frames K produced for a source signal of length N is then:

$$K = 1 + \frac{N - N_F}{N_H}$$

Since it is unlikely that the periods of all signals within the original audio will be captured within the boundaries of each frame, a spectral leakage[16] effect will likely to be introduced, wherein there will be artificially-introduced abrupt discontinuities at the ends of frames, thus distorting the frequencies produced by the STFT. A *windowing technique*, wherein a frame is multiplied sample-wise with a function of equal length that tapers towards its ends[17] may be used to reduce the amplitude of the discontinuities.

One such function is a raised cosine of amplitude 0.5. In the context of windowing, this is formally referred to as a **Hann window** [18], and is plotted in Fig. 4.

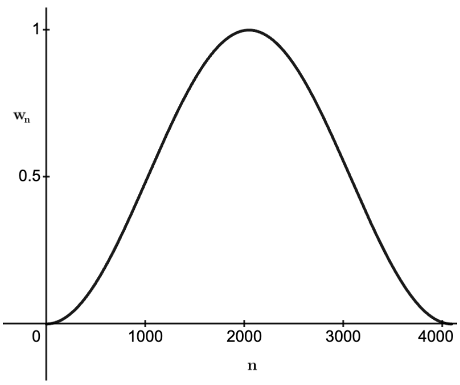


Figure 4: Sample Hann window of length $N_F = 4096$

Mathematically, a discrete-valued Hann window w_n of N_F samples is defined by equation 4.1.

$$w_n = \begin{cases} \frac{1}{2}\left[1 - \cos\left(\frac{2\pi n}{N_F}\right)\right] = \sin^2\left(\frac{\pi n}{N_F}\right) & 0 \leq n < N_F \\ 0 & \text{otherwise} \end{cases} \tag{4.1}$$

Finally, the m-th Fourier coefficient of the a-th STFT frame (a = 0, 1, 2, 3, ... K - 1)

15. Leiber, Maxime, et al. "Differentiable short-time Fourier transform with respect to the hop length". *2023 IEEE Statistical Signal Processing Workshop (SSP)*. IEEE, 2023.
16. Lyon, Douglas A. "The discrete fourier transform, part 4: spectral leakage." *Journal of object technology 8.7*, 2009.
17. Stern, Richard. "Notes on FIR Filter Design Using Window Functions". Signals and Systems 18-396. *Carnegie Mellon University Department of Electrical and Computer Engineering*. Spring 2009.
18. Ibid., 17

of width N_F, now with a constant Hann window of fixed length N_F, can be expressed as:

$$\hat{f}(a,m) = \sum_{k=aN_F}^{aN_F+N_F} (f_k w_{[k-(a-1)N_F]}) W_{N_F}^{mk} \tag{4.2}$$

All uses of the STFT in Section 4.2.1 use the above definition.

4.2 Developing a recognition system

4.2.1 System overview

Given a target audio sequence of samples of arbitrary length, we wish to correctly identify the signal in a database that the target audio is most likely to be drawn from, if any.

Based on previous[19] methods[20], we develop the following simplified process to represent an audio track:

1. The track is trimmed until its number of samples is an integral multiple of N_F.

2. As per equation 4.2, STFTs of all frames in the track are extracted, resulting in a spectrum of complex coefficients corresponding to each of the frequencies present in the frame.

3. In each frame, we divide all the frequencies into 17 groups, based on which of the following intervals (in Hz) they fall into: $[40, 80)$, $[80, 120)$, $[120, 180)$, $[180, 300)$, $[300, 500)$, $[500, 800)$, $[800, 1200)$, $[1200, 1800)$, $[1800, 2500)$, $[40, 80)$, $[80, 120)$, $[120, 180)$, $[180, 300)$, $[300, 500)$, $[500, 800)$, $[800, 1200)$, and $[1200, 1800)$.

4. For each group, we store the frequency with the highest real component in its corresponding STFT complex coefficient. Concatenating the raw values of these 17 frequencies together gives a 'fingerprint' of that frame, which is stored alongside the timestamp of the frame within the track.

19. Rijn, Roy van. "Creating Shazam in Java". *royvanjrin.com:: blog of a programmer.* 1 June, 2010. www.royvanrijn.com/blog/2010/06/creating-shazam-in-java. Accessed 20 November 2023.
20. Wang, Avery. "An industrial strength audio search algorithm." *ISMIR.* 27 October, 2003. princeton.edu/ cuff/ele201/files/Wang03-shazam.pdf. Accessed 20 November 2023.

5. Steps 3 and 4 are repeated for all frames, giving a list of fingerprint-timestamp pairs.

This fingerprinting process is conducted for all tracks in our database, and the output is stored.

To recognise a target audio track, we extract target fingerprints by the same process. Then, the database track that is considered to be a match is the one which contains the most fingerprints that are:

1. Identical to target fingerprints
2. Separated from their corresponding target fingerprints by a constant time, i.e, subtracting the timestamp of a target fingerprints from the timestamp of the corresponding database fingerprints always results in the same constant.

