Spring 2017

Competing Theories of Pitch Perception: Frequency and Time Domain Analysis

Nowell Thacher Stoddard
Bard College

Competing Theories of Pitch Perception: Frequency and Time Domain Analysis

A Senior Project submitted to
The Division of Science, Mathematics, and Computing
of
Bard College

by
Nowell Stoddard

Annandale-on-Hudson, New York
May, 2017

Abstract

Pitch perception is a phenomenon that has been the subject of much debate within the psychoacoustics community. It is at once a psychological, physiological and mathematical issue that has divided scientists for the last 200 years. My project aims to investigate the benefits and shortcomings of both the place theory and time theory approaches. This is done first by a model consistent with the long standing focus on the frequency domain, and then by expanding to a more modern approach that functions in the time domain.

Contents

Dedication

This project is dedicated to all of the amazing teachers I have had throughout my life.

Acknowledgments

I feel unbelievably lucky to be where I am today, none of it would have been possible without the support of my professors, friends and family. Of course this means the amazing and tireless professors of the Bard physics department. Specifically, I would like to thank my adviser Matthew Deady who has encouraged and supported me even when I doubted myself, his warmth and wealth and knowledge is sometimes hard to believe, thank you Matt. I want to thank Paul Cadden-Zimansky for my experience working in the lab and for teaching me to think more critically about the world around me. I would also like to thank Hal Haggard for always being so easy to talk and who always seems to be able to change my perspective on a given subject and finally I to thank Joshua Cooperman for making our quantum class fun, intriguing, mind boggling and challenging, even during the depths senior year.

I would like to thank my parents, Brook and Susan, who imparted their love of learning to be and have always supported my journey, even though I went four years without taking a single art history class. I also want to thank my big brother Percy, a constant source of smart and grounded wisdom in addition to being as good a friend as anyone could ask for. A big thank you to godparents, Kathy and Bruce, I always know I can turn to you for good sound advice.

A special acknowledgement to my cousin Amelia who passed away last spring, she showed me how full a life one can live if they really put their mind to it.

I want to acknowledge all my friends who make every day of my life better. I'd like to thank all my friends in the physics department for all the nights in Hegeman that became more goofy than productive. Of course I want to thank everyone in the Grave Street gang, Krisdee Dishmon, Aldo Grifo-Hahn, Lauren Russo, Ben Lorber and Kaiti Buchbaum, I can't imagine spending this year anywhere else or with anyone else, y'all are the best. I also want to acknowledge my good friends from back home, Sophie Simmons, Jonathan Gross, Liza Simmons, Joe Newlin, Holly Perkins and Ali Perkins, after all these years your friendship still means so much to me.

1
Introduction

In our lives we are constantly being bombarded with sound from cars' engines to the wind rustling through the trees. These sounds are coming at us in all sizes and shapes yet there are certain patterns that our ears are designed to pick up on, process and sort into pitches. We know that any sound that we hear must be a wave of pressure oscillations propagating through the air. Why then is it so much easier to identify the musical notes of a violin's string than the crash of dropped cinderblock? There must be something fundamentally different about the sound wave produced by the violin that our sensory system can pick up on. It is conceivable the brain does all the heavy lifting, receiving every single sound stimulus and assigning a sensation to each and a pitch to some. It turns out that our ears are equipped to sort out sound signals that produce pitch and begin processing them before they even reach the brain.

My interest in pitch perception began in my freshman year when I took a class in psychoacoustics. The class presented a mostly qualitative approach to the fundamentals of sound with a focus on musical applications. I was also taking my first college level physics class and although I didn't pursue the physics side of my class at the time it got me thinking about the boundaries of physics with psychology and neuroscience. As I progressed through the physics curriculum I shifted my focus to the core subjects of my courses. I began tutoring a non-major acoustics class where my responsibilities had more to do with walking students through algebraic equations and setting up physics problems than the core material of the course. Nevertheless, I read through their textbook and began to reacquaint myself with the basics of acoustics that I had encountered as a freshman. Finally, in my senior year I had the opportunity to take an advanced acoustics tutorial where we touched on the missing fundamental. The subject defied my intuition which was still largely based in Fourier analysis, so of course the only possible solution was to investigate.

In this project I will begin with some background on the structure of the ear, building up the physiological grounding that facilitates our sense of hearing. I will use this as a platform to discuss what sound is and how it propagates through the ear. This framework will allow me discuss differences between the real sound wave that is generated out in the world and the pitch that we eventually perceive.

Next I will discuss the a historical model of pitch perception. Building from the basis of Fourier analysis to a more realistic approach based on the harmonic motion of the basilar membrane. Building this picture up analytically should give insight into the physical parameters necessary for such a model to be physically plausible.

With the theoretical model established I then began to implement it in a simulation. During this process I was introduced to techniques of solving differential equations numerically which proved immensely powerful, if frustrating at times. Fine-tuning this first model took a lot of time while I revised my approach to optimize the solutions. While the model itself is seriously flawed, I learned a lot from working through it, often by trial and error.

When the first model was done to my satisfaction I immediately moved on to point out how flawed it really was. It soon becomes clear that improving on the existing model might address a few of its issues but I would need to rethink my entire approach. Here is where the historical context became very helpful, as it turns out that several nineteenth century thinkers had encountered the same problem. While there were some solutions put forth at the time, none of them held up to experimental rigor.

By the mid twentieth century a giant step forward was made by reframing the problem, shifting the focus from individual frequencies to a more macroscopic view of an entire sound wave in time. I will then re-approach pitch perception from the time theory perspective, demonstrating some of its qualities myself as well as analyzing a very successful implementation of this model.

2
Background of the Cochlea

The cochlea is a fluid filled organ located in the inner ear. It transforms the pressure variations of a sound wave into nerve signals that are transmitted to the brain for processing. The simplest picture of the cochlea is a black box that takes a time dependent pressure wave and transforms it into a frequency. This frequency is then recorded and sent on to the brain as a series of discrete electrical impulses. The field of Cochlear Mechanics aims to take a closer look at how this transformation occurs and how the transformation process might affect the final signal being transmitted.

2.1 The outer and middle ear

The majority of my work is focused on the mechanics of the inner ear so accordingly I have not devoted much attention to the intricacies of the middle and outer ear. Still, any sound that reaches the inner ear must travel through both these regions. Therefore, a basic understanding of their functions to some extent is absolutely necessary as there could be some process occurring in these regions that will affect the sound processing somewhere down the line. In the proceeding model I will treat signals entering the cochlea as being essentially unchanged by their journey through the outer and middle ear. This is undoubtedly an oversight but I believe that the effects of this oversight will be relatively small compared to the overall process inside the cochlea.

2.1.1 The outer ear

The outer ear is comprised of two pieces: the pinna and the meatus. The Pinna is the fleshy piece that protrudes outside the skull. Its function is to direct incoming sound waves into the ear canal and to support pieces of jewelry if one is so inclined. The pinna's asymmetrical and somewhat strange shape may be there to help with localizing incoming sounds. The meatus or ear canal is a channel that simply directs sound toward the tympanic membrane or "eardrum". The tympanic membrane is elastic so that the pressure of the sound wave will press on the ear drum and snap it back to its original position almost immediately. This allows it to respond quickly to high frequency pressure variations.

2.1.2 The middle ear

The far side of the eardrum begins in the region known as the middle ear. The middle ear refers to an entire chamber of the ear but for modeling purposes it can be thought of simply as three tiny bones called the *malleus* (the hammer), the *incus* (the anvil) and the *stapes* (the stirrup). Together these bones are known as the auditory ossicles (fig 2.1.1). The sounds coming into the ear are pressure variations in the air and must eventually be transmitted to pressure variations in the endolymph fluid of the cochlea. This endolymph fluid has an impedance hundreds of times higher than that of the air outside the ear. If

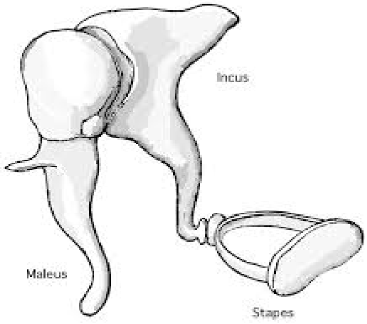


Figure 2.1.1: Auditory Ossicles

the pressure wave were to be transmitted directly from the eardrum to the endolymph fluid, around 99.9% of the sound energy would be reflected back, leaving only 0.01% to go on to the inner ear. The ossicles accomplish this impedance transformation by acting, together with the ligaments that attach them, as a lever and fulcrum. These ligaments can also tighten or loosen in order to reduce the energy going into the inner ear if one is in a loud environment. This so called "acoustic reflex" only works for sustained noise; if there is a sudden loud noise then the ligaments will not have time to loosen, and the sound will be transmitted as normal. The middle ear also contains the Eustachian Tubes which primarily serve to equalize pressure between the two sides of the tympanic membrane.

2.2 Structure of the Cochlea

The cochlea has a conical structure when stretched out, but is curled up like a snail shell inside the actual ear. It is narrower at the base and becomes progressively wider towards the apex which is called the helicotrema. Inside the coil there are three membranous tubes. In cross section(fig 2.2) on top would sit the the scala vestibuli which is connected at one end to the middle ear by a thin membrane called the oval window. On the bottom would be the scala tympani which is connected to the middle ear by another membrane known as the round window. The cochlear duct or scala media separates the two outer sections of the cochlea with scala vestubli on top and the scala tympani on the bottom. Sometimes the entire region is simply referred to as the basilar membrane although the basilar membrane is actually only a part of the cochlear duct,

In reality the scala vestibuli and tympani are connected at the apex of the spiral and so are actually one continuous tube. However, they have different effects on the cochlear duct so it is helpful to think of them as distinct structures.

The cochlear duct is where the actual sound processing occurs as it houses the auditory nerve that sends electrical impulses to the brain. The whole thing is bounded by Reissner's membrane at the top and the basilar membrane at the bottom, as can be seen below

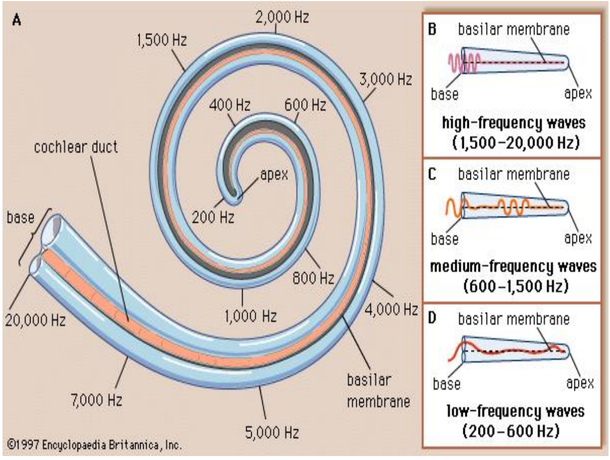


Figure 2.2.1: Top Down Cochlea

Reissner's membrane, is simply a wall. The basilar membrane is where all the action is. Resting on the bottom is the organ of Corti which is comprised of bundles of hair cells pressed up against the protruding tectorial membrane.

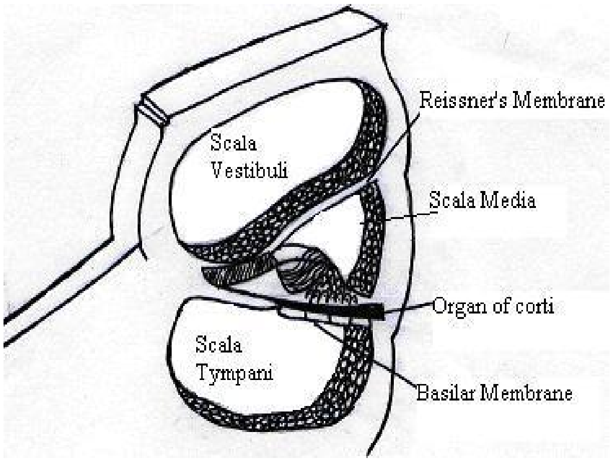


Figure 2.2.2: Cochlea Cross-section

When a pressure wave enters the cochlea through the oval window, it travels through the scala vestibuli to the apex of the cochlea, passes through the helicotrema and then travels back down the scala tympani to transfer the wave out through the round window.

The basilar membrane (BM) varies in stiffness and width along its length, going from stiffest and narrowest at the base to being pliable and wider at the apex. The varying stiffness along the BM means that particular regions will respond differently to pressure waves moving past them. The motion of the BM acts like a traveling wave like the wave produced by flicking a rope. The pressure variations of the sound wave passing by cause the BM to oscillate back and forth; this in turn pushes the inner hair cell bundles up against the tectorial membrane which does not respond to pressure variations like the BM. When the upper hair cells bend against the TM they open a gate allowing positive ions to flow into the hair cell, which triggers the release of the neurotransmitters. The particular hair cells in a bundle are of different length so that a stronger signal will excite all of them and a weaker signal will only excite the taller ones.

