Using Audio Amplifier Distortion Characteristics as a Pedagogical Tool to Explain Signal Processing Theory

Paul Crilly, Richard Hartnett
Department of Engineering, Electrical Engineering,
U.S. Coast Guard Academy, New London, CT
Paul.B.Crilly@uscga.edu

Abstract—This paper will describe how audio amplifiers specifications can be used as a pedagogical tool to help explain distortion concepts in signal processing theory. Audio entertainment systems are an ideal and practical vehicle for this subject since these are ubiquitous among students, and audio quality or lack thereof is readily perceived. We will consider various distortion types including amplitude, phase, and nonlinear. In the nonlinear distortion category we present harmonic, and intermodulation. Also discussed are distortions due to limits of dynamic range, and quantization errors. We then present how distortion concepts can be integrated into a signals and systems course.

Keywords—distortion; fidelity; signal processing; audio engineering

I. INTRODUCTION

The topic of signals and systems is a major component of electrical engineering (EE) curricula. However, this area tends to be mathematical, and thus for some students, may be perceived as being somewhat abstract as compared to other EE topics. Many EEs ask how the content in their signal processing course relates to what is encountered in their everyday world – which often consists of devices such as audio entertainment systems. An important aspect of a signal processing course is system distortion – specifically, to what degree is the output a facsimile of the input. These distortion types include, amplitude, phase, and nonlinear. While the standard signal processing course covers theory of distortion, many undergraduate electrical engineering students often cannot relate this theory to real world phenomena. For example, an electrical engineer should be able to understand and articulate why a CD player, even though quantization noise is added during the recording process, is inherently superior to reproducing a concert as compared to analog magnetic (cassette) tape.

II. SIGNAL DISTORTION

A. Distortionless Systems

The assumed goal for an audio amplifier is that it accurately replicates the original input*. In other words, it has high fidelity or is distortionless. More formally a distortionless system is where the output is a scaled and delayed version of the input. This is expressed as \( y(t) = Kx(t - t_d) \) with \( x(t) \) and \( y(t) \) the input and output, respectively, \( K \) is a constant gain factor, and \( t_d \) is a constant time delay. The system function would therefore be \( H(j\omega) = Ke^{-j\omega t_d} \), where \( \omega t_d \) is the system’s phase term. Because \( K \) is constant, we say the system has a flat response over all frequencies. The types of distortion typically encountered are amplitude, phase, and nonlinear. System distortion will affect the shape of the output signal; in addition, nonlinearity can also add artifacts to the output and/or affect the relative amplitudes of the various input components.

B. Amplitude Distortion

Amplitude distortion occurs when the \( K \) term varies with frequency. The degree of amplitude distortion is often called magnitude response. Thus a high quality system will have a “flat response.” For example, it is desirable for an audio amplifier to have a response that is constant from 20 Hz to 20 kHz, the range of the human ear. In practice, flat means the response can vary as much as 3 dB. An example of amplitude distortion would be the telephone system where the frequency range is 300 to 3000 Hz; adequate for voice, but not music. An extreme example is a system where the input is a square wave, “rich” in harmonics, and the output is a single sinusoid. System bandwidth is a measure of amplitude distortion.

C. Phase or delay distortion

Phase or delay distortion occurs when the phase term, \( \omega t_d \), is not a linear function of frequency. However, most human ears, with the exception of a musician, or the mastering engineer, are relatively insensitive to phase distortion. Therefore, phase distortion is usually not of significant concern in music systems.

* To be sure, some desire a system where distortion and/or other artifacts are intentionally added into the output. The vacuum tube amplifier is one example where the unintended distortion is pleasant to the ear.
We present a brief example of phase distortion. Consider a system with flat frequency response and linear phase with an input of 
\[ x(t) = 10 \cos \frac{\pi}{2} (t-2) + 2 \cos \frac{\pi}{6} (t-2) \] 
and transfer function of 
\[ H(j\omega) = 5e^{-j\omega} \], yields an output of 
\[ y(t) = 50 \cos \frac{\pi}{2} (t-5) + 10 \cos \frac{\pi}{6} (t-5) \]. Note the two components have equal time delays and thus meets our non-distortion criteria. The same input to a system with nonlinear phase where 
\[ H(j\omega) = 5e^{-j\omega} \] 
yields 
\[ y(t) = 50 \cos \frac{\pi}{2} (t-5.61) + 10 \cos \frac{\pi}{6} (t-1.32) \].