If no identical fingerprints are found, no match is returned.

4.2.2 Setting STFT parameters

We test this system with a database of 64 audio signals from the Open Music Archive[21], all of which are sampled at frequency $F_s = 44100$Hz. Constraints can now be placed on N_F and N_H (as defined in Section 4.1).

- To avoid zero padding, a value of $N_F = 2^x, x \in \mathbb{Z}^+$ should be selected.
- To identify audio, we must ensure that all samples from signals are used, since the target audio may be drawn from any arbitrary point in the source signal. Hence, $N_H < N_F$ must be satisfied.
- For evaluating signals produced by humans, we should ensure that we can capture the lowest frequency audible to humans, which is 20 Hz[22]. The minimum width of a window containing a 20 Hz signal is $1/20 = 0.05$ seconds. A window with width N_F has duration in seconds of $\frac{N_F}{F_s}$. $\frac{N_F}{F_s} \geq 0.05 \implies N_F \geq 2,205$ for $F_s = 44100$.

21. Simpson, Eileen, and Ben White. "Open Music Archive". 2005-24. www.openmusicarchive.org/index.php. Accessed 20 January 2024.
22. Ellinger, John. "Unit 1 Sound Basics". *MUSC 101: Music Fundamentals.* Carleton College. Spring 2012. people.carleton.edu/ jellinge/m101s12/Pages/01/01SoundBasics.html. Accessed 20 January 2024.

- To ensure that each window progresses consistently, N_H should be a constant fraction of N_F, i.e., $N_H = \frac{N_F}{x}$ for $x \in \mathbb{Z}^+$.

Because of the relatively limited range of frequencies required while dealing with human-generated signals, we set N_F to the lowest possible value that satisfies the above constraints, i.e., 4096. For the purpose of audio recognition, we do not require granular detail about the change in frequency content over time, rather, we are focused on the presence of certain sound events. Thus, N_H is set to the maximum possible corresponding value given the choice of N_F, which is 2048.

4.2.3 Methods used to evaluate recognition system

We test the ability of the system to recognise a single target audio by varying two parameters: length of target audio required and degree of background noise. To test the length of audio required, clips of lengths 0.1, 0.5, 1.0, 2.5, 5, 7.5, and 10 seconds are tested. To test the resistance of the system to background noise, this essay adds 'white noise' to the target audio, by adding values in random order from a normal distribution as defined by the probability density equation[23] below.

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

where $\mu = 0$, so that the distribution of positive and negative values symmetric around 0 is returned; and $\sigma = h \times \bar{f(t)}$, where h is a constant and $\bar{f(t)}$ is the mean noise level of the target audio $f(t)$.

7 values of h are used in testing: 0 (resulting in no background noise), 0.5, 1.0, 2.5, 5.0, 7.5, and 10.0. The resulting distributions are plotted for an example track in Fig. 5.

23. Leo, William R. *Techniques for Nuclear and Particle Physics Experiments.* 2nd Revised Edition. Springer-Verlag Berlin Heidelberg GmbH, 1994.

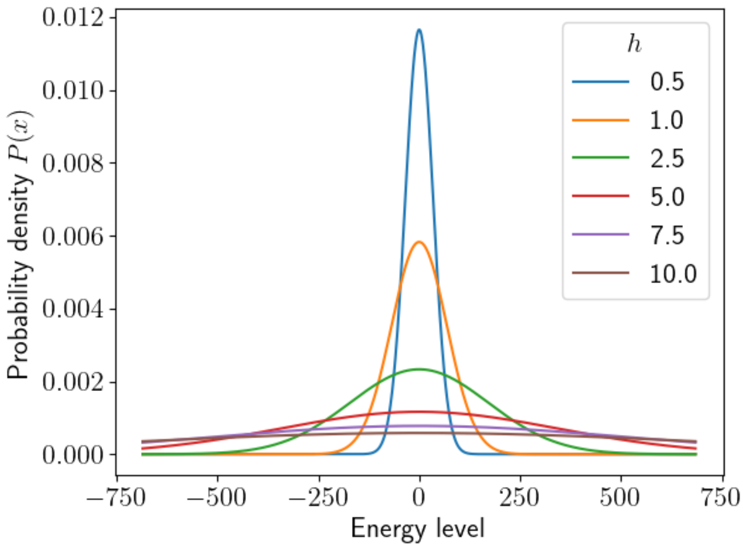
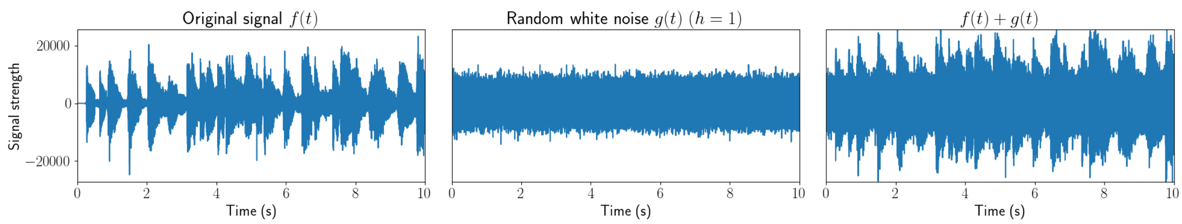
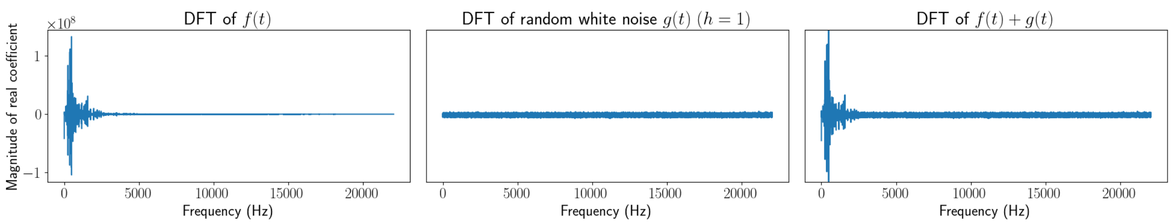


