Introduction

Concepts covered in these notes

These notes explain the following ideas related to linear systems theory.

1. the challenge of characterizing a complex systems
2. simple linear systems
   (a) homogeneity
   (b) superposition
3. shift-invariance
   (a) decomposing a signal into a set of shifted and scaled impulses
   (b) the impulse response function
   (c) use of sinusoids in analyzing shift-invariant linear systems
   (d) decomposing stimuli into sinusoids via Fourier Series
   (e) characterizing a shift-invariant system using sinusoids
4. Examples
   (a) the stereo systems
   (b) the swinging pendulum
Overview

Characterizing the complete input-output properties of a system by exhaustive measurement is usually impossible. When a system qualifies as a linear system, it is possible to use the responses to a small set of inputs to predict the response to any possible input. This can save the scientist enormous amounts of work, and makes it possible to characterize the system completely.

Representing the input: graphing waveforms; impulses

Step one is to understand how to represent possible inputs to systems. Imagine a picture that shows the structure of the physical stimulus reaching your ear. On the horizontal axis we have time, and on the vertical axis we will plot the instantaneous density of the air molecules at your ear. Thus, we plot signal strength as a function of time. In the case of a simple hand-clap, the disturbance is a short, transient burst and is aptly named an impulse. It looks like a single upwards blip on the graph: the sound pressure momentarily increases when the clap hits your ear. More complex sounds look like more complex graphs on this kind of plot. This sort of graph offers a general way to describe all of the possible auditory stimuli.

The problem: Generalizing about a system’s overall responsiveness given responses to a small set of inputs

One possible way to characterize the response of the ear to sound might be to build a look-up table: a table that shows the exact neural response for every possible auditory stimulus. Obviously, it would take an infinite amount of time to construct such a table, because the number of possible sounds is unlimited.

Instead, we must find some way of making a finite number of measurements that allow us to infer how the system will respond to other stimuli that we have not yet measured. We can only do this for certain kinds of systems with certain properties. If we have a good theory about the kind of system we are studying, we can save a lot of time and energy by using the appropriate theory about the system’s responsiveness. Linear systems theory is a good time-saving theory for linear systems which obey certain rules. Not all systems are linear, but many important ones are.
Linear systems theory: a powerful way to generalize about input-output properties of systems

To see whether a system is linear, we need to test whether it obeys certain rules that all linear systems obey. One basic test of linearity is the scalar rule; the more general test is that of superposition.

**Scalar rule**  As we increase the strength of a simple input to a linear system, say we double it, then we predict that the output function will also be doubled. For example, if a person’s voice becomes twice as loud, the ear should respond twice as much if it’s a linear system. This is called the scalar rule or sometimes the homogeneity of linear systems. Clearly, systems that obey Steven’s Power Law do not obey the scalar rule and are not linear, because they show response compression or response expansion.

**Superposition**  The scalar rule is a special, limited case of a more general principle called superposition, which is the defining characteristic of linear systems. Suppose we present a complex stimulus S1 such as the sound of a person’s voice to the inner ear, and we measure the electrical responses of several nerve fibers coming from the inner ear. Now, we present a second stimulus S2 that is a little different: a different person’s voice. The second stimulus also generates a set of responses which we write down. Finally, we perform the test to determine whether the system is linear. We present the superposition of the two stimuli S1 + S2: we present both voice together and see what happens. If the system is linear, then the measured response of each fiber will be just
the sum of its responses to each of the two stimuli presented separately. The Scalar Rule applies to a single kind of response (like the response of a single nerve fiber); the superposition principle applies more generally to a set of responses, where each response individually obeys the scalar rule (like the set of responses across a bundle of nerve fibers). Systems that satisfy this Principle of Superposition are considered to be linear systems.

**Why impulses are special** The scalar law and superposition are very important. They suggest that the system’s response to a simple impulse can be the key measurement to make. The trick is to conceive of the complex stimuli we encounter (such as a person’s voice) as the combination of a few simple stimuli (such as simple monotone sine waves). For linear systems, we can measure the system’s response to simple stimuli and we know how to predict the response to combinations of these stimuli through the principle of superposition. The impulse is particularly important because we can approximate any complex stimulus as if it were simply the sum of a number of impulses that are scaled copies of one another and shifted in time. (A digital compact disc, for example, stores whole complex pieces of music as lots of simple numbers representing very short impulses, and then the CD player adds all the impulses back together one after another to recreate the complex musical waveform.)

