Exploring the mechanism of active noice cancelation

Mathematical Exploration

Leo Chai

White Oaks Secondary School

# Introduction

I have always found music inspiring, and I always carried around a Walkman with me where ever I go when I was young. I grew up in a crowded city, and loud noises in public, such as the subway station and the sidewalk, have always bothered me when I listen to music. This is why when my parents bought me my first active noise canceling (ANC) headphones when I was in middle school, I was convinced that it functioned on magic.

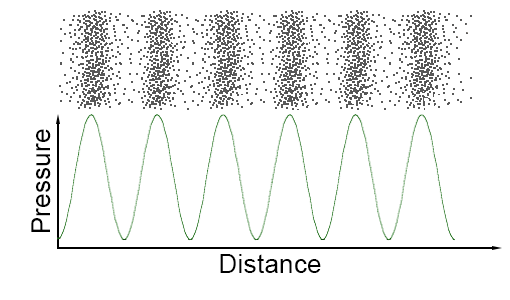
 I learned the principals which ANC function on last year in my SL physics course. Sound is a pressure wave of air with propagating numbers of air molecules as moving across the sound wave. There are points of compression created where there are more air molecules present with higher pressure, and points of rarefaction where there are fewer air molecules present with lower pressure.

Figure 1. Each dot on the top represent a air molecule, allowing the first graph to represent what a soundwave look like in real life. The second graph shows the air pressure at different points in the sound wave.

Figure one includes a graph that represents the shown sound wave in the form of a sine function. While most sound waves cannot be represented by a single sine function, this shows soundwave’s function like characteristic as each point in space only have one air pressure.

The graph in figure one denoted soundwaves by graphing distance against air pressure to better math with the visual representation of soundwave above, but this is not how soundwaves are usually recorded. Microphone record sound through the use of a diaphragm, which vibrates along the air molecule in the air as sound pass by. The microphone records the change in the physical location of the diaphragm, which is caused by the different air pressure on the soundwave passing by, allowing the computer to note down the change in air pressure over the change in time.

When two different soundwaves overlap, they interfere with each other. By viewing soundwave as a function of the change of air pressure over the change in time, it is intuitive that the result of two functions overlapping can be found by adding the two functions together. The graph shows the result of the interference, h(x), of the two functions f(x) and g(x). In the green boxes, because of that both soundwaves have the same sign, they interfere constructively, causing the amplitude of the resulting wave h(x) to be greater than either f(x) or g(x) in those regions. In boxes highlighted by red, the soundwaves interfere destructively, as their difference signs causes the two sound waves to cancel each other out. Consequentially, the resulting wave h(x) has a smaller amplitude than either f(x) or g(x) in these regions.

ANC headphones uses the interference of soundwaves to block out the noises from the outside environment. The headphone’s tight seal on the ears act as a physical barrier of sound, but there will still be some noise getting in to the headphone. The headphone then uses multiple different microphones to record the noise that was not blocked out by the headphone, and flips the signal to completely cancel it.