Figure 5: Probability distributions generated for a 10-second test sample with mean energy $\bar{f(t)} = 2656.8$. As h and the corresponding σ increases, the probability of higher-energy noise being added to samples increases.

The effect of adding white noise with $h = 1$ on a test track, both on the time-domain representation and discrete Fourier spectrum representation, is illustrated in Fig. 6.



(a) Effect of adding randomly-ordered white noise on amplitude of signal. As expected, the strength of the signal appears to change across its entire duration.



(b) Effect of adding white noise on DFT frequency spectrum of signal. The DFT of the noise appears to be evenly distributed across all frequency values. For $h = 1$, the frequencies which originally had the highest magnitude continue to be dominant. As h increases, it can be expected that their significance in comparison to all other frequencies will decrease.

Figure 6: Effect of adding white noise $g(t), h = 1$ on audio signal $f(t)$

It appears that adding random noise to all points in the time domain, i.e., adding $g(t)$

to $f(t)$, results in an addition to the magnitude of all frequencies in the frequency domain (a formal proof of this property is given in Appendix A.3). Since the 'fingerprints' used by our system consist of the frequencies with the highest magnitude in each of the chosen groups (as per Section 4.2.1), a change in the magnitudes of all frequencies is likely to change the fingerprint of the target audio, and therefore make accurate identification more difficult.

4.2.4 Results of testing

We test $7 \times 7 = 49$ combinations of parameters, resulting in 49 rounds of testing. In each round, for every track in the database, we extract a random clip of the desired length, add white noise, and apply the recognition system. We save the accuracy of the round, calculated the 64 tracks correctly recognised. The results of these experiments are provided in Table 1, and plotted in Fig. 7.

Audio length (s)	**Accuracy (%)**						
$h \rightarrow$	**0.0**	**0.5**	**1.0**	**2.5**	**5.0**	**7.5**	**10.0**
0.1	23.4	6.25	4.69	1.56	0.0	0.0	0.0
0.5	73.4	43.8	31.2	4.69	12.5	0.0	0.0
1.0	82.8	59.4	45.3	18.8	15.6	0.0	0.0
2.5	93.8	76.6	70.3	31.2	4.69	3.12	3.12
5.0	93.8	92.2	65.6	37.5	17.2	1.56	0.0
7.5	96.9	85.9	76.6	43.8	21.9	4.69	3.12
10.0	96.9	93.8	82.8	45.3	32.8	9.38	1.56

Table 1: Evaluation of recognition performance. Each row corresponds to the results for one target audio length, and each column represents one value of h tested. Accuracy is reported as a percentage to 3 significant figures.

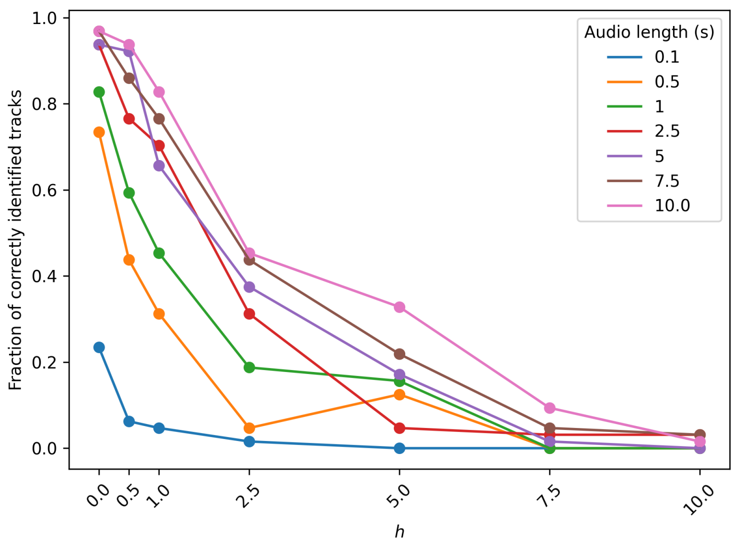


Figure 7: Plot of values from Table 1. A maximum accuracy of 96.875% is achieved. For length > 0.5 seconds, random chance recognition $\frac{1}{64} = 1.56\%$ is outperformed for all $h < 7.5$.