**Shift-invariance** Suppose that we stimulate your ear once with an impulse and we measure the electrical response, and then stimulate it again with a similar impulse at a different point in time, and again we measure the response. The response to the second
impulse will probably be the same as the response to the first impulse, except that it is shifted later in time. When the responses to the identical stimulus presented shifted in time are the same, except for the corresponding shift in time, then we have a special kind of linear system called a *shift-invariant* linear system. Just as not all systems are linear, not all linear systems are shift-invariant.

If the system is shift-invariant, then we do not need to measure the response to each kind of impulse at different points in time. Rather, the response to every impulse of a particular intensity, no matter which point in time, is the same. To characterize shift-invariant linear systems, then, we need to measure only one thing: the way the system responds to an impulse of a particular intensity. This response is called the *impulse response function* of the system.

The problem of characterizing a complex system has become simpler now. For general linear systems, we only need to measure the system’s response to various impulses at different points in time. For shift-invariant linear systems it’s even simpler: there is only a single impulse response function to measure. Once we’ve measured this function, we can (in principle) predict how the system will respond to any other possible stimulus.

The way we use the impulse response function is illustrate in Figure 1. We conceive of the input stimulus, in this case a sinusoid, as if it were the sum of a set of impulses. We know the response to an impulse scaled by each amount (i.e., the impulse response). We can predict the system’s response to each impulse. We add together all of the impulses to predict how the system will respond to the complete stimulus.
Sinusoidal stimuli

Sinusoidal stimuli have a special relationship to shift-invariant linear systems. A sinusoid is a regular, repeating curve, that oscillates around a mean level. The sinusoid has a zero-value at time zero. The cosinusoid is a shifted version of the sinusoid; it has a value of one at time zero.

The sine wave repeats itself regularly, and the distance from one peak of the wave to the next peak is called the wavelength of the sinusoid and generally indicated by the greek letter λ. The inverse of wavelength is frequency: the number of peaks in the stimulus that arrive per second at the ear. The units for the frequency of a sine-wave are hertz, named after a famous 19th century physicist, who was a student of Helmholtz. The longer the wavelength, the shorter the frequency; knowing one we can infer the other. Apart from frequency, sinusoids also have various amplitudes, which represent the distance between how high their energy gets at the peak of the wave and how low it gets at the trough. Thus, we can describe a sine wave completely by its frequency and by its amplitude. Loud, high-pitched sounds have high frequency and high amplitude.

When we write the mathematical expression of a sine-wave, the two mathematical variables that correspond to the frequency and the amplitude are A and f as in

\[ A \sin(2\pi ft) \]  

(1)

As the value of the amplitude, A, increases the height of the sinusoid increases. As the frequency, f, increases, the spacing between the peaks becomes smaller.

The response of shift-invariant systems to sine waves Just as we can express any stimulus as the sum of a series of shifted and scaled impulses, so too we can express any periodic stimulus (a stimulus that repeats itself over time) as the sum of a series of scaled sinusoids and cosinusoids at different frequencies. This expression is called the Fourier Series expansion of the stimulus. The equation describing this expansion works as follows. Suppose that \( s(t) \) is a periodic stimulus. Then we can always express \( s(t) \) as

\[ s(t) = \frac{a_0(s)}{2} + \sum_{j=0}^{\infty} A_j \sin(2\pi ft) + \sum_{j=0}^{\infty} B_j \cos(2\pi ft) \]  

(2)

(Do not memorize this equation!) You can go either way: if you know the coefficients (\( A_j \) and \( B_j \)), you can reconstruct the original stimulus \( s(t) \); if you know the stimulus, you can compute the coefficients by a method called the Fourier Transform (a way of decomposing complex stimuli into sinusoids and cosinusoids).

