Linear Systems theory Handout

Professor David Heeger

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.

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

- The challenge of characterizing a complex systems
- Simple linear systems
  - Homogeneity
  - Additivity
  - Superposition
- Shift-invariance
  - Decomposing a signal into a set of shifted and scaled impulses
  - The impulse response function
  - Use of sinusoids in analyzing shift-invariant linear systems
  - Decomposing stimuli into sinusoids via Fourier Series
  - Characterizing a shift-invariant system using sinusoids
- Examples
  - Stereo
  - Swinging pendulum

Systems, Inputs, and Responses

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.

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.
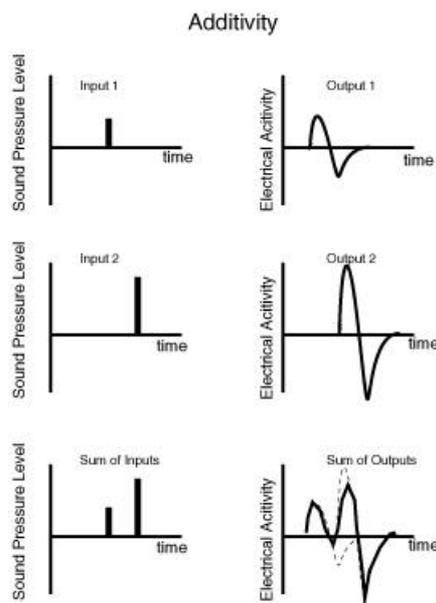

Linear Systems

To see whether a system is linear, we need to test whether it obeys certain rules that all linear systems obey. The two basic tests of linearity are homogeneity and additivity.

Homogeneity: 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 homogeneity or sometimes the scalar rule of linear systems. Clearly, systems that obey Steven's Power Law do not obey homogeneity and are not linear, because they show response compression or response expansion.
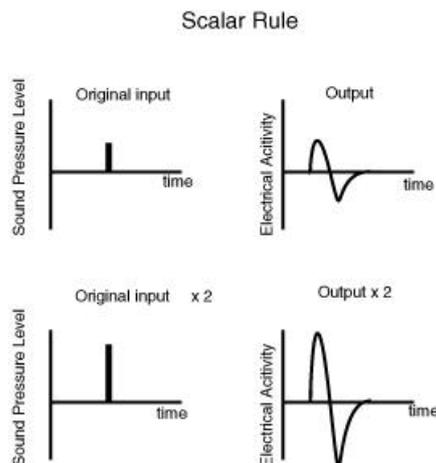

Additivity: Suppose we present a complex stimulus S1 such as the sound of a person's voice to the inner ear, and we measure the electrical responses of several nerve fibers coming from the inner ear. Next, we present a second stimulus S2 that is a little different: a different person's voice. The second stimulus also generates a set of responses which we measure and write down. Then, we present the sum of the two stimuli S1 + S2: we present both voices together and see what happens. If the system is linear, then the measured response will be just the sum of its responses to each of the two stimuli presented separately.

Superposition: Systems that satisfy both homogeneity and additivity are considered to be linear systems. These two rules, taken together, are often referred to as the principle of superposition.

Shift-invariance: Suppose that we stimulate your ear once with an impulse (hand clap) and we measure the electrical response. Then we stimulate it again with a similar impulse at a different point in time, and again we measure the response. If we haven't damaged your ear with the first impulse then we should expect that the response to the second impulse will be the same as the response to the first impulse. The only difference between them will be that the second impulse has occurred later in time, that is, it is shifted 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.
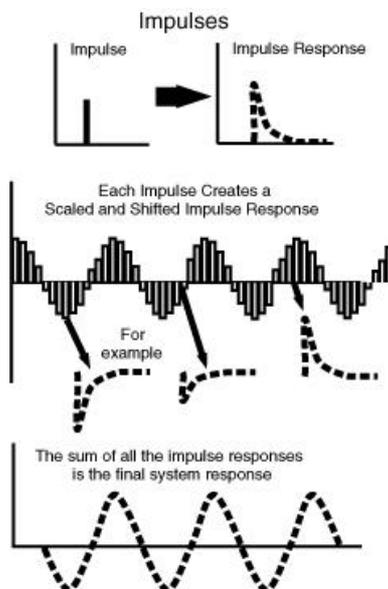
Shift-Invariance Rule
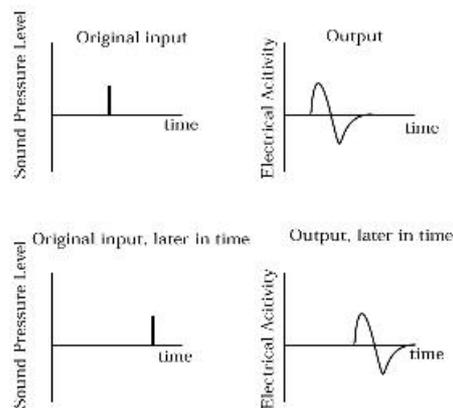
Why impulses are special: Homogeneity, additivity, and shift invariance may, at first, sound a bit abstract but they are very useful. They suggest that the system's response to an impulse can be the key measurement to make. The trick is to conceive of the complex stimuli we encounter (such as a person's voice) as a combination of impulses. 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.)

For shift-invariant linear systems, we can measure the system's response to an impulse and we will know how to predict the response to any stimulus (combinations of impulses) through the principle of superposition. 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 shift-invariant linear systems, there is only a single impulse response function to measure. Once we've measured this function, we can predict how the system will respond to any other possible stimulus.