We notice that, while the length of the target audio is consistently related to accuracy, even with only 2.5 seconds of audio, 70%+ accuracy is achieved for $h \leq 1.0$. The system appears to only deteriorate significantly for $h > 1.0$, and it appears that it is only for $h \geq 7.5$ that the dominant frequencies in each frame are changed too significantly for the recognition system to function.

5 Conclusion

In this essay, the use of Fourier transforms to represent signals in the frequency domain has been examined; and an audio recognition method using these techniques was developed, which, as per Section 4.2.3, was highly effective. We analyse below some of the potential reasons for the success of this method.

1. For every frame, only 17 frequency values were analysed; compared to the original $N_F = 4096$ time-domain values. Not only was this computationally and spatially efficient, it also meant that a relatively low number of matching values were required for recognition.

2. By focusing on the frequency domain, the system was resistant to irregular noise

to a large degree, since the most significant frequencies for two signals can be the same even if the signals are non-identical in the time domain.

3. STFTs allowed us to extract time-localised frequency information, so that clips as short as 2.5 seconds could be recognised.

Thus, as hypothesised in Section 1, a Fourier-transform based approach resulted in efficient *representation* as well as accurate *recognition* of signals.

Techniques in this study may be further developed. Aside from radix-2, other methods, such as those using prime factor properties, may be employed to further optimise the computation of the DFT. Additionally, the frequency groups used for fingerprint extraction may be further optimised by researching the kind of information revealed about an audio signal at different frequency levels.

In the future, I would be interested in seeing how this system can be extended to recognise 2-dimensional signals such as images. More generally, I am intrigued in how Fourier analysis can be used to identify the most important frequency components of a signal, such that they can be efficiently compressed and stored; or so that unwanted background noise can be removed.

Works cited

Brunton, Steve. *Data Driven Science and Engineering.* Cambridge University Press, 2019.

Brunton, Steven. "ME565 Lecture 16: Discrete Fourier Transforms (DFT)", 28 Apr 2016, youtu.be/KTj1YgeN2sY. Accessed 11 November 2024.

Cooley, James W, and John W Tukey. "An algorithm for the machine calculation of complex Fourier series". *Mathematics of computation*, vol. 19, no. 90, 1965, pp. 297–30.

Delgutte, Bertrand and Julie Greenberg. "Chapter 4 - The Discrete Fourier Transform". HST.582J Biomedical Signal and Image Processing. *Harvard-MIT Division of Health Sciences and Technology.* 1999.

Demaine, Erik. "Divide & Conquer: FFT - MIT 6.046J Design and Analysis of Algorithms", uploaded by MIT OpenCourseWare, 5 Mar 2016, youtu.be/iTMn0Kt18tg. Accessed 11 November 2023.

Ellinger, John. "Unit 1 Sound Basics". *MUSC 101: Music Fundamentals.* Carleton College. Spring 2012. people.carleton.edu/ jellinge/m101s12/Pages/01/01SoundBasics.html. Accessed 20 January 2024.

Leiber, Maxime, et al. "Differentiable short-time Fourier transform with respect to the hop length". *2023 IEEE Statistical Signal Processing Workshop (SSP).* IEEE, 2023.

Leo, William R. *Techniques for Nuclear and Particle Physics Experiments.* 2nd Revised Edition. Springer-Verlag Berlin Heidelberg GmbH, 1994.

Lyon, Douglas A. "The discrete fourier transform, part 4: spectral leakage." *Journal of object technology 8.7*, 2009.

Mattuck, Arthur et al. "Fourier Series: Definitions and Coefficients - 18.03SC Differential Equations". 18.03SC Differential Equations. Massachusetts Institute of Technology.

Mcfee, Brian. Digital Signals Theory. New York University, 2020.

Osgood, Brad. *Lecture Notes for EE261: The Fourier Transform and its Applications.* Stanford University, 2007.

Pardoe, Iain et al. “2.7 - Coefficient of Determination and Correlation Examples”. *STAT 462: Applied Regression Analysis.* Pennsylvania State University Eberly College of Science. 2018. online.stat.psu.edu/stat462/node/97. Accessed 10 Jan 2024.

Rijn, Roy van. “Creating Shazam in Java”. *royvanjrin.com:: blog of a programmer.* 1 June, 2010. www.royvanrijn.com/blog/2010/06/creating-shazam-in-java. Accessed 20 November 2023.

Rowell, Derek. 2.161 “Signal Processing: Continuous and Discrete - Lecture 11”. 2008, Massachusetts Institute of Technology.

Segdewick, Robert, et al. *Introduction to Programming in Python.* Pearson Education, Inc. 2015.

Simpson, Eileen, and Ben White. “Open Music Archive”. 2005-24. www.openmusicarchive.org/index.php. Accessed 20 January 2024.