This decomposition is important because if we know the response of the system to
Fourier Series Approximations

![Fourier Series Approximations](image)

Figure 1: Fourier Series approximation of a squarewave as the sum of sinusoids.

sinusoids and cosinusoids at many different frequencies, then we can use the same kind of trick we used with impulses to predict the response via the impulse response function. First we compute the values of the coefficients in the Fourier Series expansion, using the Fourier Transform for a complex stimulus. Then, if we know the system’s response to all possible sinusoids, we can simply sum together the responses for all the sinusoids that compose the complex stimulus, and we’ll know what the system’s response to that complex stimulus itself will be.

However, measuring a system’s responses to a lot of sinusoids and cosinusoids at different frequencies may be very time-consuming. Making these measurements can be much harder than measuring the impulse response function. What if the system’s response to a sinusoid was very complex and hard to describe? Fortunately, it isn’t for shift-invariant linear systems! When we use a sinusoidal or cosinusoidal stimulus as input to a shift-invariant linear system, the system’s responses is always a weighted sum of a sinusoid and cosinusoid at the same frequency as the input. That is, when the input is \(\sin(2\pi ft)\) the output is always of the form \(A_f \sin(2\pi ft) + B_f \cos(2\pi ft)\). Thus, measuring the response to a sinusoid or a cosinusoid for a shift-invariant linear system entails measuring only two numbers, \(A_f\) and \(B_f\). This makes the job of measuring the response to sinusoids and cosinusoids at many different frequencies quite practical.

Often, then, when scientists characterize the response of a system they will not tell you the impulse response. Rather, they will give you plots that tell you about the values of
Shift-Invariant Linear Systems and Sinusoids

Figure 2: Characterizing a system using its frequency response.

\( A_f \) and \( B_f \) for each of the possible input frequencies. This representation of how the shift-invariant system will perform is equivalent to providing you with the impulse response function. We can use these numbers to compute the response to any stimulus. This is the main point of all this stuff: a simple, fast, economical way to measure the responsiveness of complex systems that are shift-invariant. If you know the coefficients of response for sine waves at all possible frequencies, then you can determine how the systems will respond to any possible periodic stimulus.

Example 1: Stereos as shift-invariant systems

Many people find the characterization in terms of frequency response to be intuitive. And most of you have seen graphs that describe performance this way. Stereo systems, for example, are pretty good shift-invariant linear systems. Their can be evaluated by measuring the signal at different frequencies. And the stereo controls are designed around the frequency representation. Adjusting the bass alters the level of the low frequency signals, while adjusting the treble adjust the level of the high frequency signals. Equalizers divide up the signal band into many frequencies and give you finer control.
**Linear Systems Logic**

**Space/time method**
- Measure the impulse response
- Express as sum of scaled and shifted impulses
- Calculate the response to each impulse
- Sum the impulse responses to determine the output

**Frequency method**
- Measure the sinusoidal responses
- Express as sum of scaled and shifted sinusoids
- Calculate the response to each sinusoid
- Sum the sinusoidal responses to determine the output

**Figure 3:** Alternative methods of characterizing a linear system.

**Example 2: Swinging pendula as frequency analyzers**

Remember the class demonstration with two weights suspended from strings of different lengths. Each weight on a string is a pendulum. The longer the string’s length, the lower the pendulum’s natural frequency, also called it resonant frequency. By moving the stick back and forth at different frequencies corresponding to one or the other pendulum’s resonant frequency, we can get one or the other weight to move back and forth.

The swinging pendula act as frequency analyzers: they inform us about the motion of the stick and the hand that moves it. When the short pendulum swings a lot you can infer that the stick is moving at a relatively high frequency back and forth. When the long pendulum swings a lot you can infer that the stick is moving at a relatively low frequency. Thus, which pendulum is moving identifies the nature of the stick movement. In general, this sort of pendulum motion satisfies the principle of superposition and therefore the system that transforms from the input of the stick moving, to the output of the pendula swinging, is a linear system.

The cochlea transforms the complex periodic motion of the input to the simple oscillation of different frequency analyzers in the cochlea. Points along the cochlea respond based on the frequency components in the signal. This representation is used by the brain to analyze sounds for significance and meaning.