Like any simple harmonic motion, the BM's movement has frequencies that it will respond strongly to, based on its dimensions and stiffness. These resonant responses to specific frequencies are distributed over the length of the BM with high frequency resonance at the base and lower frequency responses at the apex. In general the stronger the response of the BM, the stronger the electrical signal that is sent to the brain. Where the signal comes from is largely, but not entirely, what determines which pitch the brain perceives. The regions that respond well to a particular frequency are known as critical bands. These critical bands vary in width in and how they are distributed along the cochlea. These critical bandwidths are linearly related to the center frequency to which the critical band is sensitive. As a result the critical bands are spaced farther apart for lower frequencies near the apex but are clustered much closer together for the higher frequencies at the base.

In summary, sound waves traveling through the cochlea will cause vibrations in the BM. The strength of the response is determined by the region of the BM, with different regions roughly responsible for recording different frequencies. However, it is important to keep in mind that BM response is not a binary on or off response. Any particular sound wave will excite the majority of the BM, only with certain regions having a much greater amplitude response. How quickly this larger response drops off will vary from model to model, depending what properties are being prioritized.

2.3 Historical development of Cochlear mechanics

Comprehensive study began in the Renaissance when scientists began investigating the human body through dissection, which had previously been restricted by the church. In the early sixteenth century an Italian, Andres Vessalius, described the ossicles of the middle ear. The cochlea itself was later discovered by Batholomeus Eustachius who not only was able to uncover its existence but also that it was divided into multiple channels and the tensor tympani. Surprisingly he did not have anything do to with the tube connecting the nasal cavity to the ear, although it is named for him. Further study of the anatomy of the ear was propelled by the invention of the microscope in 1590. Another great advancement came in 1822 with introduction of Fourier analysis which was applied to acoustics by Georg Ohm (1843) by proposing that the ear could pick up on the spectral frequencies of incoming sound. The pioneer place theory, the idea that a given frequency corresponds to a particular place on the BM, was introduced by Hermann Von Helmholtz (1821-1894). His theory comes from the application of Fourier analysis and is still considered by some to be a viable model, albeit with some adjustments. It was Georg Von Bekesy(1899-1972) who was finally able to observe cochlear response in live mammals and was able to show the varying amplitude responses of different parts of the BM for specific frequencies He later received a Nobel Prize in Biology and Medicine for his work. [?CM]

2.4 Sound and Pitch

2.4.1 Sound

Sound is produced from fluctuations in the air pressure relative to the atmosphere. When the fluctuations are sinusoidal, they can be represented by a single or collection of sine waves. We generally perceive these periodic fluctuations as tones. When the sound is a single sine wave I will refer to it as a pure tone, such as one that could be produced by an oscilloscope. The tone we hear will depend on the frequency of oscillation, for example, the frequency 261.4 Hz is associated with middle C. Humans are equipped to perceive frequencies ranging from about 20 Hz to 20000 Hz. The range will vary significantly with factors like age or nerve damage.

This kind of pure tone is quite rare to find in nature. Usually whatever vibrations set the air molecules in motion will have partials present, that is, the changes in pressure can only be represented by a linear combination of sine or cosine waves forming a complex wave. If these partials are multiples of some base frequency then we call them harmonic overtones. The combinations will still be periodic and so will still have an overall frequency, which will relate to the base or fundamental frequency if the partials are harmonic. Additionally, the overtones present will change the nature of the sound we hear. While it is true that middle C is associated with 261.4 Hz, the note will sound totally different depending on whether it comes from a violin or a human voice or an oscilloscope and speaker. The difference between all these sounds has to do with which overtones are present and the amplitude for each overtone. We call this difference between two tones of the same pitch timbre.

When dealing with complex waves, Fourier analysis can help extract information. Fourier's contributions here are twofold, the Fourier series and the Fourier transform. The Fourier series decomposes any periodic signal into it's constituent pure tones and assigns each with the appropriate amplitude to reproduce the original function exactly. For any periodic function $f(x)$ it takes the form

$$f(x) = a_0 + \sum_{n=1} a_n cos(nx) + \sum_{n=1} b_n sin(nx) \tag{2.4.1}$$

with a_n and b_n found by

$$a_0 = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)dx \qquad a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)cos(nx)dx \qquad b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)sin(nx)dx \tag{2.4.2}$$

Being able to look at pure sine pieces of a signal can highlight differences between two different tones of the same frequency and can help to gain some insight into things like their timbre.

The Fourier series decomposes a function of time into a linear combination of many functions of time which correspond to the spectral components of the original function. The Fourier transform, on the other hand, takes any function in the time domain, not just a periodic one, and reproduces that function in the frequency domain (In general it puts a function of x in terms of x^{-1}). It does this by putting the function into the basis of sines and cosine, or more simply as a complex exponential. It takes the form

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} g(\omega) e^{i\omega t} d\omega \tag{2.4.3}$$

with inverse Fourier transform given by

$$g(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(t) e^{-i\omega t} dt \tag{2.4.4}$$

Taking the Fourier transform of a time signal will show its spectral components. For a pure tone there will be a peak around its frequency. Figure 2.222 shows the Fourier transform of for a simple 100Hz sine wave which is peaked at its frequency

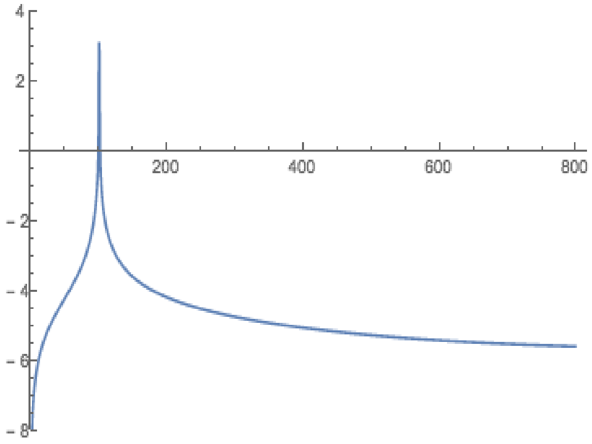


Figure 2.4.1: Pure Tone Fourier Transform

After the technique's invention it seemed like the perfect solution to those investigating the mechanics of human hearing. The human ear was thought of simply as a Fourier analyzer, transforming incoming signals into a collection of frequency spikes. The Fourier explanation is fairly comprehensive and was enough to convince some great physicists of the time such as Helmholtz and Ohm that this was the fundamental nature of human hearing, with just a few minor adjustments to be made. In fact there are those who still find this to be a convincing explanation. In Chapter 5 I will begin to examine some of the shortcomings of this approach. In the next chapters I will outline a model that mimics the Fourier results by way of physical analogy.

2.4.2 Pitch

Before we rush into investigating the mechanics of this model, we must take a moment to define our terms. Most important is the definition of pitch. In this section, I have referred to sound and tone but not to pitch. All three of these terms can be somewhat ambiguous but none more so than pitch. In my terminology, I will refer to sound as the physical phenomenon, i.e. a sound wave is a pressure oscillation occurring in a medium like air. This wave exists regardless of whether or not there is a listener present. Tone refers to the particular frequency of the sound wave and is likewise independent of an observer. Pitch is a scale of the musical notes we hear that can be organized from high to low. This definition is quite reminiscent of how frequency is thought of, the major difference being that pitch is a psychological phenomenon, it requires a human listener. It is a summary of the sound inputs received by the ear. Pitch is obviously related to sound, and particularly the frequency of that sound. I stated earlier that 261.4 is generally associated with middle C. This is not at all the same as saying that 261.4 *is* middle C. I also stated that timbre is primarily a product of the different overtones present in a particular sound, but timbre does not appear as information in the sound wave. If we were to Fourier analyze a complex wave with frequency 261.4 the result would not be some instrumental tone at middle C. Instead, the result would be a jumble of different pure tones each with different strengths.

The synthesis of all these overtones is what produces pitch. For Helmholtz and Ohm it was easy to relegate this process to higher brain function and assume the brain would handle the summation of the overtones and their amplitudes as well as the perception of timbre. We will see in chapter 5 that the other sound information will inform this summary, but for the time being we will leave pitch in the realm of psychology.

3
Modeling the Linear Cochlea

In this chapter, the response of the cochlea is approximated by a linear model in which the critical bands of the cochlea respond to various frequencies with no interaction or coupling. This can be modeled by treating each band as a filter, which is tuned such that it will have a very large amplitude response near a specific frequency and drop off quickly for other frequencies. These models can vary depending on what kind of mechanical filter is used and what the parameters of these filters are. The parameters represent the physical characteristics of the system such as the mass density of a section of the cochlea or its stiffness at a given point. They will determine how wide the band is, how quickly the amplitude will grow or shrink relative to its resonant frequency, and how long, if at all, any transient effects will persist. Before investigating sources of non-linearity I first aim to construct one of these linear models. This can hopefully form a basis for more complex models. I have opted to choose parameters that will prioritize minimizing transience in signal response, rather than having a greater and narrower amplitude response near resonance (i.e. a larger quality factor, more on this later). If there were transient signals when the cochlea first begins to respond then our perception of stable sound would change over time with no other external factors; this does not seem to be the case.

3.1 Independent Filter Bank

The filter bank I refer to is really just a set of uncoupled differential equations. Each differential equation will have the same form but different coefficients, meaning something of the approximate form

$$A\frac{dx}{dt} = x$$
$$B\frac{dx}{dt} = x$$

The coefficients will dictate the solution to each individual differential equation and where it will be maximized. In this model the cochlea can be entirely represented by a collection of these equations, provided that their critical bands agree with those found experimentally. Mapping the coefficients to each critical band corresponding to the actual ear would not be an impossible task given access to enough experimental data. However, this seems unnecessary; when looking at only a handful of regions will be easier to test and produce the same results for a linear model.

3.2 Characterizing the Bandpass Filter Bank

A simple way to have this kind of varying amplitude response is the equation for a damped, driven oscillator (DDO).

$$A\frac{d^2x}{dt^2} + B\frac{dx}{dt} + Cx = Dcos(\omega t) \tag{3.2.1}$$

This differential equation is fairly straight forward to solve analytically and the physical meaning of the coefficients is familiar. That is:

- $A \Leftrightarrow$ Inertial term
- $B \Leftrightarrow$ Damping term
- $C \Leftrightarrow$ Restoring term
- $D \Leftrightarrow$ Affects the amplitude of the steady state solution

If one solves the above equation the solutions do in fact have a resonant frequency and will have large responses around that frequency.

However, before I show that solution, I would like to reframe the problem in terms of circuitry elements. That is, solving the equation of an LRC series circuit instead of a DDO. This may seem like an arbitrary choice as the two equations look almost identical but there are a few small advantages to framing it this way. Firstly, while we are making this model mathematically, the LRC picture allows us to test this model physically with much tighter control over the elements than with a DDO. Additionally, using circuit elements establishes a common language so that if the model becomes more complicated it still may be framed in terms of electronics, which have a certain physical intuition. Once again it makes almost no difference except the notation will be slightly different.

The form of the LRC equation comes from Ohm's law $V_{in} = V_L + V_R + V_C$, or the voltage drop across all the circuit elements plus the input voltage sum to zero. Usually these voltages are written in terms of current instead of charge but recalling that $I(t) = \frac{dq(t)}{dt}$ and the formulas for the various voltage for each circuit element:

$$V_L = \frac{dI}{dt} = L\frac{d^2q}{dt^2}$$

$$V_R = IR = L\frac{dq}{dt}$$

$$V_L = \frac{1}{C}\int Idt = \frac{q}{C}$$

it is simple to write the equation in terms of charge. With the substitutions made, the equation, reads

$$V_0e^{i\omega t} = L\frac{d^2q}{dt^2} + R\frac{dq}{dt} + \frac{q}{C} \tag{3.2.2}$$

Which looks almost exactly the same as the DDO example except that

- $A \to L$, $B \to R$, $C \to 1/C$
- $Dcos(\omega t) \to V_0e^{i\omega t}$ where is the input voltage

- $x(t) \rightarrow q(t)$ where q is the total charge in the circuit as a function of time

Now there is a neat physical analog to help visualize the process. As this model eventually shows some fundamental flaws I never ended up building a physical model. Still, building the model was only part of the appeal of the LRC picture. I was also more familiar with thinking of the equation in circuitry terms. This approach would have allowed the model to be extended by coupling the different filters with other circuit elements.