Note there is no amplitude distortion, but the uneven relative time delays indicate phase distortion. An example of a system whose input is a square wave and whose output response has minimal amplitude distortion, but phase distortion is shown in Fig. 1. References [1]-[4] contain additional information on amplitude and phase distortion.

![Fig. 1. Output waveform from a system with negligible amplitude distortion, but nonlinear phase distortion. The system's input is a square wave.](image)

**D. Nonlinear distortion**

A linear system is one where proportional changes in the input result in the same proportional changes in the output. Consider the classic operational amplifier circuit where the voltage gain is 
\[ A_v = 10 \], and power supply \( V_{\text{sup}} = \pm 12 \text{ V} \). The circuit will operate linearly if the input voltage, \( |v_i| \leq 1.2 \text{ volts} \), otherwise the output flattops to the power supply rails and hence the amplifier is no longer linear. In fact, when the input is large enough such that the amplifier exceeds its linear dynamic range; it becomes overloaded causing the output to have additional frequency components or artifacts. Thus the system is nonlinear and thus exhibits nonlinear distortion. Not only can a nonlinear system cause these spurious outputs, but in the case of multiple inputs, the weaker input may be drowned out by the stronger one. As we will see later, the range of input magnitude such that the amplifier does not overload is referred to as dynamic range.

Nonlinear distortion can be modeled in the following manner: 
\[ y(t) = a_0x(t) + a_1x^2(t) + a_2x^3(t) + \ldots \]  
where one or more coefficients are significant relative to \( a_0 \). This nonlinear system may produce spurious output artifacts. Harmonic and intermodulation are two types of nonlinear distortion and are described below.

**Harmonic distortion** is where a single tone sinusoidal input generates harmonics. Consider a system with 
\[ y(t) = x(t) + 0.1x^2(t) \] 
and input \( x(t) = \cos 2\pi f_1 t \). The resulting output is 
\[ y(t) = \cos 2\pi f_1 t + 0.05 + 0.05 \cos 4\pi f_1 t \]. Neglecting the DC term, the output spectral magnitude, 
\( |Y(f)| \), is shown in Fig. 2. Note the spurious harmonic at 
\( 2f_1 \).

![Fig. 2. Harmonic distortion example for a system with a single tone input.](image)

**Intermodulation distortion** is the result when the nonlinear elements in the system cause the multiplication or modulation of the input components to generate additional, but spurious output terms. Recall that the product of two cosines results in new cosine functions whose angles are the sum and difference of the original functions. Consider the above nonlinear system, except that instead of a single input, we have two inputs at different frequencies. If the input were 
\[ x(t) = \cos 2\pi f_1 t + \cos 2\pi f_2 t \], the output becomes 
\[ y(t) = (\cos 2\pi f_1 t + \cos 2\pi f_2 t) + 0.1(\cos 2\pi f_1 t \cos 2\pi f_2 t)^2 \]. If we carry out the math, we observe four additional and spurious sinusoidal outputs at frequencies 
\( 2f_1 + 2f_2 \), \( f_1 + f_2 \), and \( f_1 - f_2 \). If \( f_1 = 2 \text{ kHz} \) and 
\( f_2 = 5 \text{ kHz} \), then in addition to an output of \( 2 \) and \( 5 \text{ kHz} \) signals, the output would also yield tones at \( 4, 10, 7 \) and \( 3 \text{ kHz} \). This is illustrated in Fig. 3. Going beyond this simple example, imagine the level of intermodulation distortion if the input were music with its many harmonics. Initially it may appear that a level of 0.1 in the squared term of \( x \) doesn’t seem significant, but 0.1 is only -20 dB. The human ear can perceive sounds that are 30 dB down from the average signal level.

![Fig. 3. Intermodulation distortion example with 2 tone input.](image)

In published specifications of some commercial audio systems, all distortion products are lumped together and constitute total harmonic distortion (THD). THD can vary from -38 to -137 dB (0.013% to 0.000013%). Cabot [5]
provides an excellent discussion on the theory and measurement of these types of distortion.

**E. Dynamic Range**

**Dynamic range** is the difference (for dB) between the loudest and weakest sounds, an example being an orchestra’s brass versus woodwind sections. Dynamic range is a limitation on a system’s linearity. The ability of an audio amplifier to preserve the relative powers of these sections is a measure of its dynamic range. Dynamic range is defined as $\text{DR} = 20 \log \left( \frac{V_{\text{max}}}{V_{\text{min}}} \right)$ with an orchestra having a DR of 120 dB. It should be noted that DR may be limited by the recording system. For example, a cassette tape’s DR is 70 dB, whereas a CD ROM’s DR is 96 dB [1]. Thus for digital plus analog systems, the DR is dictated by the number of quantization levels, the media, and the analog electronics. This is illustrated in Figure 4. In observing Figure 4, consider the following example. Let’s say we are interested in recording an orchestra with the full range of instruments. In order to “fit” the orchestra’s range of 120 dB inside the CD’s 96 dB dynamic range we have to limit the amplitude of the brass section and/or slightly increase the amplitude of the woodwind section. Therefore, in systems where the recording and/or sound system has less dynamic range than the original input, we must employ some type of amplitude compression (e.g. logarithmic compressor) so the recorded sound will fit inside the amplitude limits. In another example, to tape record a live performance, we have to compress the 120 dB input so it will fit inside the tape’s 70 dB range. Note that this amplitude compression introduces a different sort of nonlinear distortion whereby we sacrifice output signal shape fidelity in order to minimize overload and spurious output components. A manifestation of the increased dynamic range is in the transition from tape to CD players inside automobiles, where there is an increased tendency for users to adjust the sound volume.

![Fig. 4. Graph to illustrate dynamic ranges for various music sources.](image)

**F. Desirable distortion**

To give the devil his due, we conclude this section with a brief commentary of vacuum tube amplifiers. The vacuum tube, like a transistor is an amplifying device. The vacuum tube inherently has a relatively wide dynamic range as compared to the bipolar transistor. When a tube amplifier is overdriven so the output is clipped, the clipping is less abrupt than its transistor counterpart, hence more pleasant to the ear (i.e. a softer manifestation of nonlinearity). However, since the 70s, the vacuum tube amplifier has been obsolete since with today’s technology, a solid state amplifier will meet any realistic fidelity specification. Power wise the tube amp is much less efficient than its solid state counterpart, and many tube amplifiers have some residual 60 Hz background hum. These and other practicalities, suggest there are no technical reasons to use vacuum tubes in audio systems. On the other hand, many audiophiles prefer the sound from vacuum tube amplifiers since they observe that the tube amplifier introduces a degree of warmth to the listening experience, and as previously stated, clipping with tube systems is less obnoxious than solid state systems. Finally, many believe the sound quality of the tube-speaker combination is superior to that of its transistor-speaker counterpart.

**III. PEDAGOGY**

The use of audio systems as a pedagogical tool to enhance learning in the signals and systems area began several years ago when an invited speaker at an IEEE student branch meeting made a presentation on audio amplifier quality. The talk, while mostly qualitative, piqued the students’ interest in signals and system theory and gave them an appreciation that abstract signals and system theory can readily explain real world phenomena. We initially incorporated audio amplifier specification examples into the lecture material of an existing signals and systems course. In the future, it is envisioned that new laboratory projects will be developed where students can audibly experience the various types of distortions as well as observe these on an oscilloscope and dynamic signal analyzer (DSA). Examples would include the effects of overdriving an amplifier, reducing the number of quantization levels when digitizing a signal, and observing and measuring the outputs of various audio amplifiers that have various degrees of total harmonic distortion. While we could offer a separate course on audio distortion, our intent is to integrate this topic into an existing signals and systems course.

Distortion and signal overload is extremely relevant to communications systems and thus was also incorporated into our communication systems course. For example, the near-far problem whereby the receiver front end is required to have wide dynamic range in order to be able to simultaneously capture and process both weak and strong signals; a design challenge in any receiver technology, but particularly in today’s CDMA wireless phones where multiple signals are all at the same frequency.

A learning objective is that engineering students will better understand the specifications of an audio system and thus be able to make meaningful comparisons between competing systems and go beyond many of the existing electronics and signals and systems courses that assume linearity and primarily focus on maximizing bandwidth. Another objective is that by incorporating audio applications into the signals and systems courses, students are forced to confront the non-ideal behavior of real systems, and understand how this type of behavior affects the listening experience. In any case, audio systems present an interesting case study and thus are an alternative to the usual academic and theoretical black box systems encountered in the undergraduate signals and systems curriculum.
IV. CONCLUSIONS

This paper has presented a perspective on system response that goes beyond the traditional pedagogical undergraduate electrical engineering topics of amplitude distortion, step response, and even phase distortion. Using the audio amplifier example, we have introduced another set of system specifications that characterize system fidelity and performance. We have also shown that nonlinear distortion affects response by either introducing new artifacts in the output, or because of dynamic range limitations requires the input signals fall within certain signal strengths so as prevent additional output artifacts from being introduced.

References