Impulses

Impulse   Impulse Response

Each Impulse Creates a
Scaled and Shifted Impulse Response

For
example

The sum of all the impulse responses
is the final system response

The way we use the impulse response function is illustrated in the above Figure. 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 responses we would get if each impulse was presented separately (i.e., scaled and shifted copies of the impulse response). We simply add together all of the (scaled and shifted) impulse responses 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. The distance from one peak of the wave to the next peak is called the *wavelength* or *period* of the sinusoid and it is generally indicated by the greek letter lambda. The inverse of wavelength is frequency: the number of peaks in the stimulus that arrive per second at the ear. The longer the wavelength, the lower the frequency. The units for the frequency of a sine-wave are hertz, named after a famous 19th century physicist who was a student of Helmholtz. Apart from frequency, sinusoids also have various amplitudes, which represent how high they get at the peak of the wave and how low they get 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 amplitude and the frequency are A and f, respectively:

$$A \sin(2 \text{ pi} f t)$$

The height of the peaks increase as the value of the amplitude, A, increases. The spacing between the peaks becomes smaller as the frequency, f, increases.

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 (shifted and scaled) sinusoids at different frequencies. This 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 a sum of sinusoids:

$$s(t) = A0 + A1 \sin(2\ p\ f1\ t + f1) + A2 \sin(2\ pf2\ t + f2) + A3 \sin(2\ pf3\ t + f3) + ...$$

(Do not memorize this equation!) You can go either way: if you know the coefficients (the *A*'s and *f*'s), 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 its component sinusoids).

Fourier Series Approximations
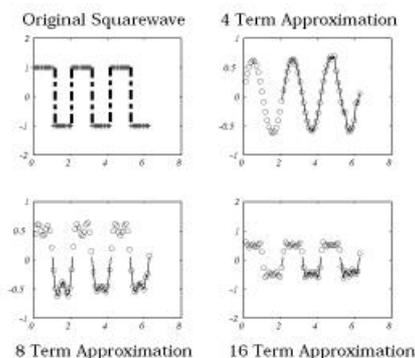
This decomposition is important because if we know the response of the system to sinusoids at many different frequencies, then we can use the same kind of trick we used with impulses to predict the response via the impulse response function. First, we measure the system's response to sinusoids of all different frequencies. Next, we take our input stimulus (a complex sound) and use the Fourier Transform to compute the values of the coefficients in the Fourier Series expansion. At this point the stimulus has been broken down as the sum of its component sinusoids. Finally, we can predict the system's response to the (complex) stimulus simply by adding the responses for all the component sinusoids.
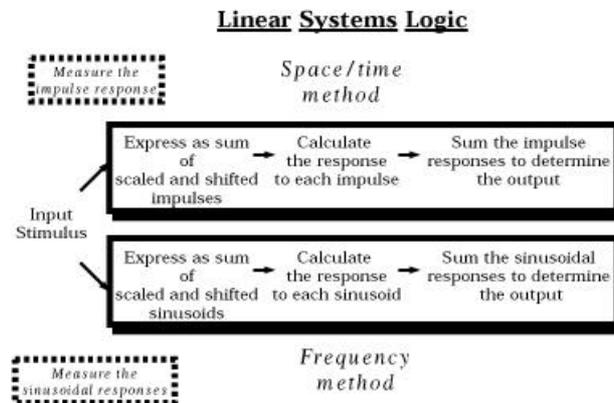
Why bother with sinusoids when we were doing just fine with impulses? The reason is that sinusoids have a very special relationship to shift-invariant linear systems. When we use a sinusoidal stimulus as input to a shift-invariant linear system, the system's responses is always a (shifted and scaled) copy of the input, *at the same frequency as the input*. That is, when the input is $\sin(2\ p\ f\ t)$ the output is always of the form $A \sin(2\ p\ f\ t + f)$. Here, *f* determines the amount of shift and *A* determines the amount of scaling. Thus, measuring the response to a sinusoid for a shift-invariant linear system entails measuring only two numbers: the shift and the scale. This makes the job of measuring the response to sinusoids at many different frequencies quite practical.

Shift-Invariant Linear Systems and Sinusoids
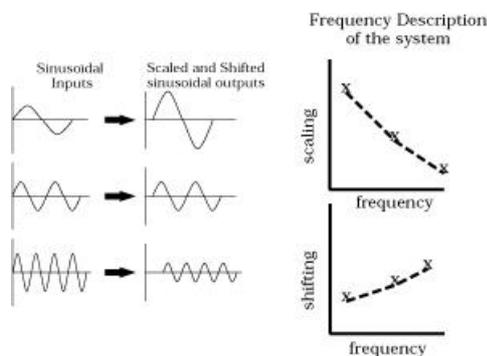
Often, then, when scientists characterize the response of a shift-invariant linear system they will not tell you the impulse response. Rather, they will give you plots that tell you about the values of the shift and scale for each of the possible input frequencies. This representation of how the shift-invariant linear system behaves 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. If you know the responses for sine waves of all frequencies, then you can determine how the system will respond to any possible periodic stimulus.

Linear Systems Logic

```
Measure the          Space/time
impulse response      method

         Express as sum      Calculate         Sum the impulse
            of            the response      responses to determine
       scaled and shifted  to each impulse        the output
          impulses
Input
Stimulus
         Express as sum      Calculate        Sum the sinusoidal
            of            the response      responses to determine
       scaled and shifted  to each sinusoid       the output
          sinusoids

Measure the           Frequency
sinusoidal responses    method
```

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 systems this way. Stereo systems, for example, are pretty good shift-invariant linear systems. They 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 into many frequency bands to give you finer control.

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 (just like the Fourier Transform): 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 at the oval window to the simple oscillation of different frequency analyzers along the basilar membrane. Different points along the basilar membrane respond based on the frequency components in the signal. This representation is used by the brain to analyze sounds for pitch and for meaning.