3.3 Solving A Single LRC Filter

In order to see the how the LRC circuit responds at different frequencies, the equation needs to be solved for $q(t)$. This can be accomplished by guessing a solution for $q(t)$, or in this case a solution for $\frac{dq}{dt}$. A standard guess for a solution like this is an exponential:

$$\frac{dq}{dt} = I(t) = I_0 e^{i\omega t+\phi} \tag{3.3.1}$$

which yields

$$q(t) = \frac{I_0}{i\omega} e^{i\omega t+\phi} \tag{3.3.2}$$

and

$$\frac{d^2q}{dt^2} = I_0 i\omega t e^{i\omega t+\phi} \tag{3.3.3}$$

If this guess is correct then the only unknown is the quantity I_0 which can found by plugging $q(t)$ back into the characteristic LRC equation and solving. I_0 represents the amplitude response of the system and will have a resonant frequency where it will be maximized .

$$V_0 e^{i\omega t} = LI_0 i\omega t e^{i\omega t+\phi} + RI_0 e^{i\omega t+\phi} + \frac{I_0}{i\omega C} e^{i\omega t+\phi} \tag{3.3.4}$$

Each term has $e^{i\omega t}$ in it, so dividing that out and regrouping the results in

$$V_0 = I_0 e^{i\phi}(Li\omega + R + \frac{1}{i\omega C}) \Rightarrow \frac{V_0}{I_0} e^{-i\phi} = [R + i(\omega L - \frac{1}{\omega C}) \tag{3.3.5}$$

then multiplying both sides by their complex conjugate gets rid of the exponential yields

$$\frac{V_0}{I_0} = \sqrt{R^2 + i(\omega L - \frac{1}{\omega C})^2} \Rightarrow I_0 = \frac{V_0}{\sqrt{R^2 + i(\omega L - \frac{1}{\omega C})^2}} \tag{3.3.6}$$

I_0 will be maximized when $\omega L - \frac{1}{\omega C} = 0$. This happens when $\omega = \sqrt{\frac{1}{LC}}$ which is the circuit's *resonant frequency* ω_0.This means that when the system is driven near to ω_0 it will have a much larger response than when it is driven at some arbitrary frequency ω. With the dependance of ω_0 on L and C known, the next step is to decide which values of L, R and C will produce the right resonant frequencies to match the critical bands under investigation.

3.4 accounting for γ factors and Q factors

When choosing values for L, R and C there are three main concerns. The most important thing is that ratio of $\sqrt{\frac{1}{LC}}$ matches the desired frequency for a particular filter. Additionally the values for L R and C should be chosen with the γ-factor and Q-factor for each filter in mind.

The solution to the ODE is periodic with a frequency determined by the driving frequency and the oscillator's own resonant frequency. The solution will eventually settle into a steady frequency at some particular amplitude. Before this happens, the amplitude envelope will start at an initial value and decay exponentially into the steady state solution. The rate of this decay is governed by the γ-factor:

$$\gamma = \frac{1}{2}RL \tag{3.4.1}$$

Another way to describe the γ factor is to say it determines how long any transience in the system will last. For pitch perception, we want as little transience in the signal as possible. If the transience decayed over a perceptible timescale, it would result in our pitch perception changing over course of a single tone or perhaps other strange auditory effects. Clearly we want to minimize transience which means a low γ-factor.

The Q-factor, or quality of the system, determines the bandwidth of the filter and to some extent the strength of the amplitude response will be. A higher Q means a narrower bandwidth and its value is determined by

$$Q = \frac{1}{R}\sqrt{\frac{L}{C}} = \frac{1}{R}\sqrt{\omega_0 L} \tag{3.4.2}$$

It should be noted that R is in the numerator for the γ-factor and in the denominator for Q-factor making them inversely proportional to each other, if we consider L and C to be static. A high Q and low γ is ideal for this model as it means low transience and more selective filters. In the actual ear, the filters actually have fairly wide bandwidths but for sake of clarity in measuring responses we will aim for a high Q and see how the system responds.

4
Implementing linear model

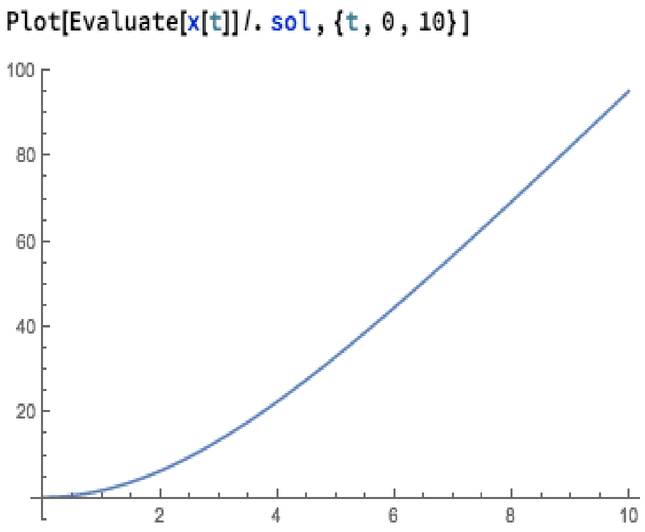
Now that the bandpass filter model has been worked out analytically the computational model can be implemented. Even though the ordinary differential equation for the LRC circuit was fairly straightforward to solve, when trying to solve several of these ODEs at once it is much easier to let a computer program solve them numerically. This has the added benefit of getting a plot of the solutions with little effort. Plotting all the solutions simultaneously gives us a quick way to compare the amplitude responses of the whole filter bank at a particular driving frequency or driving function.

4.1 Tools: NDSolve with Mathematica online

The computation was done mostly in Wolfram Mathematica using the NDSolve tool. The tool uses a method similar to Euler's method for solving differential equations to get a set of points which it then interpolates into a smooth solution. For example a fairly straightforward differnetial equation in Mathematica notation is

$$sol = NDSolve[x''[t] == f/m - c/m * (x'[t])^2, x[0] == 0, x'[0] == 0, x, t, 0, 10] \quad (4.1.1)$$

and return an approximate solution based on the values of the constants f, m and c which is shown in figure 4.1.1



```
Plot[Evaluate[x[t]]/. sol, {t, 0, 10}]
```

Figure 4.1.1: A simple DE

The tool requires the actual differential equation, as well as the independent variable and range of that variable. It also requires a number of boundary conditions equal to the number order of the differential equation.

4.2 Building up the first filter

As when constructing a differential equation like a damped driven oscillator analytically, it is prudent to start with a simplified case and add in complexity bit by bit. In this case that means starting with an undamped, non-driven oscillator and then adding a damping term and finally driving frequency.

4.2.1 Undamped Oscillation

the movement of an undamped oscillator is just the simple harmonic motion that is the familiar motion of a mass on a spring dictated by $F = -kx$ or in differential form $m\frac{d^2x}{dt^2} + kx = 0$. This just yields a periodic solution $Cos(\omega t)$ (fig 4.2.1)

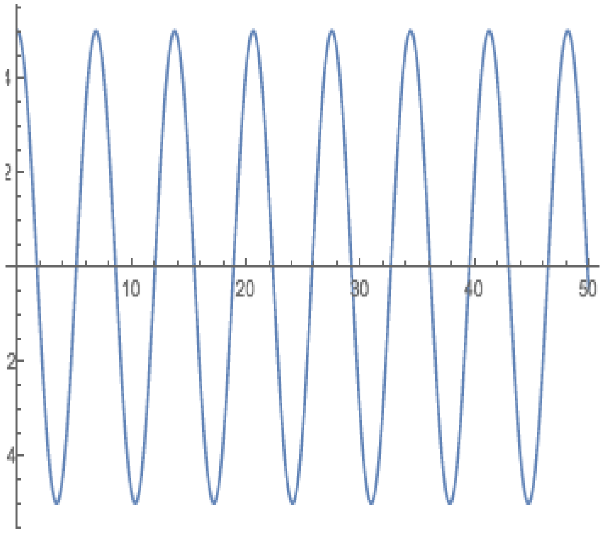


Figure 4.2.1: Undamped oscillator

4.2.2 damped oscillation

Adding in a negative damping term makes the value proportional to velocity. The negative means that the system will lose energy over time. Now the equation reads $m\frac{d^2x}{dt^2} - b\frac{dx}{dt} + kx = 0$ Or as can be seen from the solution in fig 4.2.2

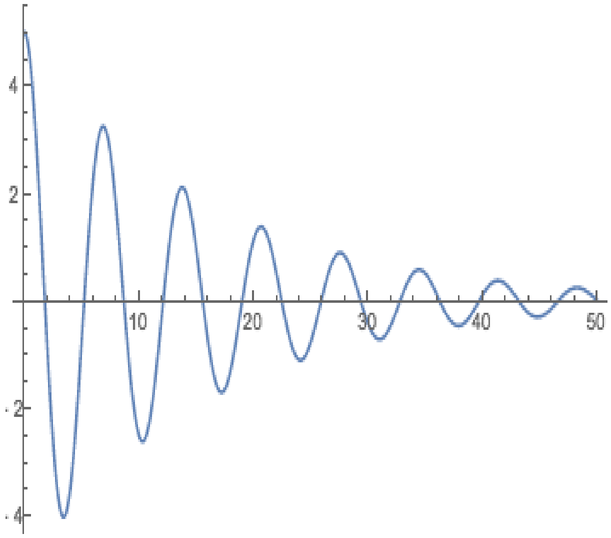


Figure 4.2.2: damped Oscillator

The amplitude of the signal decays over time exponentially. This is where the Q factor discussed earlier begins to come into play. A low Q-factor means that the system will lose energy and decay more quickly. The Q-factor will not effect the period of the signal only the amplitude envelope.

4.2.3 damped driven oscillation

Finally, to get the full bandpass filter we need to set the DE equal to a periodic function of t. In this case I used $y(t) = Acos(\omega t)$ so the final equation reads as expected

$$m\frac{d^2x}{dt^2} - b\frac{dx}{dt} + kx = y(t) = Acos(\omega t) \tag{4.2.1}$$

The first of these filters to successfully amplify a signal around its resonant frequency had values of $L = 12, R = 1.5, C = 0.1$ and took the boundary conditions $x(0) = 5$ and $x'(0) = 0$ In Mathematica it looked like

$$NDSolve[v[t] == x'[t], 12.v'[t] + 1.5v[t] + 10x[t] == y[t], x[0] == 5, v[0] == 0, x[t], t, 0, 100] \tag{4.2.2}$$

and it is being driven by

$$y[t_] := 10.Cos[wt], \omega = 2 \tag{4.2.3}$$

The last term indicates the independent variable and its range. With a solution given by

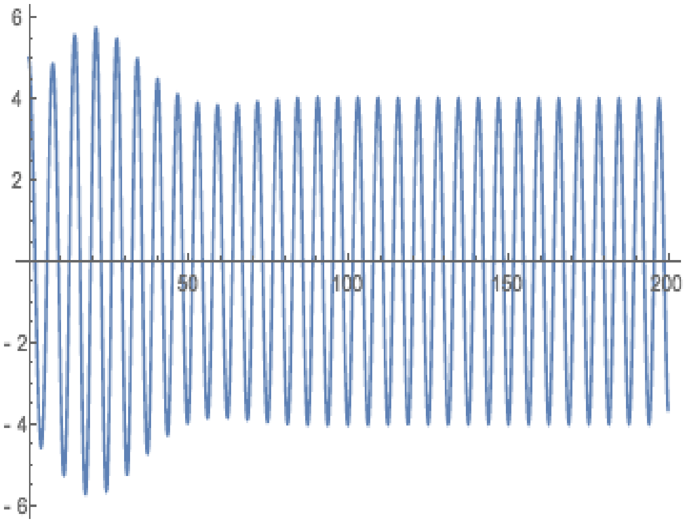


Figure 4.2.3: damped Oscillator

This particular filter has a resonant frequency at $\omega_0 = 1.9$ and is being driven at $\omega = 2$ so it has a fairly strong response. Note that the frequencies are being written in integers but could easily be scaled by $\frac{100}{2\pi}$ or $\frac{1000}{2\pi}$ to correspond to audible frequencies; however we are concerned with the *relative* amplitude responses. While the signal response is strong, there is clearly a lot of transient behavior before the signal stabilizes, meaning that we should adjust the parameters to keep the same ω_0 while increasing the γ-factor.

4.3 The Filter Bank

4.3.1 First Attempt

When I first attempted the filter bank I hadn’t begun to consider things like γ values and Q values. Instead I held on to the circuitry analogy and picked arbitrary values that seemed on the right order of magnitude for a typical circuit. For example my first filter had the form

```
sol1=NDSolve[g[t]==V[t] = (1/.01)*Q[t]+.100*Q’[t]+.0001*Q”[t] ,Q[0]==1, Q’[0]==0,Q[t],t,0,40]
```

Where

$$V[t_] := 10 * Cos[\omega * t] \tag{4.3.1}$$

Were the values for R, L and C were chosen based on the typical units I had seen for those circuit elements, i.e. tens of Ohms, millihenrys and microfarads. As might be expected, using these arbitrary values made the solution too messy to gain any real insight. While each iteration had some resonant frequency to it, there was so much transience and noise that there was little to no difference when changing the values for ω. At this point I was trying to create the filters first and then figure out the resonance after the fact. The results from this method were so inconsistent that it became clear that I would need to consider the effects of my coefficients on the solution, besides just what resonant frequency they yielded.

