Short note
The ‘a + b’ argument for teaching resonance

Emanuel Gluskin\(^1\) and Doron Shmilovitz\(^2\)
\(^1\)Electrical Engineering Department, Faculty of Engineering, Ort Braude Academic College, Carmiel, Israel
E-mail: gluskin@ee.bgu.ac.il
\(^2\)Department of Physical Electronics, School of Electrical Engineering, Tel-Aviv University, Ramat Aviv, Israel
E-mail: shmilo@eng.tau.ac.il

Abstract An objection to teaching resonance exclusively in terms of phasors is expressed.

Keywords education; relaxation time; resonance

We point out an essential matter as regards engineering education, in the teaching of the resonant phenomenon. Unfortunately this matter is often overlooked in many textbooks, which reflects, in our opinion, a misinterpretation of the nature of resonant systems which may result in erroneous application of such systems for measurement and treatment of frequencies and signals.

In general, only the established sinusoidal response is associated with the concept of resonance in the standard textbooks such as Refs 1–4 and many others. For instance, in Ref. 1, the topic of resonance appears in the section ‘Sinusoidal steady-state analysis’ and in Ref. 2 it appears in the section ‘Frequency response’. The transient associated with the establishment of the steady resonant state response is traditionally ignored in textbooks.

Evidently, it is easier to study and teach the features of only the steady state response to the sinusoidal excitation, since this part of the complete response continues, in principle, infinitely and indeed only it is needed for the ‘frequency analysis’ of periodic or stationary input signals. Nevertheless, the time needed for this establishment can not be ignored in modern fast-operating systems. This time is proportional to the same quality factor \(Q\) that defines the quality (intensity) of the resonance and for high values of \(Q\), the steady state cannot be attained in a too-fast operating system. Numerous relevant examples may be pointed out among practical circuits, such as in sensing, instrumentation, and switched mode converters.

An objection against the traditional teaching of resonance exclusively in terms of phasors\(^1\)–\(^4\) is thus expressed here in the form of a mathematically and logically very simple argument, named the ‘\(a + b\) argument’, and we even suggest some additional terminology aimed at preventing scholars from forgetting the transient period.

A second-order oscillatory system is treated below, which is not only the customary, but, perhaps, also the most important case, and is sufficient for making the point. Nevertheless, the transient should also not be ignored in higher order, more complex oscillatory systems.
Do not miss the transient!

If a sinusoidal force acts on a linear time-invariant mechanical system, or if a sinusoidal voltage is applied to an electrical circuit, etc., then the finally established steady state response, defined below as $b(t)$, is sinusoidal. Since sinusoidal function $f(t)$ does not satisfy, at any point $t$, both of the conditions $f(t) = 0$ and $df/dt = 0$, if the system (which must be at least of the second order for resonance to be obtained in it) has zero initial conditions, its solution, or output (that is, the system's response to the input) must be of the form

$$f_{out}(t) = a(t) + b(t) \quad (1)$$

where $a(t)$ is a solution of a relevant homogeneous equation. The zero initial conditions can be satisfied only when the function $a(t)$ is non-zero.

The very simple, but for our point important fact (the 'a + b argument') is that until $a(t)$ has decayed, the steady state response $f_{out}(t) = b(t)$ is not obtained.

For a second-order oscillatory system, the decay of $a(t)$ is associated (see any course of physics, or circuit theory, or mechanical oscillations) with a factor of the type

$$e^{-\gamma t}$$

where $\gamma > 0$ is the 'damping factor',\textsuperscript{1-4} such that we have for the decay time of $a(t)$

$$\frac{1}{\gamma} \sim QT_0 \quad (2)$$

where $Q$ is the 'quality factor'\textsuperscript{1-4} of the oscillatory system, and $T_0$ is the system's self-oscillations period.

Thus, higher $Q$ implies not only more intensive amplitude of $b(t)$ for the input amplitude fixed, but also longer settling time till the establishment of $b(t)$. It becomes obvious that one cannot ignore the initial transient in any serious study of resonance. The point is not only logical. In the context of modern electronics, every desirable state (here the steady state response) has to be attained fast.