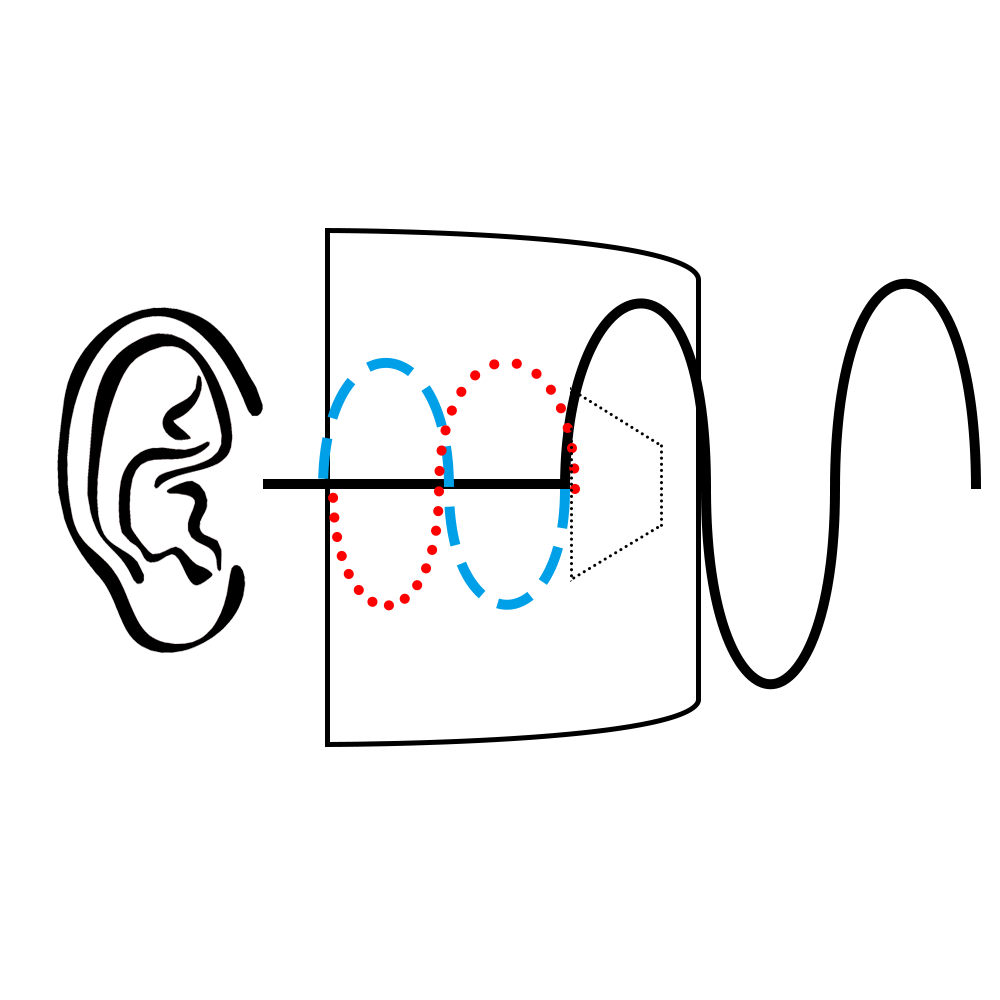


Figure 2 The black line represents the noise from outside. The blue line is the noise heard by the ears without ANC on, and the red line is the sound coming out from the speakers.

While some cheaper headphones use an analogue circuit to directly flip the signal from the microphone to cancel the sound, this type of noise cancelation does not perform as well as many would expect. Because of that high precision is required to cancel out the sound because of the high speed of sound, the uncertainties created by changes in temperature would affect the effect of the noise cancelling circuit.

Moreover, different frequencies of sound are blocked out by different amounts by the earphones themselves. Since there will always be a barrier between the microphone and the speakers, such as circuit boards even when the speaker is placed inside the earphones, the earphone needs to adjust the amplitude of different frequencies of sound.

# The Fourier Theorem

In order to be able to manipulate the amplitude of different frequencies of sound, the headphone must separate the different frequencies of sounds first. Jean-Baptiste Joseph Fourier, a French mathematician and physicist, have tackled this problem. His Fourier theorem states that any periodical function can be composed of a Fourier Series.

A Fourier Series is a summation function of an infinite amount of sine and cosine functions of different frequencies. Since the Fourier theorem is commonly regarded as an intuitive assumption in math, and its proof goes beyond the scope of this essay, I will be providing an example to show the theorem’s affect instead of going through a tedious proving process.

This graph illustrates the function . As n increases from 1 to infinity, the graph approaches the desired shape of the function which it is trying to construct. If a recording of sound was imagined as a self-repeating function, it could also be expressed as a Fourier Series. By expressing soundwaves in the form of a Fourier Series, to be able to adjust the amplitude of each frequency, the headphone will be better able to cancel out noises actively.

# Fourier Analysis

I have established that a soundwave can be represented using a Fourier Series, and now I will investigate the concept of the Fourier analysis, which transforms of the soundwave as a function of time into a function of frequencies.

