Why Class-D Audio Amplifiers Work Well: A Theoretical Explanation

Kevin Alvarez[1], Julio C. Urenda[1,2], and Vladik Kreinovich[1]
[1]Department of Computer Science
[2]Department of Mathematical Sciences
University of Texas at El Paso
El Paso, TX 79928, USA
kalvarez9@miners.utep.edu, jcurenda@utep.edu, vladik@utep.edu

Abstract

Most current high-quality electronic audio systems use class-D audio amplifiers (D-amps, for short), in which a signal is represented by a sequence of pulses of fixed height, pulses whose duration at any given moment of time linearly depends on the amplitude of the input signal at this moment of time. In this paper, we explain the efficiency of this signal representation by showing that this representation is the least vulnerable to additive noise (that affect measuring the signal itself) and to measurement errors corresponding to measuring time.

1 Formulation of the Problem

Most current electronic audio systems use class-D audio amplifiers, where a signal $x(t)$ is represented by a sequence $s(t)$ of pulses whose height is fixed and whose duration at time t linearly depends on the amplitude $x(t)$ of the input signal at this moment of time; see, e.g., [1] and references therein. Starting with the first commercial applications in 2001, D-amps have used in many successful devices.

However, why they are so efficient is not clear. In this paper, we provide a possible theoretical explanation for this efficiency.

We will do it in two sections. In the first of these sections, we explain why using pulses makes sense. In the following section, we explain why the pulse's duration should linearly depend on the amplitude of the input signal $x(t)$.

2 Why Pulses

Let us start with some preliminaries.

There are bounds on possible values of the signal. For each device, there are bounds $\underline{s}$ and $\overline{s}$ on the amplitude of a signal that can be represented by this device. In other words, for each signal $s(t)$ represented by this device and for each moment of time t, we have $\underline{s} \le s(t) \le \overline{s}$.

Noise is mostly additive. In every audio system, there is noise $n(t)$. Most noises are *additive*: they add to the signal. For example, if we want to listen to a musical performance, then someone talking or coughing or shuffling in a chair produce a noise signal that adds to the original signal.

Similarly, if we have a weak signal $s(t)$ which is being processed by an electronic system, then additional signals from other electric and electronic devices act as an additive noise $n(t)$, so the actual amplitude is equal to the sum $s(t) + n(t)$.

What do we know about the noise. In some cases – e.g., in a predictable industrial environment – we know what devices produce the noise, so we can predict some statistical characteristics of this noise.

However, in many other situations – e.g., in TV sets – noises are rather unpredictable. All we may know is the upper bound n on possible values of the noise $n(t)$: $|n(t)| \le n$.

We want to be the least vulnerable to noise. To improve the quality of the resulting signal, it is desirable to select a signal's representation in which this signal would be the least vulnerable to noise.

Additive noise $n(t)$ changes the original value $s(t)$ of the signal to the modified value $s(t)+n(t)$. Ideally, we want to make sure that, based on this modified signal, we will be able to uniquely reconstruct the original signal $s(t)$.

For this to happen, not all signal values are possible. Let us show that for this desired feature to happen, we must select a representation $s(t)$ in which not all signal values are possible.

Indeed, if we have two possible values $s_1 < s_2$ whose difference does not exceed $2n$, i.e., for which $\frac{s_2 - s_1}{2} \le n$, then we have a situation in which:

- to the first value s_1, we add noise $n_1 = \frac{s_2 - s_1}{2}$ for which $|n_1| \le n$, thus getting the value $s_1 + n_1 = \frac{s_1 + s_2}{2}$; and
- to the second value s_2, we add noise $n_2 = -\frac{s_2 - s_1}{2}$ for which also $|n_2| \le n$, thus also getting the same value $s_2 + n_2 = \frac{s_1 + s_2}{2}$.

Thus, based on the modified value $s_i + n_i = \frac{s_1 + s_2}{2}$, we cannot uniquely determine the original value of the signal: it could be s_1 or it could be s_2.

Resulting explanation of using pulses. The larger the noise level n against which we are safe, the larger must be the difference between possible values of the signal. Thus, to makes sure that the signal representation is safe against the strongest possible noise, we must have the differences as large as possible.

On the interval $[\underline{s}, \overline{s}]$, the largest possible difference between the possible values is when one of these values is $\underline{s}$ and another value is $\overline{s}$. Thus, we end up with a representation in which at each moment of time, the signal is equal either to $\underline{s}$ or to $\overline{s}$.

This signal can be equivalently represented as a constant level $\underline{s}$ and a sequence of pulses of the same height (amplitude) $\overline{s} - \underline{s}$.

3 Why the Pulse's Duration Should Linearly Depend on the Amplitude of the Input Signal

Pulse representation: reminder. As we concluded in the previous section, to make the representation as noise-resistant as possible, we need represent the time-dependent input signal $x(t)$ by a sequence of pulses.

How can we encode the amplitude of the input signal. Since the height (amplitude) of each pulse is the same, the only way that we can represent information about the input signal's amplitude $x(t)$ at a given moment of time t is to make the pulse's width (duration) w – and the time interval between the pulses – dependent on $x(t)$. Thus, we should select w depending on the amplitude $w = w(x)$.

Which dependence $w = w(x)$ should we select: analysis of the problem. Which dependence of the width w on the initial amplitude should we select?

Since the amplitude x of the input is encoded by the pulse's width $w(x)$, the only way to reconstruct the original signal x is to measure the width w, and then find x for which $w(x) = w$.

We want this reconstruction to be as accurate as possible. Let $\varepsilon > 0$ be the accuracy with which we can measure the pulse's duration w. A change Δx in the input signal – after which the signal becomes equal to $x + \Delta x$ – leads to the changed width $w(x + \Delta x)$. For small Δx, we can expand the expression $w(x + \Delta x)$ in Taylor series and keep only linear terms in this expansion:

$$w(x + \Delta x) \approx w(x) + \Delta w,$$

where we denoted $\Delta w \stackrel{\text{def}}{=} w'(x) \cdot \Delta x$. Differences in width which are smaller than ε will not be detected. So the smallest difference in Δx that will be detected comes from the formula $|\Delta w| = |w'(x) \cdot \Delta x| = \varepsilon$, i.e., is equal to $|\Delta x| = \dfrac{\varepsilon}{|w'(x)|}$.

In general, the guaranteed accuracy with which we can determine the signal is thus equal to the largest of these values, i.e., to the value

$$\delta = \max_x \frac{\varepsilon}{|w'(x)|} = \frac{\varepsilon}{\min\limits_x |w'(x)|}.$$

We want this value to be the smallest, so we want the denominator $\min_x |w'(x)|$ to attain the largest possible value.

The limitation is that the overall widths of all the pulses corresponding to times

$$0 = t_1, t_2 = t_1 + \Delta t, \ldots, t_n = T$$

should fit within time T, i.e., we should have

$$\sum_{i=1}^{n} w(x(t_i)) \le T.$$

This inequality must be satisfied for all possible input signals. In real life, everything is bounded, so at each moment of time, possible values $x(t)$ of the input signal must be between some bounds $\underline{x}$ and $\overline{x}$: $\underline{x} \le x(t) \le \overline{x}$.

We want to be able to uniquely reconstruct x from $w(x)$. Thus, the function $w(x)$ should be monotonic - either increasing or decreasing.

If the function $w(x)$ is increasing, then we have $w(x(t)) \le w(\overline{x})$ for all t. Thus, to make sure that the above inequality is satisfied for all possible input signals, it is sufficient to require that this inequality is satisfied when $x(t) = \overline{x}$ for all t, i.e., that $n \cdot w(\overline{x}) \cdot T$, or, equivalently, that $w(\overline{x}) \le \Delta t = \dfrac{T}{n}$.

If the function $w(x)$ is decreasing, then similarly, we conclude that the requirement that the above inequality is satisfied for all possible input signals is equivalent to requiring that $w(\underline{x}) \le \Delta t$.

Thus, to find the best dependence $w(x)$, we must solve the following optimization problem:

Which dependence $w = w(x)$ should we select: precise formulation of the problem. To find the encoding $w(x)$ that leads to the most accurate reconstruction of the input signal $x(t)$, we must solve the following two optimization problems.

- Of all increasing functions $w(x) \ge 0$ defined for all $x \in [\underline{x}, \overline{x}]$ and for which $w(\overline{x}) \le \Delta t$, we must select a function for which the minimum $\min_x |w'(x)|$ attains the largest possible value.
- Of all decreasing functions $w(x) \ge 0$ defined for all $x \in [\underline{x}, \overline{x}]$ and for which $w(\underline{x}) \le \Delta t$, we must select a function for which the minimum $\min_x |w'(x)|$ attains the largest possible value.

Solution to the above optimization problem. Let us first show how to solve the problem for the case when the function $w(x)$ is increasing. We will show that in this case, the largest possible value of the minimum $\min_x |w'(x)|$ is attained for the function $w(x) = \Delta t \cdot \dfrac{x - \underline{x}}{\overline{x} - \underline{x}}$.

Indeed, in this case, the dependence $w(x)$ on x is linear, so we have the same value of $w'(x)$ for all x – which is thus equal to the minimum:

$$\min_x |w'(x)| = w'(x) = \frac{\Delta t}{\overline{x} - \underline{x}}.$$

Let us prove that the minimum cannot attain any larger value. Indeed, if we could have

$$m \stackrel{\text{def}}{=} \min_x |w'(x)| > \frac{\Delta t}{\overline{x} - \underline{x}},$$

then we will have, for each x,

$$w'(x) \geq m > \frac{\Delta t}{\overline{x} - \underline{x}}.$$

Integrating both sides of this inequality by x from $\underline{x}$ to $\overline{x}$, and taking into account that

$$\int_{\underline{x}}^{\overline{x}} w'(x)\,dx = w(\overline{x}) - w(\underline{x})$$

and

$$\int_{\underline{x}}^{\overline{x}} m\,dx = m \cdot (\overline{x} - \underline{x}),$$

we conclude that $w(\overline{x}) - w(\underline{x}) \geq m \cdot (\overline{x} - \underline{x})$. Since $m > \dfrac{\Delta t}{\overline{x} - \underline{x}}$, we have $m \cdot (\overline{x} - \underline{x}) > \Delta t$ hence

$$w(\overline{x}) - w(\underline{x}) > \Delta t.$$

On the other hand, from the fact that $w(\overline{x}) \leq \Delta$ and $w(\underline{x}) \geq 0$, we conclude that $w(\overline{x}) - w(\underline{x}) \leq \Delta t$, which contradicts to the previous inequality. This contradiction shows that the minimum cannot attain any value larger than $\dfrac{\Delta t}{\overline{x} - \underline{x}}$, and thus, that the linear function $w(x) = \Delta t \cdot \dfrac{x - \underline{x}}{\overline{x} - \underline{x}}$ is indeed optimal.

Similarly, in the decreasing case, we can prove that in this case, the linear function $w(x) = \Delta t \cdot \dfrac{\overline{x} - x}{\overline{x} - \underline{x}}$ is optimal.

In both cases, we proved that the width of the pulse should linearly depend on the input signal – i.e., that the encoding use in class-D audio amplifiers is indeed optimal.

Acknowledgments

This work was supported in part by the US National Science Foundation grants 1623190 (A Model of Change for Preparing a New Generation for Professional Practice in Computer Science) and HRD-1242122 (Cyber-ShARE Center of Excellence).

References

[1] B. Santo, “Bringing big sound to small devices: One module revolutionized how we listen”, *IEEE Spectrum*, 2019, No. 10, p. 17.