Smith, Steven W. *The Scientist and Engineer’s guide to digitial signal processing.* California Technical Publishing. 1998.

Stern, Richard. “Notes on FIR Filter Design Using Window Functions”. Signals and Systems 18-396. *Carnegie Mellon University Department of Electrical and Computer Engineering.* Spring 2009.

Strang, Gilbert. *Computational Science and Engineering.* Cambridge University Press, 2007.

Wang, Avery. “An industrial strength audio search algorithm.” *ISMIR.* 27 October, 2003. princeton.edu/ cuff/ele201/files/Wang03-shazam.pdf. Accessed 20 November 2023.

“Numeracy, Maths and Statistics - Academic Skills Kit - Odd and Even Functions”. www.ncl.ac.uk/ webtemplate/ask-assets/external/maths-resources/core-mathematics/pure-maths/functions/odd-and- even-functions.html. Newcastle University, 2008. Accessed 20 June 2023.

A Appendices

A.1 Proofs for vanishing integrals of sinusoids

The proof for equation 2.2 is as follows:

$$\begin{aligned}\int_{-L}^{L} \sin(\frac{n\pi x}{L}) &= \frac{-L}{n\pi} cos\left(\frac{n\pi x}{L}\right)\Bigg|_{-L}^{L} \\ &= \frac{L}{n\pi}\left[-\cos(n\pi) + \cos(n\pi)\right] \\ &= 0\end{aligned}$$

The proof for equation 2.3 is as follows:

$$\int_{-L}^{L} \sin(\frac{n\pi x}{L}) \cdot \cos(\frac{m\pi x}{L})\, dx = 0$$
$$= \frac{1}{2}\int_{-L}^{L} 2\sin(\frac{n\pi x}{L}) \cdot \cos(\frac{m\pi x}{L})\, dx$$
$$= \frac{1}{2}\int_{-L}^{L} \sin\left(\frac{n\pi x}{L} + \frac{m\pi x}{L}\right) + \sin\left(\frac{n\pi x}{L} - \frac{m\pi x}{L}\right)\, dx$$

Since $2\sin a \cos b = \sin(a+b) + \sin(a-b)$

$$= \frac{1}{2}\int_{-L}^{L} \sin\left(\frac{n\pi x}{L} + \frac{m\pi x}{L}\right) + \frac{1}{2}\int_{-L}^{L} \sin\left(\frac{n\pi x}{L} - \frac{m\pi x}{L}\right)$$

We solve the first integral as follows. Let u $= \frac{n\pi x}{L} + \frac{m\pi x}{L}$. Then, $du = \frac{n\pi}{L} + \frac{m\pi}{L}\, dx$

$$\frac{1}{2} \cdot \frac{1}{\frac{n\pi}{L} + \frac{m\pi}{L}} \int_{-L}^{L} \sin(u)\, du$$
$$= \frac{L}{2(n\pi + m\pi)} \cdot \left[\cos(\frac{n\pi x}{L} + \frac{m\pi x}{L})\Bigg|_{-L}^{L}\right]$$
$$= \frac{L}{2(n\pi - m\pi)} \cdot \left[\cos\left(n\pi + m\pi\right) - \cos\left(-n\pi + m\pi\right)\right]$$

Since $\cos(-a) = \cos(a)$, this is equal to:

$$= \frac{L}{2(n\pi - m\pi)} \cdot [\cos(n\pi + m\pi) - \cos(n\pi + m\pi)]$$

$$= 0$$

We solve the second integral similarly. Let u $= \frac{n\pi x}{L} - \frac{m\pi x}{L}$. Then, $du = \frac{n\pi}{L} - \frac{m\pi}{L}\, dx$

$$\frac{1}{2} \cdot \frac{1}{\frac{n\pi}{L} - \frac{m\pi}{L}} \int_{-L}^{L} \sin(u)\, du$$

$$= \frac{L}{2(n\pi - m\pi)} \cdot \left[\cos(\frac{n\pi x}{L} - \frac{m\pi x}{L}) \Big|_{-L}^{L} \right]$$

$$= \frac{L}{2(n\pi - m\pi)} \cdot [\cos(n\pi - m\pi) - \cos(-n\pi + m\pi)]$$

$$= \frac{L}{2(n\pi - m\pi)} \cdot [\cos(n\pi - m\pi) - \cos(n\pi - m\pi)]$$

$$= 0$$

Thus, the sum of the two integrals is also 0.

The proof for equation 2.4 is as follows.

When $n = m = 0$

$$\int_{-L}^{L} \sin(\frac{n\pi x}{L}) \cdot \sin(\frac{m\pi x}{L})\, dx = \int_{-L}^{L} \sin^2 0\, dx = \int_{-L}^{L} 0 = 0$$

When $n = m \neq 0$,

We know that $\sin^2 a = \frac{1 - \cos 2x}{2} = \frac{1}{2}(1 - \cos(2a))$

$$\Longrightarrow \int_{-L}^{L} \sin(\frac{n\pi x}{L}) \cdot \sin(\frac{n\pi x}{L})\, dx = \frac{1}{2} \int_{-L}^{L} 1 - \cos\left(\frac{2n\pi x}{L}\right) dx$$

$$= \frac{1}{2} \left[\int_{-L}^{L} 1\, dx - \int_{-L}^{L} \cos\left(\frac{2n\pi x}{L}\right) dx \right]$$

Let u $= \frac{2n\pi x}{L}$. Then $du = \frac{2n\pi}{L}$

$$\Longrightarrow \frac{1}{2} \left[\int_{-L}^{L} 1\, dx - \frac{L}{2n\pi} \int_{-L}^{L} \cos u\, du \right]$$

$$= \frac{1}{2}\left[x - \frac{L}{2n\pi}\sin\left(\frac{2n\pi x}{L}\right)\Bigg|_{-L}^{L}\right]$$
$$= \frac{1}{2}\left[L - \frac{L}{2n\pi}\sin\left(\frac{2n\pi L}{L}\right)\right] - \frac{1}{2}\left[-L - \frac{-L}{2n\pi}\sin\left(\frac{-2n\pi L}{L}\right)\right]$$
$$= \frac{L}{2}\left[\left(1 - \frac{1}{2n\pi}\sin(2n\pi)\right) - \left(-1 + \frac{1}{2n\pi}\sin(-2n\pi)\right)\right]$$
$$= \frac{L}{2}\left[1 + 1 - \frac{1}{2n\pi}\sin(2n\pi) - \frac{1}{2n\pi}\sin(-2n\pi)\right]$$

Since $\sin(-a) = -\sin(a),\ \sin(-2n\pi) = -\sin(2n\pi)$

$$= \frac{L}{2}\left[2 - \frac{1}{2n\pi}\sin(2n\pi) + \frac{1}{2n\pi}\sin(2n\pi)\right]$$
$$= L$$

To solve this for when $n \neq m$, we subtract the trigonometric identity

$\cos(a+b) = \cos(a)\cos(b) - \sin(a)\sin(b)$ from

$\cos(a-b) = \cos(a)\cos(b) + \sin(a)\sin(b)$

$\implies \cos(a+b) - \cos(a-b) = 2\sin(a)\sin(b)$

$\implies \sin(a)\sin(b) = \frac{1}{2}\left[\cos(a+b) - \cos(a-b)\right]$

$$\therefore \int_{-L}^{L}\sin\left(\frac{m\pi x}{L}\right)\cdot\sin\left(\frac{n\pi x}{L}\right)\,dx = \frac{1}{2}\int_{-L}^{L}\cos\left[\frac{(m+n)\pi}{L}x\right]\,dx - \frac{1}{2}\int_{-L}^{L}\cos\left[\frac{(m-n)\pi}{L}x\right]\,dx$$
$$= \frac{L}{2\pi}\left[\frac{1}{m+n}\sin\left(\frac{(m+n)\pi x}{L}\right) - \frac{1}{m-n}\sin\left(\frac{(m-n)\pi x}{L}\right)\right]_{-L}^{L}$$
$$= \frac{L}{2\pi}\cdot 0 = 0$$

Similarly, the proof for equation 2.5 is derived as follows.

When $n = m = 0$,

$$\int_{-L}^{L}\cos\left(\frac{n\pi x}{L}\right)\cdot\cos\left(\frac{m\pi x}{L}\right)\,dx = \int_{-L}^{L}\cos^2 0\,dx = \int_{-L}^{L}1\,dx = 2L$$

When $n = m \neq 0$,

$$\int_{-L}^{L}\cos\left(\frac{n\pi x}{L}\right)\cdot\cos\left(\frac{m\pi x}{L}\right)\,dx$$

$= \int_{-L}^{L} \cos^2\left(\frac{n\pi x}{L}\right) \, dx$

Since $2\cos^2 a - 1 = \cos(2a) \implies \cos^2 a = \frac{\cos 2a + 1}{2}$, this can be simplified to

$= \frac{1}{2}\int_{-L}^{L} \cos\left(\frac{2m\pi x}{L}\right) + 1 \, dx$

$= \frac{L}{4\pi m} \sin\left(\frac{2m\pi x}{L}\right)\Big|_{-L}^{L} + \frac{2L}{2}$

$= \frac{L}{4\pi m}\left[\sin(2m\pi) - \sin(-2m\pi)\right] + L$

$= L$ since $m \in \mathbb{Z}^+$

Finally, in the case where $n \neq m$, we use the below property to simplify the integral.

$$\cos a \cos b = \frac{1}{2}\left[\cos(a+b) + \cos(a-b)\right]$$

Using this, the integral becomes:

$\int_{-L}^{L} \cos\left(\frac{n\pi x}{L}\right) \cdot \cos\left(\frac{m\pi x}{L}\right) \, dx$

$= \frac{1}{2}\int_{-L}^{L} \cos\left(\frac{\pi x(m+n)}{L}\right) \, dx + \frac{1}{2}\int_{-L}^{L} \cos\left(\frac{\pi x(m-n)}{L}\right) \, dx$

$= \frac{L}{2(m+n)\pi} \sin\left[\frac{(m+n)\pi}{L}x\right]\Big|_{-L}^{L} + \frac{L}{2(m-n)\pi} \sin\left[\frac{(m-n)\pi}{L}x\right]\Big|_{-L}^{L}$

$= \frac{L}{2(m+n)\pi}\left[\sin(\pi(m+n) + \sin(\pi(m+n))\right] + \frac{L}{2(m-n)\pi}\left[\sin(\pi(m-n) + \sin(\pi(m-n))\right]$

$= \frac{L\sin(\pi(m+n))}{(m+n)\pi} + \frac{L\sin(\pi(m-n))}{(m-n)\pi}$

$\because (m+n) \in \mathbb{Z}^+, (m-n) \in \mathbb{Z}^+$ and $\sin(k\pi x) = 0$ for $k \in \mathbb{Z}^+$

$\therefore \frac{L\sin(\pi(m+n))}{(m+n)\pi} + \frac{L\sin(\pi(m-n))}{(m-n)\pi} = \frac{0}{(m-n)\pi} + \frac{0}{(m-n)\pi}$

$= 0$

A.2 Verification of result of complex Fourier series for square function

We can try to recalculate the Fourier series of the odd square wave function $f(x)$ defined in Section 2.3.3 using the expressions for c_0 and c_m. over the interval $(-\pi, \pi)$. Here, $L = \pi$, therefore, since $\omega = \frac{\pi}{L}$, $\omega = 1$. First, we try to find the general term for the coefficient c_m, as per equation 2.12:

$$
\begin{aligned}
c_m &= \frac{1}{2L}\int_{-\pi}^{\pi} e^{-\omega imx} f(x)\,dx \\
&= \frac{1}{2\pi}\int_{-\pi}^{0} e^{-imx}(-1)\,dx + \int_{0}^{\pi} e^{-imx}(1)\,dx \\
&= \frac{1}{2\pi}\left[\frac{1}{im}e^{-imx}\bigg|_{-\pi}^{0} - \frac{1}{im}e^{-imx}\bigg|_{0}^{\pi}\right] \\
&= \frac{1}{2\pi im}\left[e^{0} - e^{im\pi} - e^{-im\pi} + e^{0}\right] \\
&\because e^{-ik\pi} = e^{k\pi} \\
&\therefore = \frac{1}{\pi im}\left[1 - e^{im\pi}\right]
\end{aligned}
$$

Note that

$$
1 - e^{im\pi} = \begin{cases} 0 & m \text{ even} \\ 2 & m \text{ even} \end{cases}
$$

$$
\sum_{m \text{ odd}} \frac{2}{\pi im} e^{imx}
$$

As described in Section 2.4, we can group the positive and negative terms. Let m be a positive integer. Then, we have terms in the series for m and $-m$:

$$
\begin{aligned}
m\colon \ & \frac{2}{\pi im} e^{imx} \\
-m\colon \ & \frac{-2}{\pi im} e^{-imx}
\end{aligned}
$$

Adding them gives:

$$
\sum_{m \text{ odd}} \frac{2}{\pi im} e^{imx}
$$

$$\begin{aligned}
&\frac{2}{\pi i m}\left[\cos(mx) + i\sin(mx) - \cos(-mx) - i\sin(-mx)\right] \\
&= \frac{2}{\pi i m}\left[2i\sin(mx)\right] \qquad\qquad (\cos(-k) = -\cos(k),\ \sin(-k) = -\sin(k)) \\
&= \frac{4}{\pi m}\sin(mx)
\end{aligned}$$

We can write this only for positive odd integers $n = 2k+1$:

$$\frac{4}{\pi}\sum_{k=0}^{\infty}\frac{1}{2k+1}\sin((2k+1)x)$$

We arrive at the same expression as in equation 2.10. Thus, we show that complex coefficients of a Fourier series resolve to real coefficients for a real function.

A.3 Additivity of DFT

We wish to prove that

$$\text{DFT}(f(t) + g(t)) = \text{DFT}(f(t)) + \text{DFT}(g(t))$$

In words, we wish to prove that the DFT of the sum of discrete samples sourced from functions $f(t)$ and $g(t)$ is equal to the sum of the DFTs of the samples taken separately[24]. Assume we have N samples of both f_k and g_k, where f_k and g_k represent the k-th sample drawn from $f(t)$ and $g(t)$ separately.

$$\begin{aligned}
\text{DFT}(f(t) + g(t)) &= \sum_{k=0}^{N-1}(f_k + g_k)e^{-i\frac{2\pi}{N}mk} \\
&= \sum_{k=0}^{N-1} f_k e^{-i\frac{2\pi}{N}mk} + \sum_{k=0}^{N-1} g_k e^{-i\frac{2\pi}{N}mk} \\
&= \text{DFT}(f(t)) + \text{DFT}(g(t))
\end{aligned}$$

Hence, adding two signals in the time domain results in their frequency-domain DFT

24. Smith, Steven W. *The Scientist and Engineer's guide to digitial signal processing.* California Technical Publishing. 1998.

representation also being added together.

A.4 Data tables

A.4.1 Comparison of DFT and FFT

Below, the times taken to compute the DFT and FFT of signals of varying lengths with a computer a program are listed. A computer system with 16 gigabytes of memory and a 3228 MHz processor was used.

Length of signal	Time taken (seconds)	
	Naive DFT	FFT
500	0.0046	0.0079
1000	0.0318	0.0153
1500	0.0509	0.0549
2000	0.1193	0.0583
2500	0.1693	0.0940
3000	0.2225	0.0896
3500	0.2604	0.0874
4000	0.3383	0.1201
4500	0.3895	0.1848
5000	0.4820	0.1816
5500	0.5566	0.1726
6000	0.6637	0.2133
6500	0.7581	0.2055
7000	0.9265	0.2440
7500	0.9843	0.2577
8000	1.0912	0.2427
8500	1.2416	0.3659
9000	1.8771	0.3592
9500	1.7557	0.3238

10000	1.6624	0.3511
10500	2.3248	0.3465
11000	2.6621	0.3613
11500	2.6435	0.3397
12000	3.0212	0.3505

A.4.2 List of audio tracks used

All 64 tracks used in experiments, listed below, are sourced under a public domain licence from the Open Music Archive[25].

Track Name	Duration (seconds)	Track Name	Duration (seconds)
Ommie Wise	193.2	Oi ya nestchastay	200.67
In The Dark Flashes	197.69	Umbrellas To Mend	220.37
Titanic Blues	189.23	55 By Beatrice Dillon	88.87
Peg And Awl	184.89	April Kisses	186.51
John Hardy By Karen Gwyer	56.53	Night Latch Key Blues	185.5
Dry Bones	180.98	ATL BEAT 10	63.37
Ole Gray Beard By Karen Gwyer	64.24	Jump Steady Blues	198.53
My Name Is John Johanna	197.4	House Carpenter	419.89
Daddy Wouldnt Buy Me A Bow Wow	171.55	At The Ball Thats All	246.94
Eddies Twister	173.8	Little Bits	187.27
Struggling	163.19	One Dime Blues	167.13
Waiting For A Train	165.9	Sail Away Lady	178.36

25. Simpson, Eileen, and Ben White. "Open Music Archive". 2005-24. www.openmusicarchive.org/index.php. Accessed 20 January 2024.

Henry Lee	212.19	Six Cold Feet In The Ground	182.88
Georgia Stomp By Dj Assault	182.49	Rolls Royce Papa	174.63
Drunkards Special	201.98	Poor Me Blues	194.52
Casey Jones By Karen Gwyer	67.55	Whitehouse Blues	212.9
Charles Giteau	186.93	Willie Moore	198.64
Deep Blue Sea Blues	210.87	O Patria Mia From Aida	261.62
Evil Minded Blues	168.91	A Lazy Farmer Boy	182.91
Sugar Ba	178.21	For Months And Months And Months	170.95
Barragan Sound System - Jonny Hill	1000.07	Pinetops Blues	175.36
Frankie	207.55	The House Carpenter	198.27
The Butchers Boy	184.92	Pinetops Boogie Woogie	204.72
Mississippi Boweavil Blues	191.76	Old Dog Blue	184.4
Old Lady And The Devil	188.06	The Wagoners Lad	185.95
Ragtime Annie	192.94	King Kong Kitchie Kitchie Kimio	195.71
Wake Up Jacob	175.49	I'm Gonna Get Me A Man Thats All	184.53
Intro And Tarantelle	274.99	Don't Go Way Nobody	244.98
Frankie And Johnny By Karen Gwyer	53.45	Don't Let Your Deal Go Down Blues	171.26
The Wild Wagoner	197.43	I'm Sober Now	191.29
Butchers Boy By Karen Gwyer	193.31	Court House Blues Take 1	198.93

The Wagoners Lad By Karen Gwyer	81.29	White House Blues	216.84

A.5 Code used to evaluate DFT and FFT

Code was written using programming language Python (version 3.11).

```
def dft(x):
    N = x.shape[0]
    n = np.arange(N)
    k = n.reshape((N, 1))
    # Create the DFT matrix. j is used instead of i.
    M = np.exp(-2j * np.pi * k * n / N)
    return np.dot(M, x) # Multiply the matrix by the signal



def fft(x):
    N = len(x)
    zeros = np.zeros((2**(N-1).bit_length() - N), dtype=complex)
    x = np.concatenate((x, zeros))
    N = len(x)

    if N == 1:
        return x
    else:
        X_even = fft(x[::2])
        X_odd = fft(x[1::2])

        X = np.zeros(N, dtype=complex)
        for m in range(N):
```

```
        m_new = m if m < (N//2) else m - (N // 2)
        w_N = np.exp(-2j * np.pi * m / N)
        X[m] = (X_even[m_new]) + (np.multiply(w_N, X_odd[m_new]))
    return X
```