## Fourier Transformation

Starting simple, human’s understanding of what defines a sinusoidal wave needs to translated into mathematics. Take the function f(x) = sin(x) for example, the amplitude of the function is just f(x) for any given point on the x axis. We determine the period of the function by first finding an x-interception of the function. We then fallow the function, as it increases above the original point, return to the height of the original point, decreases to below the original point and finally return again to the height of the original point. We then find the distance the we have traveled in the x-axis, and that value will be the period of the function.

While we can replicate the identification process exactly with the use of f’(x) and f(x), which finds the slope of the function and x-interceptions, the process can be dramatically simplified with another way of approaching the problem.

The characteristic of a positive hump directly followed by a negative hump both with the exact same shape can be taken advantage of to describe a period. The opposite nature of the humps cancels each other out when trying to find the average or the integral of the graph f(x) within the boundaries of its period. When the boundaries do not match up exactly with the period, not all of the parts above and below the x-axis with match up perfectly. Causing the average and the integral of the graph f(x) to have an absolute value greater than 0.

Starting simple, if we simplify the Fourier series to only whole number frequencies, we get:

Since the Fourier series is the sum of an infinite amount of sinusoidal functions, n will be approaching infinity, allowing us to re-write the function g(x):

A(n), B(n) and C(n) are functions to represent the constants a, b and c in each sinusoidal function within the series instead of listing them one by one. For instance: A(1)=a1. This equation can also be further modified to better suit our context. The variables n and x while looks simple, does not help us model soundwave. We can thus replace them with frequency (f) if oscillations in a given time interval (t). The “n” in the expression represent the n’th sinusoidal function in the series, and we can instead label each sinusoidal function using their frequency. For example, the cosine function with a frequency of f would have a weight of A(f). We also need to modify the cosine function in this specific term to have the frequency (oscillations for 1 unit of time) f in the specific point in time t. The function is first horizontally compressed by 2π, to set its period to 1, and then expanded by its period. Since the inverse of the period is just the frequency, the cosine function can be written as cos(2πft), and the function g(x) can thus be represented as fallowing:

Since the Fourier series is actually continuous, we can use an integration to represent the function instead:

This expression uses two functions A(f) and C(f) to represent the amplitude and phase shift of each sinusoidal function in the series. Therefore, when trying to find the Fourier series of a soundwave, we need to determine the amplitude and phase shift of each frequency.

A useful tool to represent the amplitude and phase shift is complex numbers, as they can be represented in polar form: |z|cis(θ). |z| is the absolute value of the complex number, which is also the amplitude of the cis function, and θ can be used with to determine the phase shift of the function.

How can we then incorporate complex numbers into our process of converting a soundwave to a function that outputs the amplitude and phase shift when given a certain frequency? The special characteristics of averages of a sine function within a certain interval, which we have investigated before, inspired me to utilize the concept of centroid in geometry.

A radial symmetrical shape has its centroid located in the center of the shape. The centroid can be interpreted as the average location of all the points on the shape, which is why it is also the center of gravity of the shape. The radial segments of the radial symmetrical shape cancels each other out, causing the average of the location of points on the shape to be in the middle.

A symmetrical shape can be created by repeating a certain pattern multiple times to form its sides. Since sinusoidal functions have a repeating pattern, a shape exhibiting radial symmetry can be created with the sinusoidal function as its side and twisting the sine function around a certain point.