It needs to be seen that the situation with the technical and purely physical applications of resonance are different. In physics, the energy pumping into the system, clearly seen via the amplitude increase during the transient, can mean, for instance, very significant resonant absorption of photons, and the resonance is sometimes interpreted (defined in Ref. 5) as this transient, i.e. as the initial, linear increase of the amplitude.

The role of $Q$ is thus perfectly seen not only through $b(t)$ but also in $a(t)$, i.e. through the duration $1/\gamma$ of the transient phase during which the envelope of $f_{out}(t)$ essentially increases from 0 to $\max\{b(t)\}$.

This oscillations' 'envelope' is shown in Fig. 1, for excitation precisely at the resonant frequency (which avoids the minor complication otherwise caused by the possible beats between $a(t)$ and $b(t)$). The envelope increases linearly during a period of about $QT_0$, and then gradually saturates at a height (amplitude level) which is proportional to $Q$, and thus is established.
Fig. 1  The schematic resonant response (precisely at the resonant frequency, otherwise we would have some beats of a(t) + b(t) during the transient period) starting from zero conditions, presented by the saturating envelope of the oscillations that finally become sinusoidal. The higher the final amplitude, the longer the time of its establishment, in agreement with eqn (1).

Obviously, the existence of the transient is relevant to any use of a resonant curve/circuit in instrumentation and measurement. Any technical idea of associating a measurement process or a signal treatment with a transfer from one point on the resonance curve to another such point, must take account of the time requirements; such a transfer requires time of the order of QT₀ because of the role of e⁻ᵀ in the relevant a(t).

However, in the absolute majority of the circuit-theory and technical textbooks, for instance Refs 1–4, the topic of resonance is found treated only in terms of phasors related solely to the steady-state response b(t).

We consider this neglect of the basic ‘a + b’ argument, based on the elementary equation (1), in the standard textbooks to be strange and the resulting situation as regards the teaching of resonance unsatisfactory. The correct outlook on the resonant response has to be introduced into the elementary curriculum. This criticism is quite in the spirit of the ‘old’ pedagogical advice by Ernst Guillemin⁶ not to hurry with the frequency-domain analysis, letting the physical reality be first well understood in the time domain.

It may be suggested that the transient part of the resonant response (that is, a(t) + b(t) as long as a(t) is significant) be named the ‘Q-t part’, and the steady-state part, that is, b(t), the ‘Q-a part’. These notations mean that in the first case the ‘quality factor’ Q is expressed via the time (t) parameters, and in the second case via the amplitude (a) parameters. These connections with the main (basic) parameter Q characterising resonance stress the physical importance of both parts of the response, which is the point of the terminology.
Concluding the point, we would also like to warn against reversing the emphasis during the teaching of technical students, because such reversing can unjustifiably influence the very definition of resonance appropriate for such students. Although, as was noted earlier, in physics where the transient, i.e., the start of intensive energy absorption, can be very important per se, and thus resonance can be defined just as the linear increase in the amplitude, in electrical engineering the concept of the quality factor $Q$ is most important. The linear increase can, in principle (i.e. ideally), also exist for $Q = \infty$ (no losses in the circuit), when it is never finished, and the steady state never obtained. However, in electrical engineering, the value of $Q$ that defines the finite amplitude of $b(t)$ is always the focus, and the idealisation of $Q = \infty$ is never relevant. Thus, in technical education, the definition of resonance has to be done in terms of the ‘frequency response’ $b(t)$, i.e. as usual, using the resonant curve.

Last but not least, the teacher of basic courses must look forward and prepare the students for modern technical reality. We have been speaking only about linear systems for which eqn (1) is relevant, but when passing on to nonlinear oscillatory systems, for which ‘frequency response’ is not a very good concept (since Fourier transform is poorly effective for such systems), the time-domain analysis actually becomes the main analytical tool. This enhances, in the methodological plane, the importance of not missing this analysis, starting from teaching resonance in simple linear systems.

After all, the simple logic of our argument should not be forgotten:

1. Since the sinusoidal function cannot satisfy zero initial conditions, the response must include two terms.
2. The response cannot become only one of the terms if the other has not decayed.

Acknowledgement

We are grateful to unknown reviewers for some helpful comments.

References