4.3.2 Second Attempt

Even when accounting for the γ-factor and Q-factor, having the correct resonant frequency is still the primary focus of any particular filter. So on the next attempt I started with the particular ω_0, recalling that $\omega = \sqrt{\frac{1}{LC}}$. In order to keep the solutions to the different filters as similar as possible, I elected to keep C at a constant value of 0.1 for ease of computation. I then changed the value of L to suit ω_0 for each particular filter. This is where I started using smaller ω's that could be scaled if needed. For example, for a filter with $\omega_0 = 2$ where $L = 5$ would solve

$$\omega = \sqrt{\frac{1}{LC}} = 2 \tag{4.3.2}$$

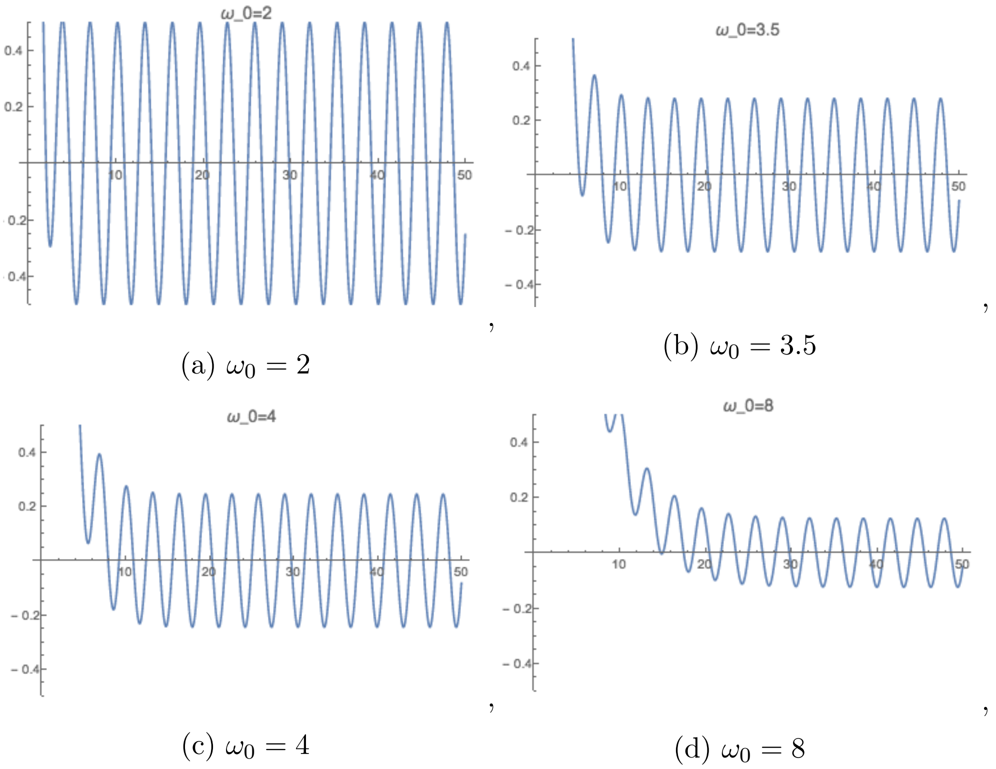
With $C = 0.1$

$$\sqrt{\frac{1}{(.1)L}} = 2 \Rightarrow 4(0.1) = \frac{1}{L} \Rightarrow L = 2.5 \tag{4.3.3}$$

This did produce a filter that responded well at the resonance frequency but of course I still needed some R values for each filter. I knew that a high Q factor would sharpen my signal so I picked an arbitrary value of $Q = 50$ and solved solved for R based on the value of L for that particular circuit. To follow the example at $\omega_0 = 2$

$$Q = \frac{\omega_0 L}{R} = 50 \Rightarrow R = \frac{\omega_0 L}{50} \Rightarrow R = 0.1 \tag{4.3.4}$$

Doing the same calculations for filters at $\omega_0 = 3.5, 4, 8$ produced a bank of four filters tuned to different frequencies(fig 4.3.1) at $\omega = 2$. The signal response was fairly good, although there was a long transient period. Transience aside the main problem was that the filter bank didn't respond as well at higher driving frequencies. For example fig 4.3.2 where the system is being driven at $\omega = 8$ filter with resonance at $\omega_0 = 8$ has the least response!



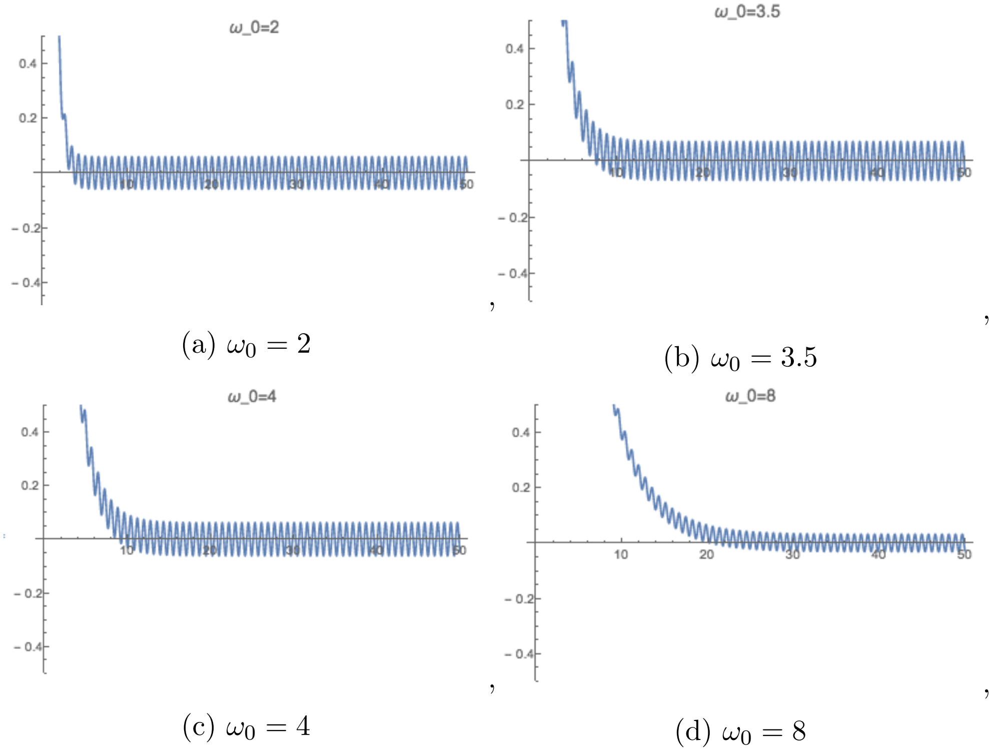
(a) $\omega_0 = 2$

(b) $\omega_0 = 3.5$

(c) $\omega_0 = 4$

(d) $\omega_0 = 8$

Figure 4.3.1: Q=50 Response at $\omega = 2$



(a) $\omega_0 = 2$

(b) $\omega_0 = 3.5$

(c) $\omega_0 = 4$

(d) $\omega_0 = 8$

Figure 4.3.2: Q=50 Response at $\omega = 8$

4.3.3 Third Attempt

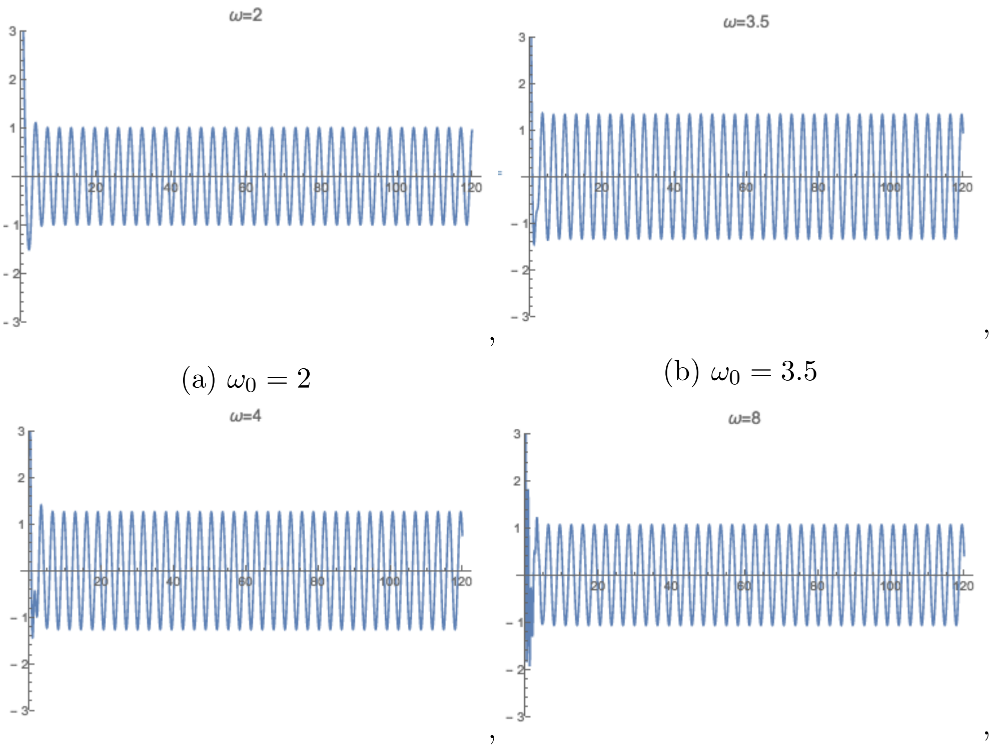
It was not yet clear why the higher frequency filters responded so poorly, so I decided to try and tackle the transience. Knowing that a low γ-factor would decrease transience the next filter bank was constructed with R values such that the $\gamma = 1$. The calculation is similar to setting a constant Q-factor, illustrating once again with $\omega_0 = 2$ with $C = 0.1$ We know that

$$\omega_0 = 2 = \sqrt{\frac{1}{(.1)L}} \rightarrow L = 2.5 \qquad \gamma = \frac{R}{2L} \qquad \gamma = 1$$

$$R = 2L = 5$$

These filters did have much less transience (see fig. 4.3.3 & fig 4.3.4) and much stronger responses near resonance. However their responses became much less distinct at lower frequencies, almost disappearing entirely.

The problem was with the choice to hold C fixed and change L in order to match ω_0. L appears in both γ and Q and in both cases, the value of L decreased with ω_0, which would generally decrease Q. This not only changes the values of γ and Q but also amplifies Q's dependance on ω_0. This explains why in the second case the quality factor is actually increasing for the higher frequency filters even though the values L are lower.



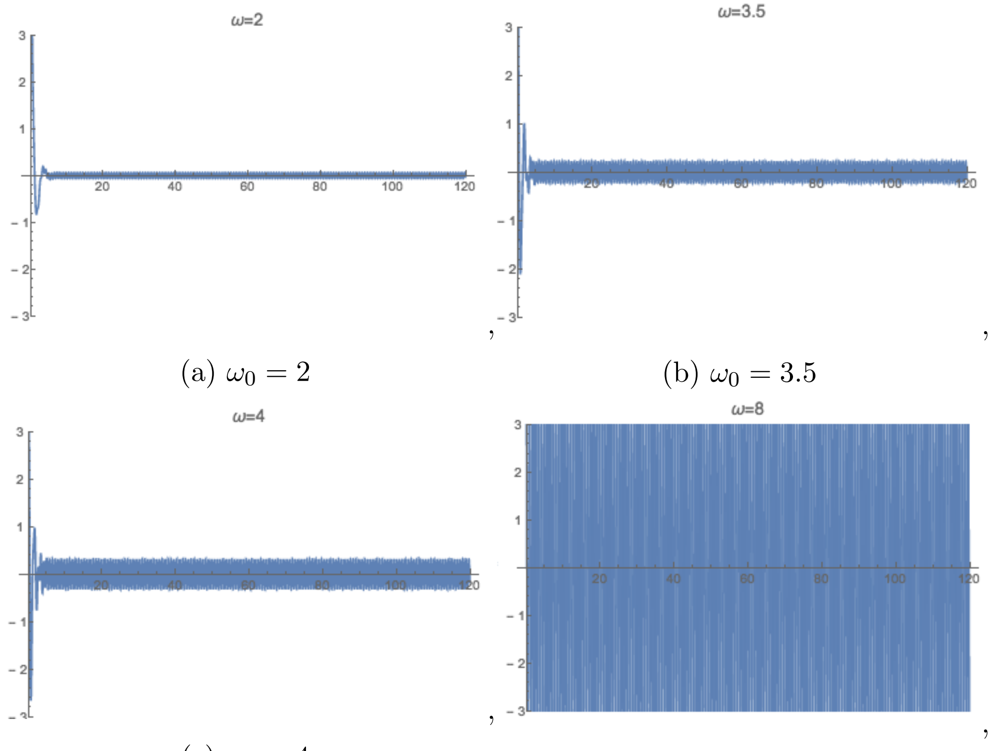
(a) $\omega_0 = 2$

(b) $\omega_0 = 3.5$

(c) $\omega_0 = 4$

(d) $\omega_0 = 8$

Figure 4.3.3: $\gamma = 1$ Response at $\omega = 2$



(a) $\omega_0 = 2$

(b) $\omega_0 = 3.5$

(c) $\omega_0 = 4$

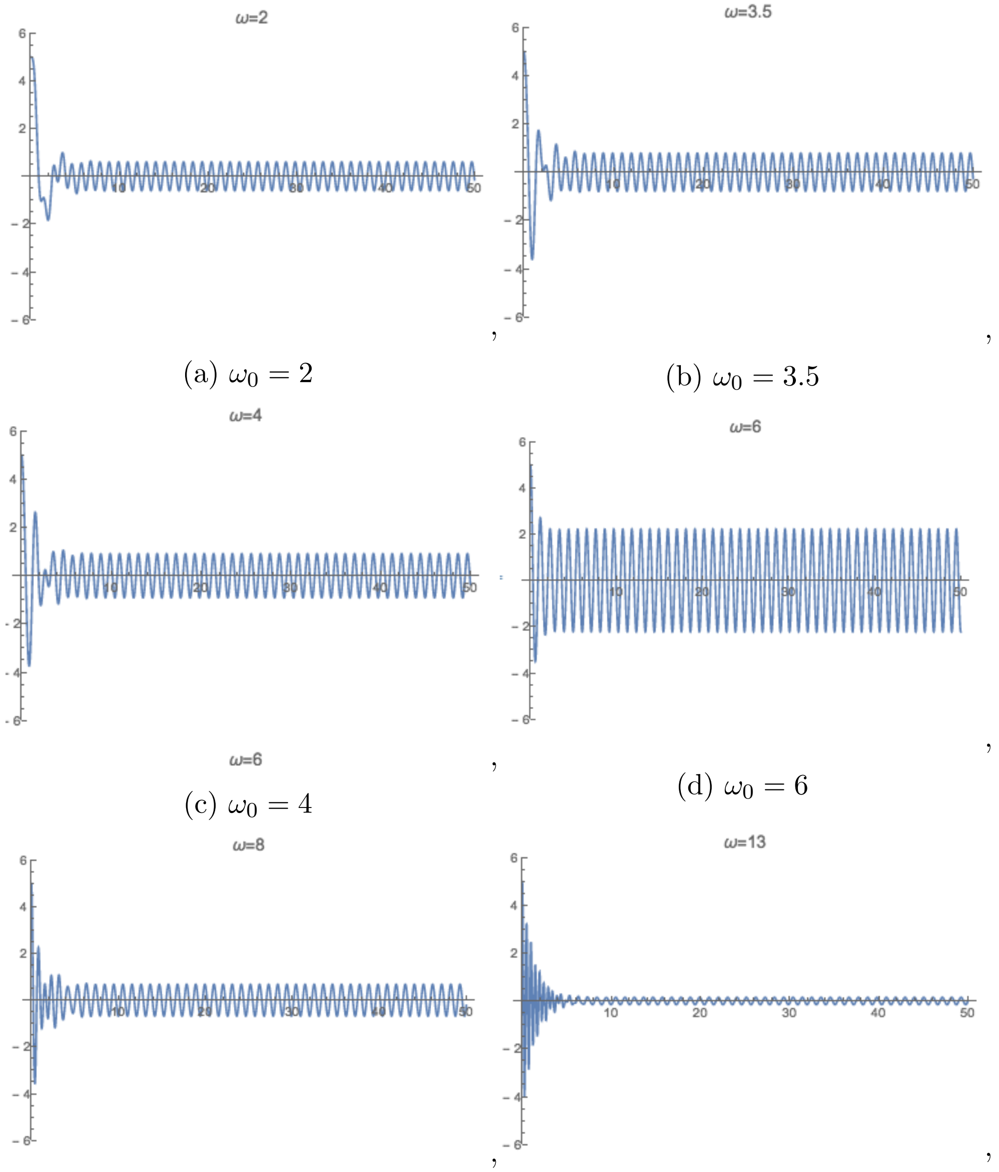
(d) $\omega_0 = 8$

Figure 4.3.4: $\gamma = 1$ Response at $\omega = 8$

4.4 The Fnisished Filter Banks

Instead of fixing C at 0.1, L was fixed at a value of 1 to simplify the calculations and to minimize its effect on Q and γ. This approach lets C be $\frac{1}{\omega_0^2}$ and lets R be chosen to optimize the solution. It is tempting to make the value of R arbitrarily small to maximize Q and γ, however as R decreases, it will again amplify the differences in Q caused by the different values of ω_0. I found that an R value of 1.5 worked well with this configuration. While the Q values vary a certain amount with ω_0, it is much more stable than the previous iterations and each filter shows a distinct response to its resonant frequency. There is some transience but it dies out fairly quickly compared to some of the earlier attempts. The Final code is written below as well as examples of the filter bank being driven at various pure tones with resonance for least one the filters.

```
w=6
y[t_] := 10Cos[w t]
solA=NDSolve[{v[t] == x'[t], v'[t]+1.5*v[t]+4x[t] == y[t], x[0]==5, v[0]==0},
  {x[t]},{t,0,100}];
solB=NDSolve[{v[t] == x'[t], v'[t]+1.5 v[t]+12.25x[t] == y[t], x[0]==5, v[0]==0},
{x[t]},{t,0,100}];
solC=NDSolve[{v[t] == x'[t], v'[t]+1.5 v[t]+16x[t] == y[t], x[0]==5, v[0]==0},
 {x[t]},{t,0,100}];
solD=NDSolve[{v[t] == x'[t], v'[t]+1.5 v[t]+36x[t] == y[t], x[0]==5, v[0]==0},
{x[t]},{t,0,100}];
solE=NDSolve[{v[t] == x'[t], v'[t]+1.5 v[t]+64x[t] == y[t], x[0]==5, v[0]==0},
{x[t]},{t,0,100}];
solF=NDSolve[{v[t] == x'[t], v'[t]+1.5 v[t]+169x[t] == y[t], x[0]==5, v[0]==0},
{x[t]},{t,0,100}];
```



(a) $\omega_0 = 2$

(b) $\omega_0 = 3.5$

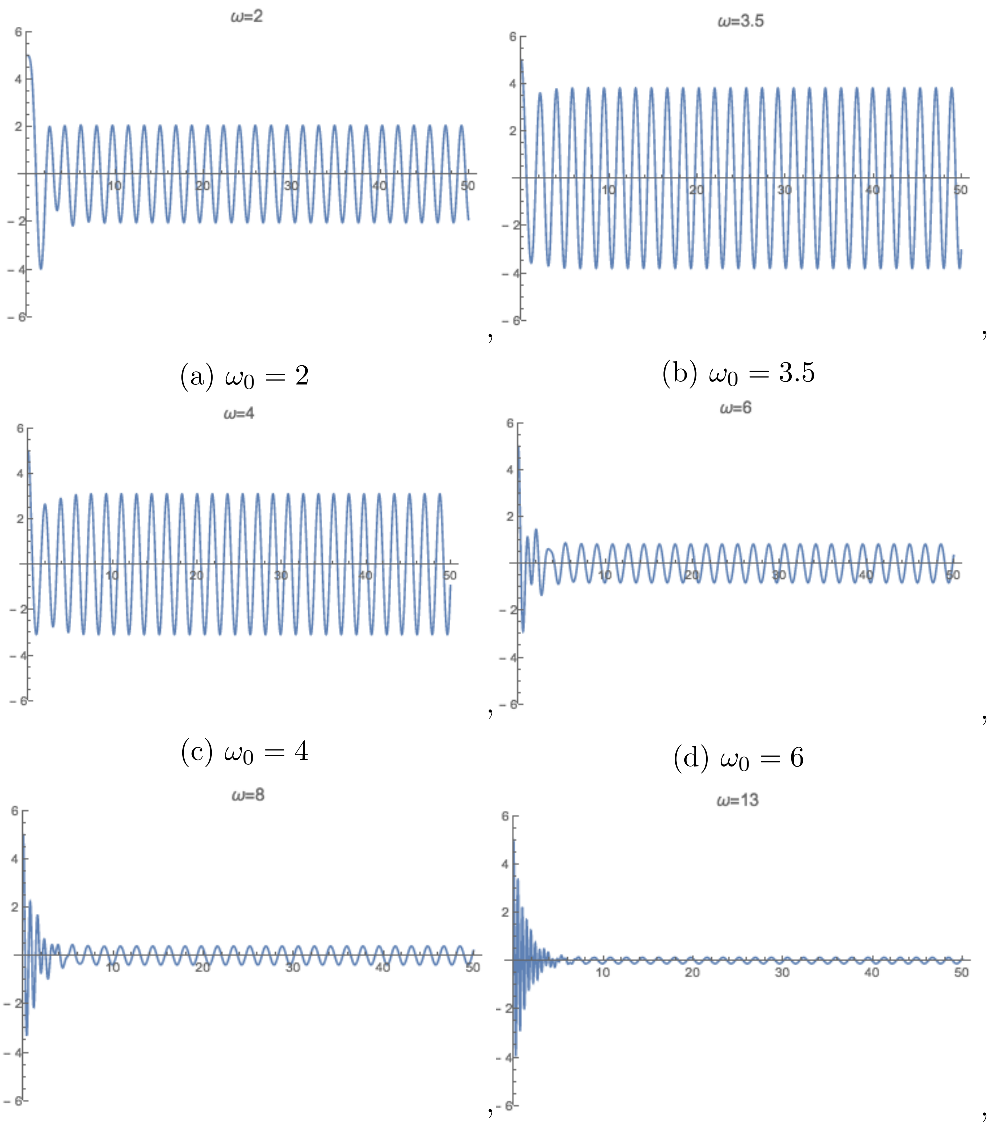
(c) $\omega_0 = 4$

(d) $\omega_0 = 6$

(e) $\omega_0 = 8$

(f) $\omega_0 = 13$

Figure 4.4.1: Final Filter Bank Response at $\omega = 6$



(a) $\omega_0 = 2$

(b) $\omega_0 = 3.5$

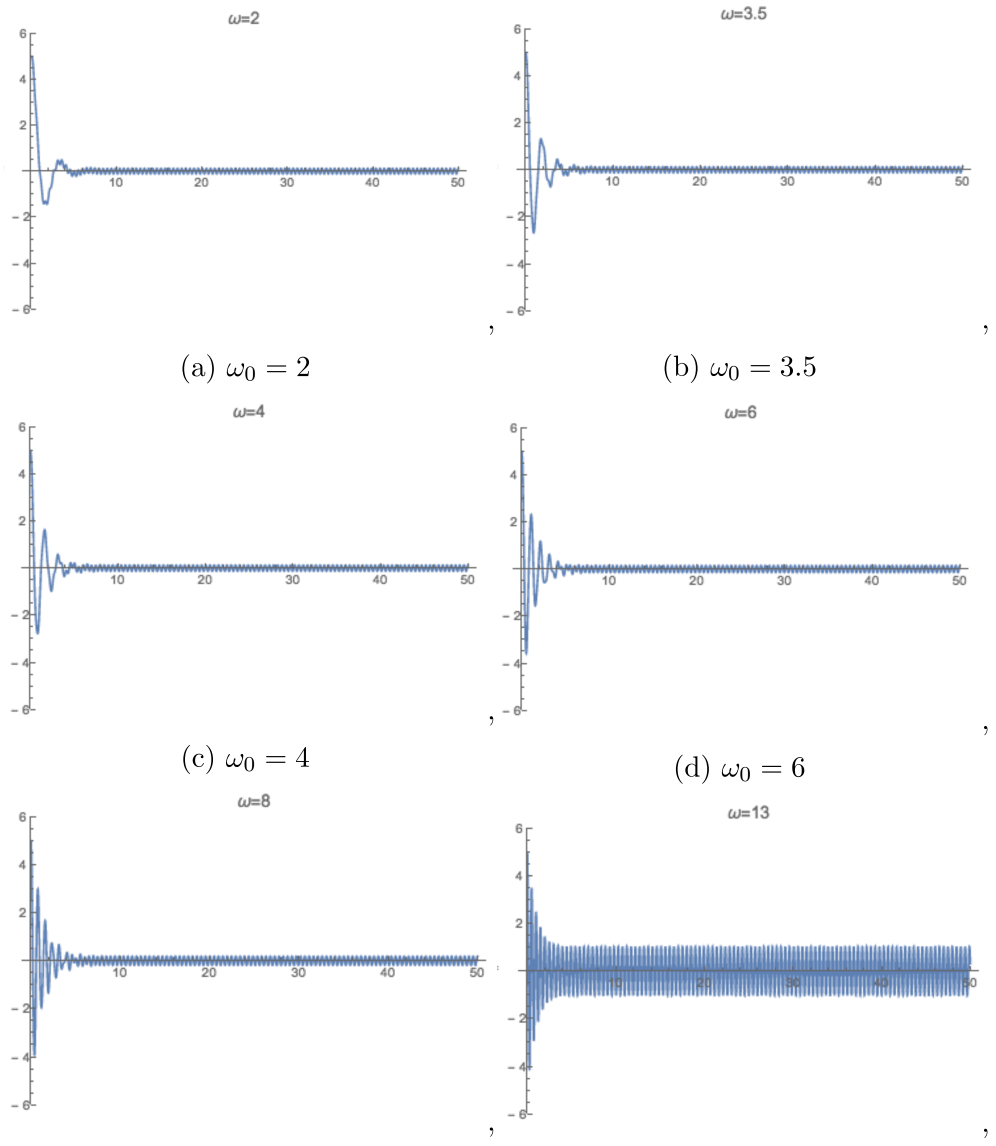
(c) $\omega_0 = 4$

(d) $\omega_0 = 6$

(e) $\omega_0 = 8$

(f) $\omega_0 = 13$

Figure 4.4.2: Final Filter Bank Response at $\omega = 3.5$



(a) $\omega_0 = 2$

(b) $\omega_0 = 3.5$

(c) $\omega_0 = 4$

(d) $\omega_0 = 6$

(e) $\omega_0 = 8$

(f) $\omega_0 = 13$

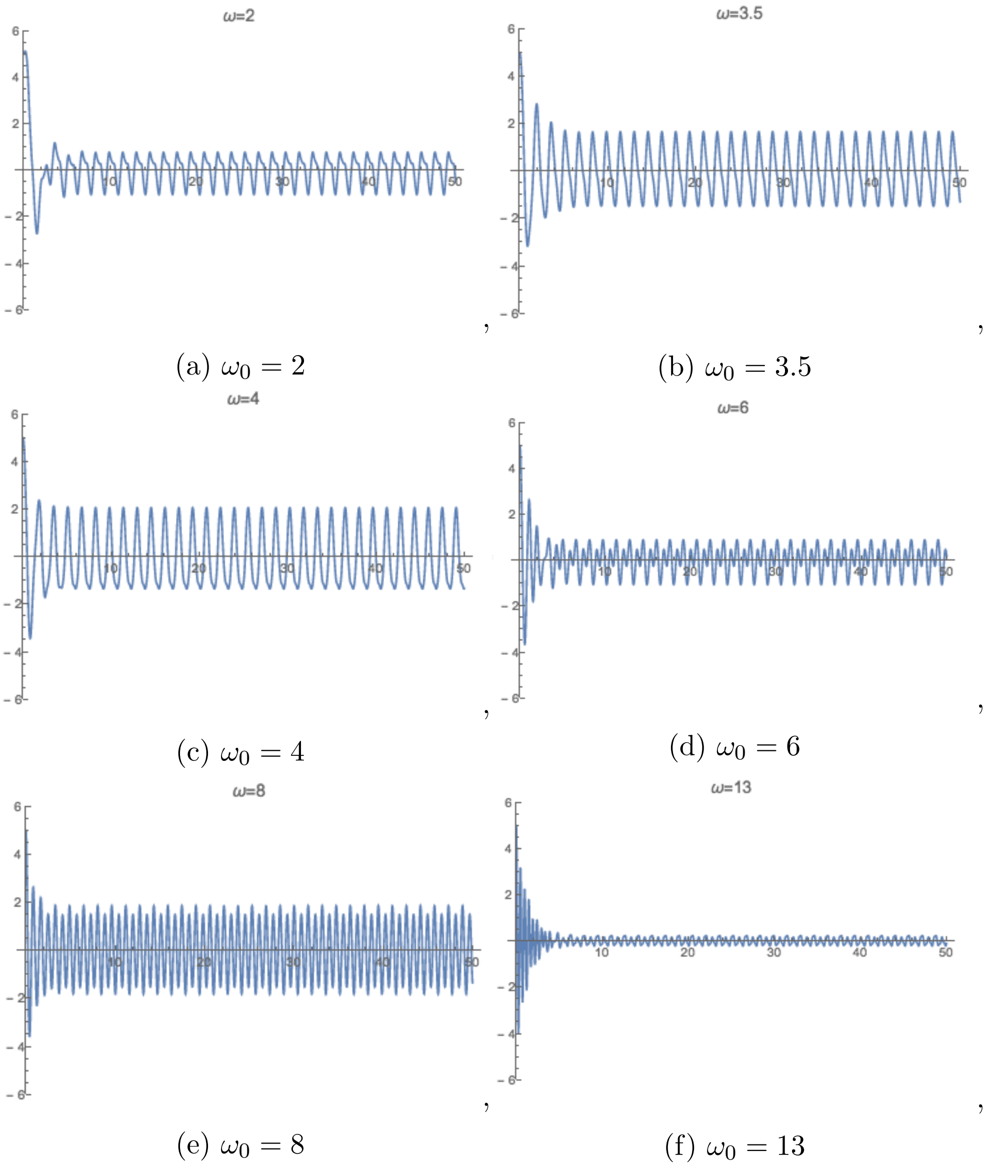
Figure 4.4.3: Final Filter Bank Response at $\omega = 13$

4.5 Testing more complex signals

So far the filter bank has only been applied to pure tone signals; notice that $\omega = 4$ and $\omega = 3.5$ respond almost identically with only slight variations and so would have overlapping critical bands. It should also be noted that higher frequencies still have slightly less amplitude response that lower frequencies. That is, when driven at their respective resonant frequencies $\omega = 13 < \omega = 6 < \omega = 3.5$

Realistically any model should be able to handle more complex signals, Figure 3.5.1 shows the filter bank's response to a wave given by

$$10Cos(4t) + 10Cos(8t) + 10Cos(20t)$$



(a) $\omega_0 = 2$

(b) $\omega_0 = 3.5$

(c) $\omega_0 = 4$

(d) $\omega_0 = 6$

(e) $\omega_0 = 8$

(f) $\omega_0 = 13$

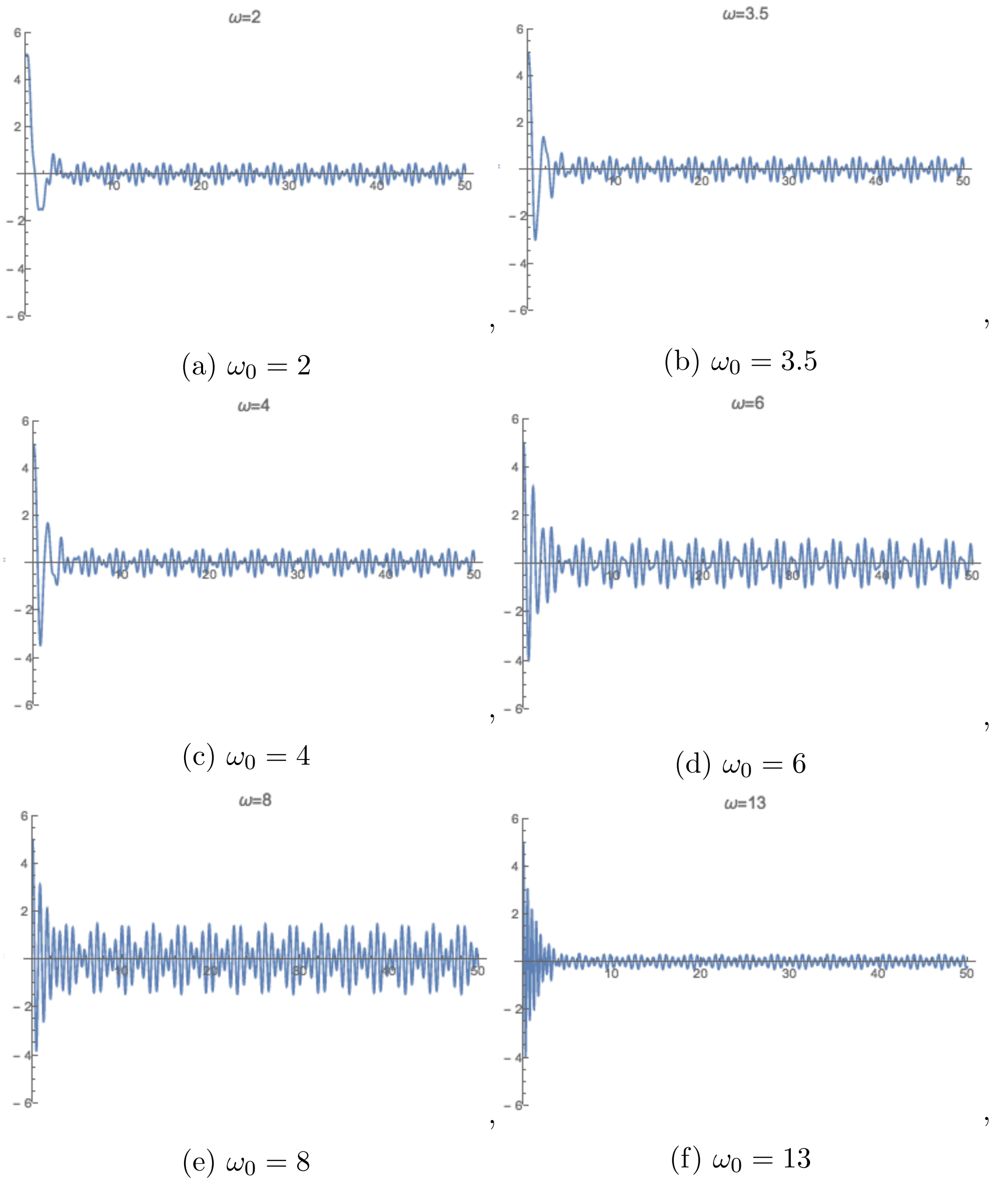
Figure 4.5.1

As expected the filters corresponding to $\omega_0 = 4, 8$ respond maximally with $\omega_0 = 3.5$ responding strongly as well. The other filters are mostly unaffected which is to be expected since there is no coupling between the filters. While there is no coupling, in this model several partials of complex wave can illicit a stronger response in a single filter. For example a signal of the form

$$10Cos(7t) + 10Cos(9t) + 10Cos(20t)$$

Should affect the $\omega_0 = 8$ filter the most

Figure 4.5.2 shows that the $\omega_0 = 8$ filter has the strongest response, i.e. it is close enough in bandwidth to get resonance from both the $\omega = 7$ and the $\omega = 9$ signal.



(a) $\omega_0 = 2$

(b) $\omega_0 = 3.5$

(c) $\omega_0 = 4$

(d) $\omega_0 = 6$

(e) $\omega_0 = 8$

(f) $\omega_0 = 13$

Figure 4.5.2

5
Beyond a Simple Filterbank

In this chapter we will examine the physical significance and accuracy of the linear filter bank model and how we might expand on it.

5.1 Interpretation

The results from the end of chapter 3 are meant to represent the strength of a particular frequency within a signal that is being sent to the brain. The amplitude of the response corresponds to the number of hair cells firing in a particular region of the BM. The number of hair cells firing tells the brain how strong that particular frequency is within the incoming signal while still picking up on some of the weaker spectral pieces. This is how we can register the overtones of a sound which give it a particular timbre even if we only hear a single tone. In this model the brain takes all the frequency inputs and compiles them into a single pitch based off the strength of each incoming signal. Since every bit of the BM will be excited to some extent by a particular sound wave, the brain has to look at the peaks in a region. Here each section of the BM responds only to the stimulus of the sound and is not affected by vibrations of other regions, i.e. there is no spatial coupling for nearby regions.

In the physical ear, the values for R, L and C are determined by the physical characteristics of the cochlea. L is the mass density of a particular section, R is the impedance of the endolymph fluid that acts as a friction coefficient to the BM's motion and C is the the stiffness of the BM. In hindsight it seems obvious that the major variations of the filter should be variations in the value of C rather than L as it is known that stiffness of the BM varies varies significantly from one end to the other.

The linear filter bank model could be improved to match some physical aspects of the ear which effect things like frequency dependent loudness or variations of the middle ear. For example, our concept of loudness is not only dependent on the sound pressure level of the sound wave but also on the frequency. Fig 5.1.1 shows the so called Fletcher Munsen curve, the necessary sound pressure level (SPL) for a particular frequency to be just audible. In general, the lower the value on this curve, the more sensitive people are to that particular frequency. Notice that the scale is logarithmic as is the distribution of critical bands along the cochlea.

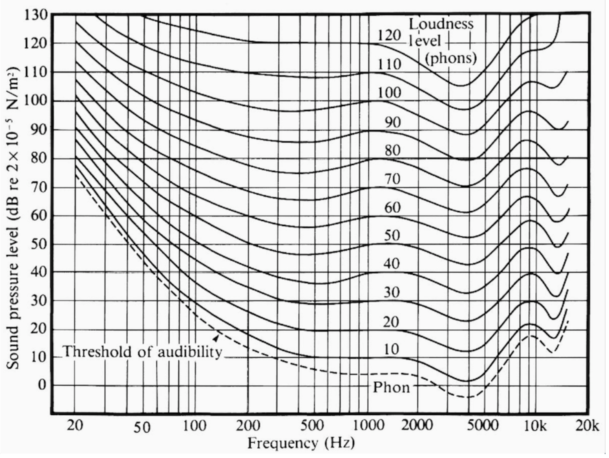


Figure 5.1.1: Fletcher Munsen Curve

Interestingly, the dip around 3400 Hz is not actually a psychological effect but is an amplification that comes from the ear canal's length. The meatus is about 25mm long is open at one end which gives it a resonance 3400 Hz so that frequencies around this value will be amplified before the they even reach the cochlea.

Adjustments could be made to the model using this data and weighing the filters correspondingly. Likewise the there are some mechanical effects of the middle ear's impedance transformation that were ignored and could be accounted for in this model. However, even with these adjustments, there are some deeper problems with the human-as-Fourier-analyzer picture.

5.2 Problems with the Filter Bank Picture

This picture of the ear as a Fourier analyzer was dominant in the mid nineteenth century and was popularized by Herman von Helmholtz and Georg Ohm. These two, as well as some of their contemporaries, used a model similar to the one I put forth in the last chapter. Although they used the analogy of taut strings instead of circuit elements the calculations yield the same result. As we found out in the previous section it is difficult to balance the quality factor such that there is a narrow bandwidth, making the model much rougher than an ideal Fourier analyzer. The lack of resolution in the model helped explain away some inconsistencies like roughness or the attack and decay of the sound

This model played into their larger theory of pitch perception which was summarized by Plack and Oxenham [?Oxenham]

> " (a) Among the many periodic vibrations within a given period, only those containing a nonzero fundamental partial evoke a pitch with related period;
>
> (b) other partials might also evoke pitches related to that period;
>
> (c) Relative partial amplitude affects the quality (timbre) of the vibration, but not its pitch, as long as the amplitude of the fundamental is not zero;
>
> (d) Relative phases of partials (up to a certain rank) affect neither quality nor pitch."

In other words, any partial that informs the pitch and is present in an incoming sound wave will produce a pitch related to the fundamental. For this model there must always be some fundamental present in the Fourier decomposition of that signal. It seems that Ohm and Helmholtz believed that partials contained in a sound only really affected the timbre of the fundamental pitch. To them, each partial produced their own pitch, if very faintly[?Heller]. This is true to some extent in that the relative amplitude of the partials does affect timbre of a sound and that a well trained ear can pick out some overtones that are present in the sound. It is also true that the phase of neuron firing tends to be locked into the phase of the incoming partials. The fundamental error comes from the assumption that any complex signal needs to have a nonzero fundamental present in the signal. It is, in fact, possible to have a combination of partials that add up to a pitch that has no associated fundamental. This phenomenon is commonly known as the 'Missing fundamental.' which is a slightly misleading name as the fundamental pitch isn't missing at all. It's exactly what is being heard. However there is no actual spectral component with fundamental's frequency. In light of this confusion some older literature refers to the phenomenon as a 'Residue pitch' which, in my opinion, doesn't make it simpler as it is not clear what is leaving the residue.

5.3 The Missing Fundamental

In general, the overtone pattern for a particular pitch comes in multiples of that pitch. Musically we know that for a pitch perceived at 200 Hz, which is between a G_3 and a $G_3^{\#}$, the first overtone will be a partial of 400 Hz which is an octave above the original note. The next overtone will be at 600 Hz and then 800 Hz and so on forever. Most notes that are produced physically will have several overtones present with their amplitudes dependent on how they were originally generated. The pitch of the note will be that of the fundamental, even with many overtones present. This is true even if some of the amplitudes of the overtones are as large as the amplitude of the fundamental.

As stated above, the phenomenon of the missing fundamental dominates even when it is not present in the sound! So a signal containing partials at 400Hz, 600Hz and 800Hz will still be perceived as something like a G_3-$G_3^{\#}$. This directly contradicts the theories of Helmholtz and Ohm. Helmholtz knew of this phenomenon but his faith in Fourier analysis was so great that he decided to stand by his theory and try to add in elements that would explain the phenomenon. He eventually settled on the idea that the fundamental could arise physically within the cochlea itself due to some non-linear wave interference. That is, that wave interference within cochlea could produce another real partial at the fundamental's frequency. There has been no real evidence to support this claim, although that is not to say that there are no non-linear interactions within the cochlea. There is fluid coupling, after all.

Helmholtz's work is the basis for almost all place theory. Place theory uses Fourier's theorem to look at the individual partials present and claim that each one has an associated pitch so that each pitch is processed individually. The modern theory of pattern recognition builds directly off of Place theory. It still describes a cochlea that focuses on individual partials instead of the full signal, however it assumes that the brain has a set of defined templates for pitch. For a particular pattern of partials the brain will assign a pitch based on what template it is most similar to, which could explain the missing fundamental. This approach is also concerned entirely with pure tone partials as they are the components of these mental templates. One issue with this approach is that it cannot account for unresolved partials, more on those later. [?Oxenham] [?Cheveigne]

Place theory was eventually challenged by William Rutherford (1886) who, among others, proposed the frequency or "Telephone" theory of pitch perception. In this approach the ear simply transmits any incoming vibrations to the brain exactly as they are, without any processing within the ear or the auditory nerve. This may sound familiar as it is exactly the mechanism that telephones and speakers use to transmit sound. This theory was eventually discredited as it was found experimentally that the hair cells were incapable of firing fast enough to replicate the higher frequencies that humans can hear. This problem was eventually corrected by volley theory which allows groups of hair cells to fire in sequence in order to produce a higher firing rate. This approach is still being investigated but I believe it has some inherent flaws that I hope will become clear. [?Oxenham]

The other modern model is that of time theory, this approach was much less popular in Helmholtz's day but is now a dominant theory of pitch perception. Temporal theories focus on events (or spikes) occurring within the ear. They compare the signals over some time interval within the ear. The temporal method shifts the focus from the properties of individual partials to properties of the complete waveform. This addresses the issue of the missing fundamental since a signal of harmonic overtones will have the same period wether or not the fundamental is present. The initial difficulty with time theories was that it is difficult to define what the 'event' should be to base the theory on. For a pure tone one could pick easily choose the largest peak or the zeroes to characterize the period. With complex waves, the choice is more difficult since the waveform could be radically different and still produce the same pitch. In the mid 1950's Joseph Licklider came across a method that addressed this issue: Autocorrelation.

6
Autocorrelation

Autocorrelation is essentially a measure of self similarity; it compares a function to itself at some later time. More precisely it compares some point in $f(t)$ to a later point $f(t+\tau)$. Here τ is the 'jump' or 'shift' that will be compared to the original point. The idea is to check for self similarity over a range of jumps as a way to look for periodic behavior. A simple test for self similarity is through a difference function of the form

$$d(\tau) = \frac{1}{2}\int_{t_1}^{t_2} [f(t) - f(t-\tau)]^2 dt \tag{6.0.1}$$

The function will be zero when τ=0 and again if the function is periodic, so when $f(t+\tau) = f(t+T)$. The integral checks every possible τ value on the time interval and the square keeps all the responses positive. So for a simple periodic function such as $Cos(2\pi\cdot 100t)$, with a period of 10 milliseconds, the difference function is zeroed whenever $\tau = n10ms$ (fig 5.4.1)

The bounds of integration must be large enough that τ can range over at least one full period and ideally it will range over several periods. The range of this integral places a delay into the model since a pitch cannot be perceived for several periods of an incoming signal. However, even at lower frequencies the delay can be ignored when compared to the time limit imposed by the physical response of the hair cells.

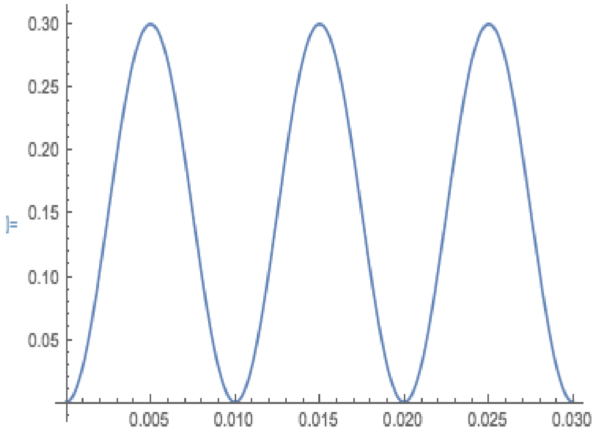


Figure 6.0.1: simple difference function

The other approach is for the autocorrelation function to spike when τ is equal to the period instead of falling to zero. These autocorrelation functions take the form

$$s(t) = \int_{t_1}^{t_2} f(t) * f(t+\tau)dt \tag{6.0.2}$$

So that $s(t)$ is maximized when τ is zero or nT. There is some debate as to which form of autocorrelation is better suited to handling nonlinear effects that may arise in more complicated models. However, for this project I will make the somewhat arbitrary choice to favor the second type, as it is what the early adopters tended to use.

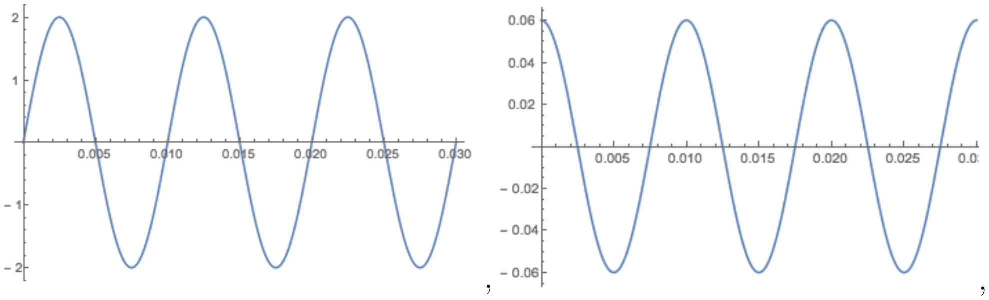
The autocorrelation approach also does a good job at dealing with the the higher, *unresolved* harmonics. By unresolved I mean that the particular parts of the BM with a strong amplitude response are not separated by a region of low amplitude response. This happens when several critical bands overlap and so tend to produce more ambiguous pitch. Unresolved harmonics are generally higher order harmonics of a signal since the spacing of critical bands is much closer together towards the apex of the cochlea. For example a complex 20 0Hz signal might have distinct partials at 200 Hz, 400 Hz and 800 Hz would be resolved partials with very little BM response between them but at higher frequencies, say 2000-3000 Hz there may be much less distinct BM responses. These unresolved harmonics will add a lot of noise into the pattern matching model, but even the unresolved harmonics can be autocorrelated individually to gain information about their period and the overall waveform.

For a similar reason, autocorrelation does a much better job of addressing phase sensitivity as compared to earlier temporal models. This is because the signal will be analyzed over several periods, so any phase factor will be irrelevant. Phase does become an issue when dealing with unresolved harmonics. Since the unresolved harmonics essentially share neural channels any phase factor will actually change how signal being processed.

These two factors seem to set autocorrelation above the other modern theories like pattern matching and volley theory. However, in the next section I will be comparing it to the much simpler Fourier approach of Helmholtz.

6.1 A simple ACF model

For a pure tone the ACF proposed by Licklider looks very similar to a cosine function graphed in time, but of course it is really being graphed with respect to τ rather than t. For example, compare the plot of $Sin(100t)$ to its autocorrelation function



(a) Pure Sine wave

(b) Autocorrelation

Figure 6.1.1

Notice the autocorrelation peaks when $\tau = T$ and the sine is zero when $t = T$ The original model proposed this type of response for each individual hair cell and was later modified so that it describes each channel of hair cells corresponding to a specific critical band.

For pure tones the autocorrelation has no discernible advantage to the filter bank/Fourier transform approach. However, given a complex signal autocorrelation gives a better indication of the pitch properties of the signal as a whole. For a set of harmonic overtones, both with and without the fundamental, we can compare the autocorrelation function method to Helmholtz's Fourier analyzer picture.

Take the series of harmonic overtones for 100 Hz, with each overtone at equal amplitude

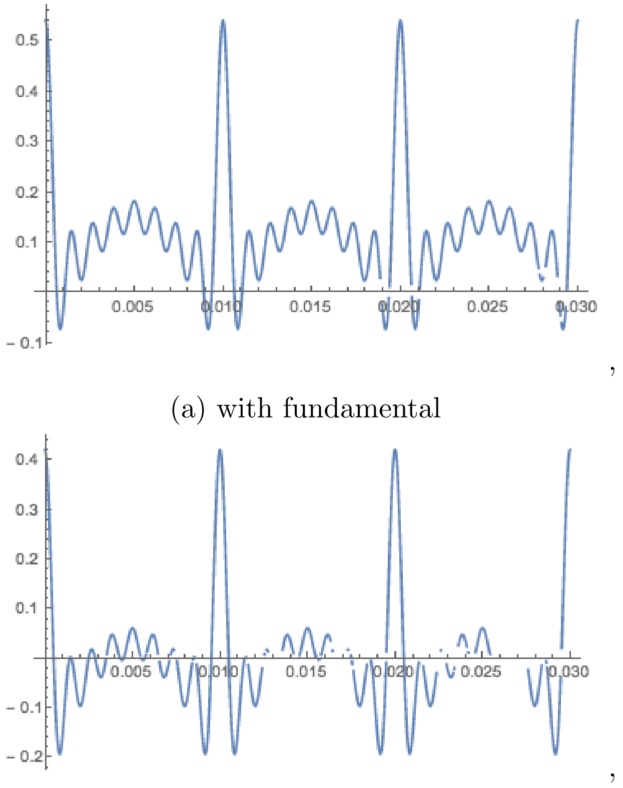
$$g(t) = 2Cos(2\pi\cdot 100t) + 2Cos(2\pi\cdot 200t) + 2Cos(2\pi\cdot 300t) + 2Cos(2\pi\cdot 400t)$$

$$+2Cos(2\pi\cdot 500t) + 2Cos(2\pi\cdot 600t) + 2Cos(2\pi\cdot 700t) + 2Cos(2\pi\cdot 800t)$$

Applying the autocorrelation function to $g(t)$ we

$$d(\tau) = \int_{t_1}^{t_2} [g(t) - g(t+\tau)]^2 dt \qquad d'(\tau) = \int_{t_1}^{t_2} [g'(t) - g'(t+\tau)]^2 dt \tag{6.1.1}$$

Where $g'(t)$ is just $g(t)$ with the first partial removed. Fig 5.5.1 shows the result of the autocorrelation functions for both $g(t)$ and $g'(t)$

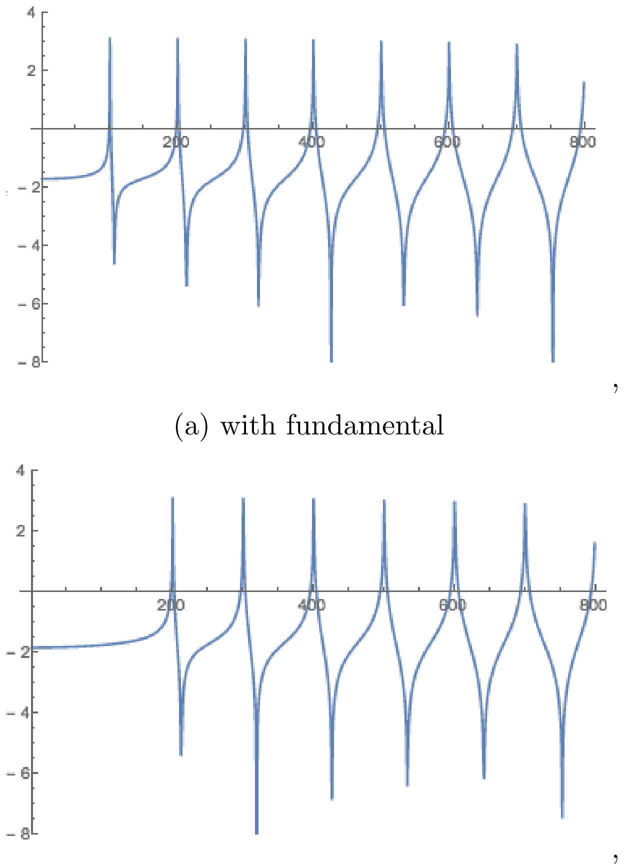


(a) with fundamental

(b) without fundamental

Figure 6.1.2: Harmonic series for 200 Hz tone

Although the graphs differ slightly, they have the same inter-spike interval regardless of whether or not the fundamental was actually present. Since in this model, the inter-spike interval determines the period of the pitch being perceived, both signals would produce the same pitch. If instead we took the Fourier transform of the signals we can see the clear difference (fig 5.5.2).



(a) with fundamental

(b) without fundamental

Figure 6.1.3: Harmonic series for 200 Hz tone

The Fourier transform misses the fundamental frequency of 100Hzl which is the dominant tone being perceived!

6.2 Implementing autocorrelation a la Meddis and Hewitt

6.2.1 Structure of Meddis and Hewitt Model

Licklider's autocorrelation model was eventually implemented by Ray Meddis and Michael J. Hewitt in their 1990 paper "Virtual pitch and phase sensitivity of a computer model of the auditory periphery: I Pitch Identification". At the time pattern matching theories were popular but were still unable to extract pitch from unresolved harmonics; however it was already known that pitch extraction could occur even when dealing entirely with unresolved harmonics. The necessity of unresolved harmonics was an initial cue to turn to autocorrelation. Where Licklider had looked at autocorrelation for individual hair cells, Meddis and Hewitt looked at autocorrelation for individual hair cell channels and then summed the results over the full range of the cochlea.

Their model began with a series of filter banks. The first filter bank was implemented to account for the amplification due to the ear canal, as discussed in section 4.1 and the second was used to recreate the pressure gain due to the ossicles of the middle ear. Next they put their signal through a filter bank to simulate the mechanical frequency sorting of the BM. Instead of a series of RLC circuits they used a series of 128 gamma-tone filters of the form

$$g(t) = at^{n-1}e^{-2\pi bt}cos(2\pi ft + \phi) \tag{6.2.1}$$

which can be tuned to amplify specific frequencies in much the same way as in chapter 3.

Instead of sending the selected frequencies directly onto the brain for processing they modeled the probability of a transmitter fluid spike for each channel. This probability was based on how much transmitter fluid was released from the hair cells firing, the refractory and re-uptake periods of the hair cells and the amount of free transmitter fluid that would be present near the hair cell channel. This probability function $p(t)$ was used as the basis function for each channel's autocorrelation function and is directly related the channel's amplitude response.

$$h(t, \tau) = p(t)p(t - \tau)dt \tag{6.2.2}$$

Where $p(t)$ is the probability and τ ranges from 0.05ms to 16.647ms. For this computational model the functions sum instead of integrating. Meddis and Hewitt also imposed a time constant Ω to kill the response after several periods. So for any particular channel k and period T, the autocorrelation function is

$$h(t, \tau, k) = \sum_{n=1}^{\infty} p(t - T)p(t - T - \tau)e^{\frac{-T}{\Omega}}dt \tag{6.2.3}$$

With $T = ndt$, so that the contribution at time t from $p(t - T)$ diminishes for larger T. Meddis and Hewitt used the value $\Omega = 2.5$ms that was proposed by Licklider.

Finally, these functions are summed and averaged

$$s(t, \tau) = \sum_{k=1}^{128} \frac{h(t, \tau, k)}{128} \tag{6.2.4}$$

To get the final signal from which the pitch is extracted.

6.2.2 Results and Implications

Running sound signals through this model the pitch extraction works well with and without a fundamental. In section 5.5 I demonstrated a crude model of one of these summary ACFs. I only tested the ACF with a simple series of resolved harmonic overtones. With the greater range of filters available for the Meddis and Hewitt's model, they were able to test for pitch perception at a greater range of frequencies including unresolved harmonics that appear at higher frequencies. They found that the peaks tend to smear out around the fundamental period but still show a clear pattern under autocorrelation.

6.3 Where does the Meddis and Hewitt model leave pitch perception

The autocorrelation model proposed by Meddis and Hewitt is still fairly relevant to the study of pitch perception, with different authors arguing for different formulations of the autocorrelation functions and others defending the pattern matching model. The strength of autocorrelation is that, unlike pattern matching models, it still produces pitch for unresolved harmonics. This is also a major flaw of the theory since it tends to produce pitch for unresolved harmonics too well, giving more definite pitch than has been found experimentally.

The next step is to test these models vigorously and hold them up to experiment. After all, these are only models, and they are only as good as the results that they produce.

6.4 Nonlinear Complications

There are several factors that were largely ignored in the previous discussion that point to some nonlinearity in the traveling wave of the BM. This nonlinearity could be affecting the results and may account for some of the theory's inconsistencies. Below are some examples of auditory phenomena that are not encompassed in the theory of pitch perception so far. They may be cues to nonlinearity or due to some other factor entirely.

6.4.1 Beating

When two tones of almost identical frequency are played together, a listener will hear a tone at the average of the two frequencies. This tone's perceived loudness will oscillate from high to low at a frequency determined by how different the two initial frequencies are. This effect is known as a *beating* and only occurs for frequencies separated by about 15 Hz or less. The narrow band in which beat frequencies are perceived leads me to believe that they may simply be the result of spatial coupling between nearby regions of the cochlea. This is not to be confused with the missing fundamental, for which the tones are harmonically related and there are enough overtones to characterize the summary ACF.

6.4.2 Difference Tones

Difference tones arise when two, generally non-harmonic, tones are played at the same time. If the tones are distinct enough in frequency then the listener will perceive a third pitch that is the difference between the two tones. i.e. $f_d = f_1 - f_2$. These are not be confused with the missing fundamental in which the various tones are harmonically related. Difference tones may arise from wave interference within the cochlea and there are various approaches to modeling it [?CM]

6.4.3 Oto-acoustic Emissions

Oto-acoustic emissions are faint sounds that are produced from somewhere inside the ear. The can be stimulated by sending tones into the ear which sends out a tone that is dependent on the input frequency. Opinions are divided on what causes this effect [?CM] [?Heller], but it certainly cannot be explained simply by the models I have discussed.

7
Conclusion

Pitch perception is a difficult subject to approach as it draws from so many fields of study. Especially when constructing complicated models of the inner ear's function it can be hard to remember that however complex the analysis is, there is more processing to be done "up the line." When I began working on this project I was still unsure as to how much of the process was really present in the inner ear. I knew that at the very least the frequency and most likely the amplitude of the sound wave had to be encoded into neural signals. I knew that for a complex wave there must be some kind of decoding so that we could register things like timbre.

One of the major benefits of autocorrelation, as opposed to pattern matching, is that at least some of our pitch occurs within the ear itself. For each channel, the signal is sent based on autocorrelation, not just a 1-to-1 response. These signals are then together within the auditory nerve. This means that at least some of the processing is occurring within the ear. This is a clue that an irregularity like the missing fundamental is not an illusion in the sense that it is not mistake of our nervous system. It is an example of our auditory system functioning exactly as it is meant to. There are still large gaps in this theory that may be answerable with a bit of mathematical rigor. [?CM]

My early focus on the linear filter bank may have slowed me down to some extent but in the process I gained some insight into the physicality of the model as it relates to the Basilar membrane, as well as its limitations. The familiarity of that approach tempted me to extend the circuit analogy to correct some of its flaws. The thought was that by extending the analogy and adding in more circuit elements I could account for some spatial coupling between different regions of the BM. This idea came from some fairly recent literature but it was mostly to correct smaller nonlinearities of the BM's response. I find it unlikely that any spatial coupling could explain a phenomenon like the missing fundamental which has to do with several harmonically related regions across the length of the cochlea.

This realization is what prompted the shift in focus to autocorrelation. Although I was not able to recreate an autocorrelation model to the sophistication of Meddis and Hewitt, I was able to gain a solid foundation in the tools and implications of autocorrelation as a path to pitch. The focus on autocorrelation came later on in the project but once I started to investigate it I was hooked. It just *resonated* with me.

Bibliography

[1] Hendrikus Duifhuis, *Cochlear Mechanics*, Springer, New York, NY, 2012.

[2] Eric J Heller, *Why You Hear What You Hear: An Experiential Approach to Sound, Music, and Psychoacoustics*, Princeton University Press, Princeton, NJ, 2013.

[3] Christopher J. Plack, Richard R. Fay, Andrew J. Oxenham, and Arthur N. Popper, *Pitch, neural coding and perception*, Springer, New York, NY, 2005.

[4] Ray Meddis and Michael Hewitt, *Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: pitch identification*, journal of the acoustical society of america **89** (1991), 2866–2873.

[5] Luis Robles and Mario A. Ruggero, *Mechanics of the Mammalian Cochlea*, Physological Reviews **81.3** (2001), 1305.

[6] Guangjian Ni, Stephen J. Elliott, Mohammad Ayat, and Paul D. Teal, *Modeling Cochlear Mechanics,* BioMed Research International **2014** (2014), 42 pages.

[7] Licklider J.C.R., *An optimum processor theory for the central formation of the pitch of complex tones*, journal of the acoustical society of america **23** (1951).

[8] Alain Cheveigne, *Pitch perception-a historical review*, CNRS-Icram, Paris, France (2004).

[9] *Pitch perception and grouping*, available at `http://www.isle.illinois.edu/sst/courses/ece538/MIT_pitch.pdf`.

[10] *Combination Tones*, available at `https://ccrma.stanford.edu/CCRMA/Courses/150/combination_tones.html`.

[11] Julius L. Goldstein, *An optimum processor theory for the central formation of the pitch of complex tones*, journal of the acoustical society of america **54** (1973), 1496–1516.