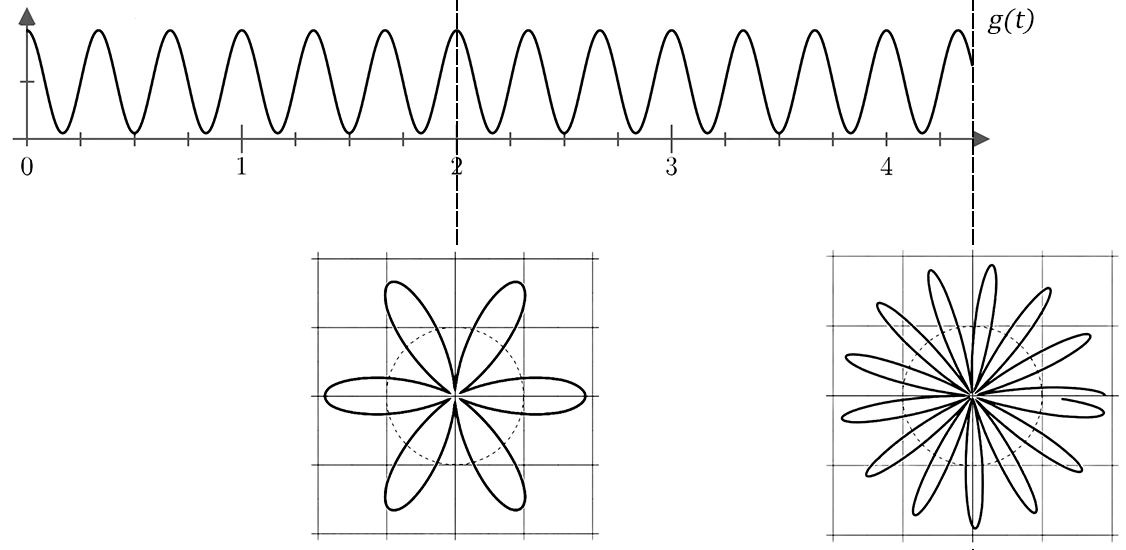


Figure 3. Winding the function g(t)=sin(6πt)+1 for t∈[0,2] and t∈[0,4.3] around a certain point

In figure 3, I chose to use the function sin(6πt)+1 so that I do not need to deal with negative values on the visual representation, as they might be distracting. The figure suggests the importance of the domain chosen for the winding: if the domain is an integer multiple of the period of the graph, (2 is an integer multiple of by 6 times) then the shape will symmetrical. If the domain is not an integer multiple of the period of the sin function, then it will not be a radial symmetrical shape.

While a 2D Cartesian plane can be used to plot the shape, and find the location of the centroid of the shape, finding an equation that describes the shape will be difficult. Instead, we can use the useful tool of complex numbers’ polar form to graph this function.

Recall function g(t) from figure 3. The process of transitioning the graph into a shape can be seen as changing the x-axis into the degree of rotation and y-axis to the length from a certain point to the origin. Since the x-axis values are just t and y-axis values are g(t), we can write the complex expression . The expression within the cis function is the same as the sine function in g(x) because we also need to perform the same horizontal compression the cis function to make sure that both function represents the same graph.

The expression can be further generalized by using the variable f to represent the frequency. The transition that we are essentially performing on the graph is compressing a normal sine wave horizontally by 2π, and then horizontally expanding it by its period. Notice that the transition that we are performing here are as same as the transitions that we have performed to a Fourier series, which prompts us that we are on the right path. The expression thus becomes: .

Since we need to find the average location of the graph, which is a continuous set of data, we can divide its integral by its domain. After this modification, we can create a function of f:

Now that we have built this complex expression , which we know have some sort of connection to the abundance of sinusoidal functions with different frequencies, but what does it actually do and how does it behave? I claimed that when the frequency f is a whole number multiple of the actual frequency of the function, the centroid will be in the middle of the shape, which in this case is the origin. This is actually a false claim because there is one exception of this rule when f is exactly the same as the frequency of the function g(x). This is because the whole number of multiples actually indicator of the amount of segments that is radially symmetric. This can be seen from the figure 3, where the whole number multiple is , which is also the amount of segments in the shape which it produces. When the whole number multiple is one, there is only one segment, and there is nothing else to balance it out, making it so that the centroid of the shape being far away from the center. This can be seen in figure 4:

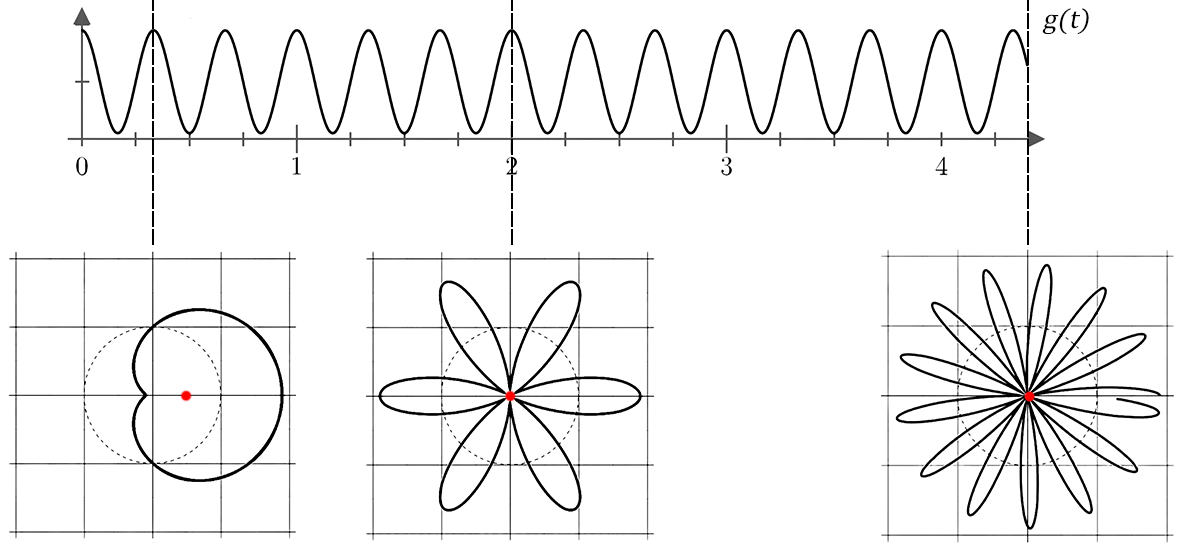


Figure 4. Winding the function g(t)=sin(6πt)+1 for t∈[0,1], t∈[0,2] and t∈[0,4.3] around a certain point and their centroids respectively

There is something special happening to the shape on the very right: even though it is not radial symmetrical, its centroid still appears to be so close to the origin that it is almost indistinguishable by eye. This is because of the large amount of oscillation that happens with in the period. It contains 12.8 phases, meaning that each phase only account for of the location in the centroid. This causes the lack of radial symmetry of the shape to have a small effect on the location of the centroid of the shape. The influence of the lack of symmetry on the shape also decreases as a larger interval is considered. This can be seen if we use a simple function to represent the influence of length of the interval on the effect of a single period on the centroid: where f is the frequency of the wave. Since x is in the denominator, . This means that the larger the range selected, the smaller affect the lack of symmetry on the centroid.

The distance between the centroid of the shape and the origin can thus be connected to the abundance of a sinusoidal function with the period as same as the time interval. This is because the effect on the lack of radial symmetry on the distance between the centroid and origin can be eliminated by increasing the boundary of the domain to ∞ and -∞. After whipping out the effect of radial symmetry on the distance, the only factor which affects the distance of the two point is whether the period matches with time interval.

However, given the domain that is being used, it is impossible for the sinusoid to have the same period as the period must be a finite value. If the period of the sinusoid is used to control the amount of oscillations, we are limiting ourselves to just one cycle in the graph, which jams the entire set of real numbers into the domain 0 to 2π for the cis function.

This “bug” in our formula can be fixed by adding a small patch. Instead of using the period to control the frequency of oscillations in the shape, we can use the frequency itself as a variable. This increases the domain of the cis function from only 0 to 2π into -∞ to ∞. This way, if the frequency of winding the graph matches with the frequency of the sinusoid oscillating, the period of the sinusoid will match up with the domain of one cycle in the graph, causing the centroid of the shape to deviate from the origin.

This modification is essentially making the cis function to have the same frequency as the sinusoid function, and can be carried out by expanding the sinusoid horizontally by f when plotting it onto the complex plane. This modifies the expression to: