# NOT YOUR ORDINARY SOUND EXPERIENCE: A NONLINEAR-ACOUSTICS PRIMER

**Anthony A. Atchley**
*Graduate Program in Acoustics, The Pennsylvania State University*
*University Park, PA 16802*

Ordinarily, sound is a very weak mechanical process. So weak, in fact, that the majority of acoustic phenomena encountered in everyday life can be modeled to a high degree of accuracy by assuming that the amplitude of the sound is infinitesimally small. Our intuition about sound is shaped, to a large extent, by the consequences of this infinitesimal-amplitude, or linear, theory. However, there are many situations where the sound amplitude is so large that linear theory breaks down. This high-amplitude regime is characterized by acoustic phenomena unanticipated from linear theory. Far from just an academic exercise, nonlinear acoustic theory has led to widespread practical applications in the areas of medical imaging, biomedical ultrasound, nondestructive testing, and aircraft design, to name but a few.

*"Nonlinear acoustic theory has led to widespread practical applications in the areas of medical imaging, biomedical ultrasound, nondestructive testing, and aircraft design."*

As implied by the title, the purpose of this article is to introduce readers unfamiliar with nonlinear acoustics to a (very) few of its most fundamental concepts and to use these concepts to discuss a handful of nonlinear acoustic phenomena. No attempt whatsoever is made to encompass the breadth of the field. In fact we will make use of plane waves in an ideal fluid for most of this discussion. This choice serves our purposes, for it permits brevity and simplicity. However, with sincere apologies to experts, a remarkable range of phenomena will not be touched upon at all, particularly nonlinear acoustics in solids. References 1–3 and the works cited within them provide extensive coverage of the field.

We begin by reviewing the most basic underpinnings of linear acoustics and how these concepts are modified under high-amplitude conditions. Two examples are discussed: the propagation of an initially-single-frequency sound wave and the interaction of two such sound waves. The first example forms the basis of wave steepening, harmonic generation, shock formation, and sonic booms. The second example leads to sum and difference frequency generation and the parametric array.

Four basic concepts of *linear* acoustics are essential to our discussion:

1) The speed $v$ with which sound propagates is independent of the amplitude and frequency of the sound wave. As a consequence, in the absence of absorption and geometrical spreading, the shape of a wave, (e.g., pressure vs. time) remains unchanged as it propagates.

2) The medium through which a sound wave propagates oscillates longitudinally back and forth along the direction of propagation. The speed of the moving medium $u$ is assumed to be much, much less than the speed of sound.

3) The medium returns to its initial state after the wave passes.

4) The principle of superposition holds for linear acoustic processes. This principle states that the net effect of multiple sound waves interacting at a given location and time is the sum of the effects caused by each individual wave. Much of our intuition about acoustics is shaped by this principle.

A single cycle of a continuous sound wave is depicted in Fig. 1. This figure shows a graph of the fluid speed $u$, normalized to its peak value, as a function of time $t$, relative to the acoustic period $T$. The graph represents, for example, the output of a sensor as a function of time as would be seen, for instance, on the display of an oscilloscope. According to linear acoustics, each point on the waveform travels with a speed equal to the infinitesimal-amplitude speed of sound $c_0$. However in nonlinear acoustics, the speed of propagation $v$ depends on the instantaneous amplitude of the signal. In other words, every point on the waveform travels with a different speed. In general, $v$ can be expressed as $v = c_0 + \beta u$, where $c_0$ is the infinitesimal-amplitude speed of sound, what we ordinarily call the speed of sound, and $\beta$ the coefficient of nonlinearity. The amplitude dependence of the speed of propagation lies at the origin of nonlinear acoustics. As indicated on the graph in Fig. 1, $v$ is greater than $c_0$ during times when $u$ is positive, and less than $c_0$ when $u$ is negative. When $u$ is zero, $v = c_0$. Imagine traveling along with this wave at speed $c_0$. The sound in regions of positive values of $u$ travels faster than you. Hence, it will gradually catch up with you as it propagates. In contrast, the sound in regions of negative $u$ travels slower than you and so you tend to catch up with it. The net result of the differences in propagation speed is that the waveform distorts as it travels. Notice however, that because the zero-crossings travel at speed $c_0$, the acoustic period remains constant as the wave propagates. This behavior is a property of continuous waveforms. As we will see later, there are waveforms for which the period changes with propagation distance.

One may be tempted to recover linear acoustic behavior by setting $\beta = 0$, in which case, $v = c_0$, the linear acoustic value. However, $\beta$ is not zero, even for an ideal

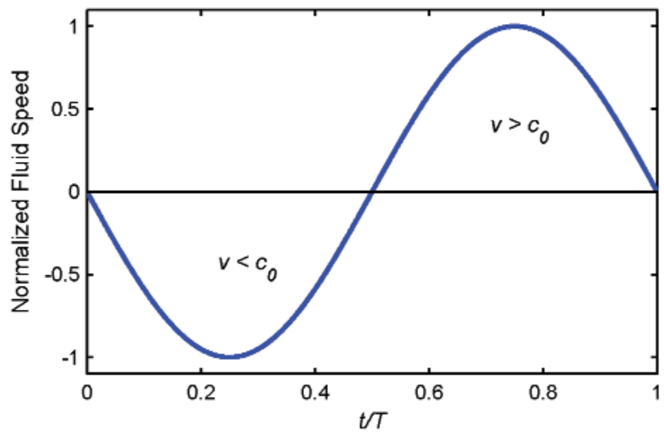


*Fig. 1. Graph of fluid speed u, normalized to its maximum value, as a function of time t, relative to the acoustic period T for one cycle of a continuous wave. The speed of sound v is greater than $c_0$ at times when u is positive, and less than $c_0$ when u is negative. The zero-crossings travel at speed $c_0$. The net result of the amplitude dependence of the speed of sound is that the waveform distorts as it propagates, while maintaining a constant acoustic period.*

fluid. It is a physical property of the fluid. For example, in a fluid, $\beta = 1 + B/2A$. Table 1 contains representative values of the parameter $B/A$, referred to as "b-over-a." For an ideal gas, $B/A = (\gamma - 1)$, and so $\beta = (1 + \gamma)/2$, where $\gamma$ is the ratio of specific heats. $\gamma = 1.4$ for air. Extensive tables of the parameter $B/A$ for gases, fluids, and tissues, including the values listed in Table 1, can be found in Reference 4 (Chapter 2 of Reference 1) and Reference 2.

*Table 1*

Values of the parameter $B/A$ for selected fluids. Values are from Tables I, II, and III of Ref. 4 and assume an ambient pressure of 1 atm.

| Fluid | *B/A* |
|---|---|
| Acetone | 9.2 |
| Distilled Water | 5.2 |
| Human Liver | 7.6±0.8 |
| Liquid Nitrogen | 7.7 |
| Mercury | 7.8 |
| Seawater | 5.25 |

Linear acoustics is instead recovered, not by setting $\beta = 0$, but rather by setting $u = 0$. Of course, $u$ is zero only if the sound amplitude is zero. However, from a practical point of view, linear-acoustic propagation occurs if the ratio $u/c_0$, also known as the acoustic Mach number $M$, is much less than unity. How much less? A few basic principles of nonlinear acoustics need to be introduced, before this essential question can be answered.

There are two physical causes for nonlinearity in a fluid, i.e., two reasons why $\beta$ is non-zero. One reason is that the speed of sound in a fluid moving with speed $u$ is $c_0 + u$. A, perhaps, familiar consequence of this result is that sound traveling up wind is slower than that traveling down wind. In our case, the fluid motion is not caused by wind, but by the passage of the sound wave itself. During the passage of a sound wave, the speed of the fluid medium $u$ at a particular location changes continuously. Consequently, the instantaneous speed of sound changes continuously as well.

The second cause of nonlinearity in a fluid is that the speed of sound depends upon the local thermodynamic conditions of the fluid. In the case of an ideal gas the adiabatic speed of sound depends on temperature according to $c=\sqrt{\gamma RT/M}$, where $R$ is the universal gas constant, $T$ the absolute temperature, and $M$ the molar mass of the gas. Just as $u$ changes continuously during the cycle, so does the temperature. The combination of convection and the dependence on the thermodynamic state are both captured in the parameter $\beta$.

The evolution of a high-amplitude wave as it propagates is illustrated in Fig. 2. The upper graph, Fig. 2a, shows the normalized acoustic amplitude of a plane wave as a function of time at three propagation distances. The initially-sinusoidal waveform is depicted in blue. The green and red curves show the waveform at successively greater propagation distances, respectively. Because the positive-amplitude regions of the wave travel faster than $c_0$, they will arrive successively earlier during the cycle. In contrast, the negative-amplitude regions travel slower than $c_0$ and arrive succes-

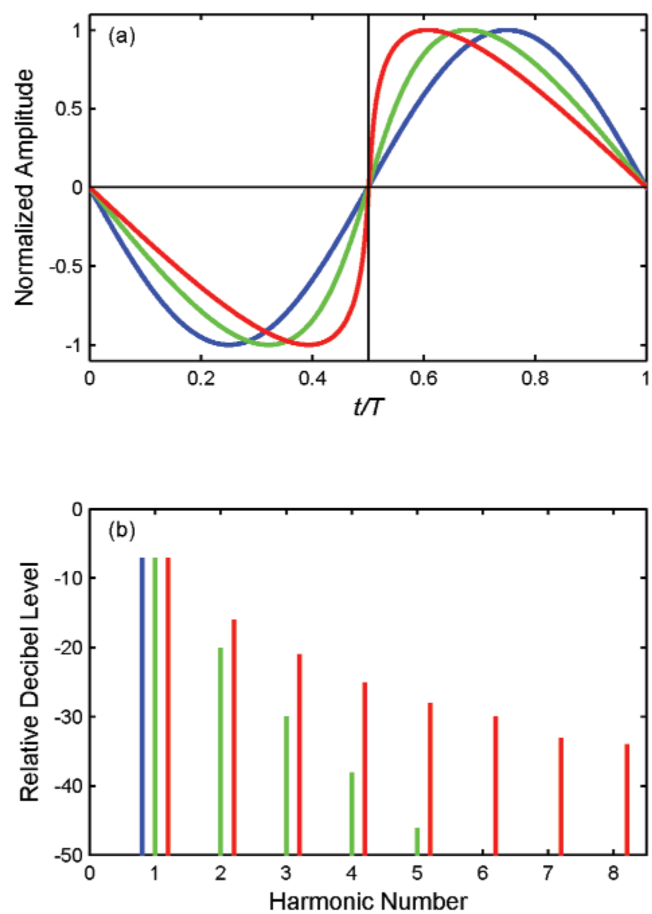


*Fig. 2. a) Graph of normalized acoustic amplitude as a function of t/T at three propagation distances. The initially-sinusoidal waveform is depicted in blue. The green and red curves show the waveform at successively greater propagation distances, respectively. The waveform becomes more and more distorted as it propagates. b) Graph of the frequency spectrum of the waveforms shown in a). Increased waveform distortion is accompanied by increased harmonic content.*

sively later in the cycle. As the wave propagates, it experiences more and more distortion. The continuously increasing level of distortion demonstrates one of the important properties of nonlinear acoustic processes—the effect is cumulative. This is why nonlinear processes may need to be considered, even for seemingly small acoustic Mach numbers.

Figure 2b shows that the distortion is accompanied by a spread of energy from the original frequency *f* to its harmonics. In other words, nonlinear distortion is accompanied by harmonic generation. The two effects are not independent. In fact, they imply one another. To see this, consider a continuous initially-sinusoidal sound wave of frequency *f*. As we have learned, the waveform distorts as it propagates, owing to the amplitude-dependence of *v*. Nevertheless, the waveform is a periodic function of time, the period *T* being 1/*f*. Fourier's theorem tells us that any periodic waveform can be represented in terms of an infinite series of harmonically related sinusoids. The number of harmonics needed to represent the waveform depends on the amount of distortion. As indicated by the blue line in Fig. 2b, only one term in the series is required to represent the initially-sinusoidal wave. As the waveform becomes distorted, however, a single term is not sufficient. Increasingly more harmonics are necessary to reproduce the waveform.

A practical application of harmonic generation can be found in medical imaging using ultrasound. A typical imaging system uses a transducer to generate a pulse of ultrasound. As this pulse propagates through tissue, it reflects off interfaces between different types of tissue, bone, organs, tumors, blood vessels, etc. The reflected signals travel back to the transducer where they are processed to produce an image. The resolution of the image depends on the frequency of the ultrasound. Higher frequencies give higher resolution. Because body tissue is nonlinear, harmonics are generated as the ultrasonic pulse propagates. If these harmonics are used to process the image, rather than the fundamental frequency, the resolution of the image is greatly enhanced. This is the basis for a technique known as tissue harmonic imaging. A number of remarkable images can be found by searching for "tissue harmonic imaging" on the Web.

We deviate here slightly to introduce the concepts of linear and nonlinear systems. In this context, a "system" is anything that responds to a stimulus, i.e., the proverbial black box. The term "system" could refer to a piece of sound recording or sound-reproducing equipment (or both), a microphone, the ear, or, as in our case, a small volume of fluid exposed to sound. One consequence of linear theory is that, given sufficient time, a system excited by a stimulus of frequency *f*, responds at and only at frequency *f*, as depicted in Fig. 3a. This concept lies at the heart of a branch of study known as linear systems theory. Linear acoustics is governed by this property. An example of a (nearly) linear system is a very well designed microphone operating well within its specified input range. The output of an ideal (i.e., linear) microphone is an electrical signal that faithfully captures the amplitude and phase of the sound to which it is exposed. If

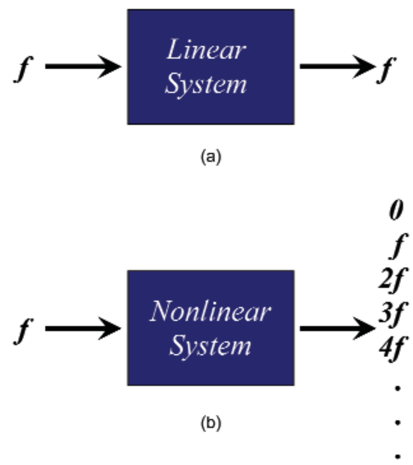


*Fig. 3. a) A stimulus of frequency f applied to a linear system results in an output at, and only at, the same frequency f. b) In contrast, a nonlinear system responds to an input of frequency f, with an output consisting of harmonics of f.*

the input is a single-frequency, constant-amplitude, constant-phase sinusoid, the output is also a single-frequency, constant-amplitude, constant-phase sinusoid.

A nonlinear system responds differently. As before, suppose that the input is a sinusoid of single-frequency *f*. If the system is nonlinear, then in general the output will consist of a combination of sinusoids of frequencies *nf*, where *n* = 0, 1… A common, undesirable example of this phenomenon is harmonic distortion in audio amplifiers. An ideal audio amp is linear; it amplifies the input without changing the shape or frequency content of the waveform. However, no amplifier, no matter how expensive, is truly linear. In practice, the output is never an exact replica of the input. The higher the quality of the amplifier, the more nearly linear its performance, as characterized by lower harmonic distortion. Returning the discussion to the acoustics of fluids, at very low amplitudes the fluid responds approximately as a linear system. However at higher amplitudes, the nonlinear nature of the fluid becomes observable, e.g., harmonic distortion

As we have seen, a high-amplitude wave becomes increasingly distorted as it propagates. Therefore, one might expect it would follow the progression depicted in Fig. 4a, in which the cumulative distortion leads from the blue to the green to the red curves at increasingly larger propagation distances. However, the green and red waveforms are unphysical because they predict that the fluid has three different values of $u$ at the same instance during parts of the cycle. This cannot happen; a given volume of fluid can move at only one speed at a time. Therefore, our simple description of waveform distortion breaks down when it begins to predict unphysical results such as multi-valued physical quantities. An oversimplified, yet useful resolution of this problem is to eliminate the multi-valued region and make the waveform discontinuous. The green waveform from Fig. 4a is reproduced in Fig. 4b, with the multi-valued region shaded magenta. The discontinuity is placed according to the "equal area" rule, by which the area of the multi-valued region ahead of the discontinuity equals that behind it. The discontinuity is called a shock. In reality, a true discontinuity never exists. The abrupt change in amplitude occurs over a finite, yet short, period of time.

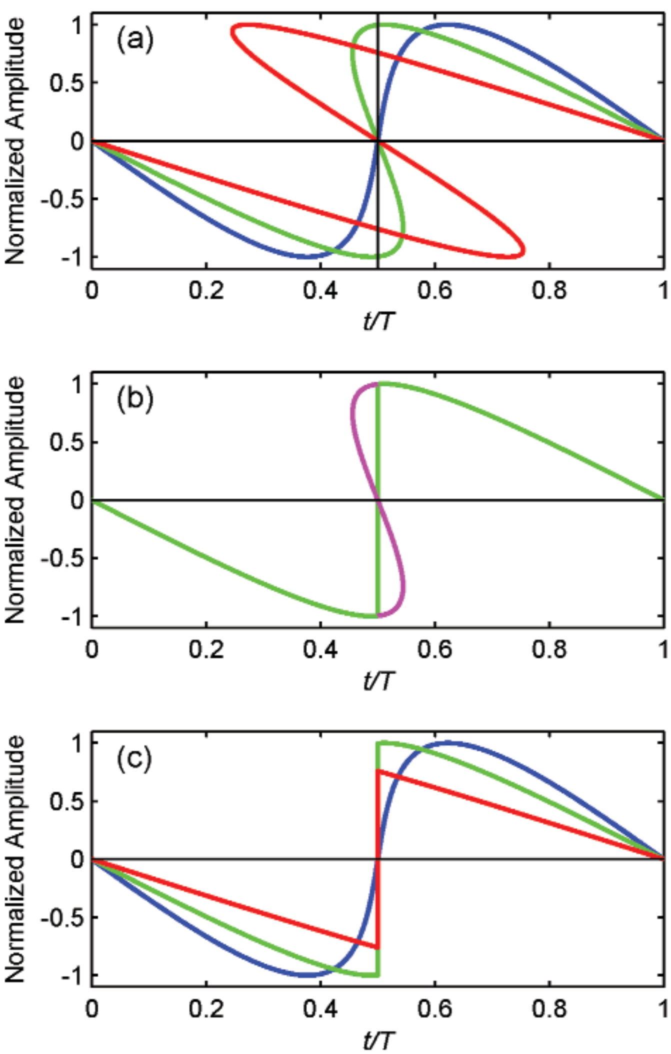


*Fig. 4. a) One might expect that if the waveforms depicted in Fig. 2a were allowed to propagate farther, they would take on waveforms that are multi-valued. b) However, multi-valued waveforms are unphysical. The multi-valued sections of the waveforms are eliminated by replacing them with a discontinuity, or a shock. c) The physical progression of the waveforms depicted in a) leads to shock formation. As the propagation distance increases, the waveform approaches a sawtooth shape, in the absence of attenuation.*

The propagation distance required for an initially-sinusoidal plane wave to form a shock is termed the shock formation distance $\bar{x} = \lambda/2\pi\beta M$, where $\lambda$ is the wavelength of the initially-sinusoidal wave. According to this equation, the longer the wavelength the farther a wave has to propagate before a shock is formed. This is because nonlinear effects are cumulative. Also, shocks form sooner in more nonlinear fluids and for higher amplitudes (larger values of $\beta$ and $M$).

The three waveforms from Fig. 4a are shown again in Fig. 4c with the multi-valued regions replaced by shocks. As indicated by the red waveform, as a wave propagates it steadily evolves towards a sawtooth shape. The discussion leading to this point has assumed an initially-sinusoidal waveform. However, a somewhat non-intuitive consequence (linearly speaking) of nonlinear acoustics is that the fate of any waveform is a sawtooth, neglecting absorption.

Table 2 (after Table 3-2 of Reference 2) contains values of $\bar{x}$ for a few frequencies and Mach numbers for air and water. In addition, the absorption length $L_a$ (the reciprocal of the absorption coefficient in Np/m) is listed for the same frequencies. This table provides two rough indicators of when nonlinear effects may be important. If the propagation distance is significantly less than $\bar{x}$, then there is insufficient distance for the cumulative effects of nonlinearites to become relevant. If the distances are comparable to or greater than $\bar{x}$, then the relevant length scale against which to compare is $L_a$. If the ratio $L_a / \bar{x}$, also known as the Gol'dberg number, is greater than unity, then a shock can form before significant absorption has occurred. In this case, nonlinear effects need to be considered. If, on the other hand, the Gol'dberg number is less than unity, then absorption attenuates the waveform

*Table 2*

| | Air | | | | Water | | | |
|---|---|---|---|---|---|---|---|---|
| | $\overline{x}\ (m)$ | | | $L_a\ (m)$ | $\overline{x}\ (m)$ | | | $L_a\ (m)$ |
| $\lvert P \rvert$ (dB) | 100 | 120 | 140 | | 220 | 220 | 240 | |
| $M$ | $2x10^5$ | $2x10^4$ | $2x10^3$ | | $6x10^6$ | $6x10^5$ | $6x10^4$ | |
| $f\ (Hz)$ | | | | | | | | |
| 1k | 2300 | 230 | 23 | 1900 | 11000 | 1100 | 110 | 37000 |
| 10k | 230 | 23 | 2.3 | 55 | 1100 | 110 | 11 | 3700 |
| 100k | 23 | 2.3 | 0.23 | 2.6 | 110 | 11 | 1.1 | 370 |
| 1M | 2.3 | 0.23 | 0.023 | 0.053 | 11 | 1.1 | 0.11 | 37 |

before nonlinear effects have become significant.

Once a shock forms, propagation is never the same, as illustrated by the following example. According to linear acoustics, if the amplitude of a source is doubled, the amplitude of the sound is doubled. According to linear theory, which of course eventually breaks down, there is no limit to the amplitude of sound. This is not true in the presence of a shock. Once a shock is formed, there is an upper limit to the amplitude that the sawtooth can reach. In a process known as acoustic saturation, any additional power delivered by a source is dissipated in the shock. It does not go into increasing the amplitude of the wave.

Referring back to Fig. 3b, one may wonder about the significance of the $n = 0$ term in the output of a nonlinear system. Consider the nonlinear system to be a small volume of fluid exposed to a passing high-amplitude sound wave. At low acoustic amplitudes, the response of this volume of gas will be very nearly linear. For instance, the displacement of the gas in the volume will be a linear function of the pressure amplitude. As the pressure undergoes a complete acoustic cycle, the displacement of the volume undergoes a complete acoustic cycle, returning to its starting point after that cycle.

However, at high amplitudes the volume of gas no longer behaves as a linear system; it behaves nonlinearly. According to Fig. 3b, the response to a single-frequency sinusoid pressure, will be a displacement that is described by a combination of sinusoids of frequencies $nf$. The $n = 0$ term corresponds to a net displacement of the gas volume after one cycle. In other words, the volume of gas does not return to its starting point after a complete cycle. This behavior carries over to other variables as well, the general conclusion being that the gas volume does not return to its original state after the passage of a high-amplitude acoustic wave.

The amplitude-dependent speed of sound also gives the space shuttle its characteristic double-boom, heard during reentry to the atmosphere. The waveform in the near vicinity of a supersonic vehicle is composed on a sequence of shocks radiated from various parts of its surface, the nose, wings, tail, etc., as depicted in Fig. 5. As this sequence propagates towards the ground, the higher-amplitude, faster shocks overtake and “consume” the lower amplitude, slower ones, *b*. Given sufficient propagation distance, i.e., sufficient time, eventually only two shocks remain, *c*, resulting in the characteristic N-wave. Notice that the duration of the sequence of shocks lengthens as it propagates. The reason for this stretching is that the front shock travels faster than the back shock. As a result the time interval between the

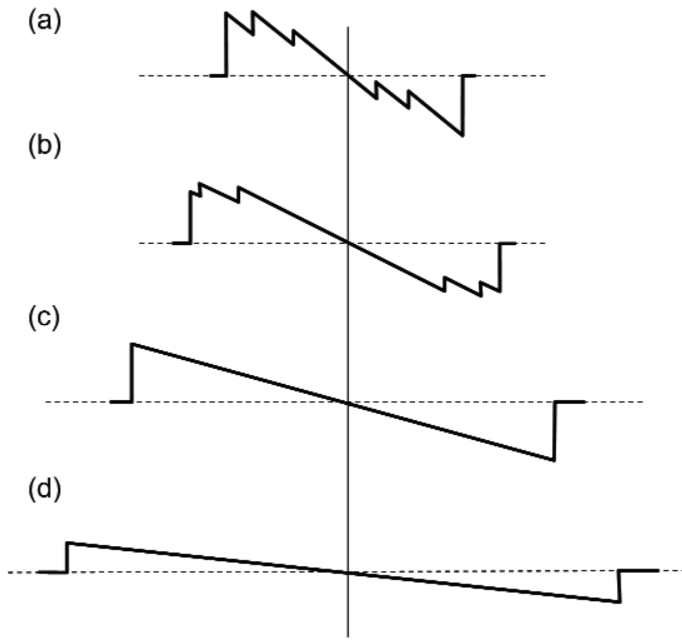


*Fig. 5. A notional diagram depicting the propagation of a waveform initially containing numerous shocks, such as might be generated near a supersonic aircraft. As the propagation distance increases (a – d), the duration of the event increases and the shocks coalesce, forming an N-wave (c).*

two steadily increases with propagation distance. By the time the wave reaches a listener on the ground, the front and back shocks are sufficiently separated in time to be perceived as two distinct events, boom-boom. Because these booms come without warning, they can be startling and their impact annoying. So annoying, in fact, that overland supersonic flight is currently banned.

Advances in computer modeling technology have, however, led recently to improved aircraft designs that could, perhaps, significantly reduce the annoyance caused by supersonic flight. By carefully shaping the aircraft surface, the shape of the sonic boom can be modified, hence reducing its perceived annoyance. This concept was demonstrated during the August 2003 test flights of Northrop Grumman’s Shaped Sonic Boom Demonstrator, a modified U.S. Navy F-5E. At least two companies, Aerion and Supersonic Aerospace International (working with Lockheed Martin), have publicly announced plans to build supersonic business jets.

Up to this point we have considered the nonlinear response to a single input. Now suppose that two initially-sinusoidal sound waves propagate together. Linear theory predicts that the resulting action is simply the sum, or superposition, of the two individual signals acting alone. That is, the sum of a sound wave at $f_1$ and one at $f_2$. However, nonlinearity causes a much more complicated output as depicted in Fig. 6. Based on the previous discussion, one would expect the output to consist of harmonics of the two individual signals, $nf_1$ and $mf_2$. In addition, however, the output contains the sum and difference of all the possible harmonics, $\pm nf_1 + \pm mf_2$. For instance suppose

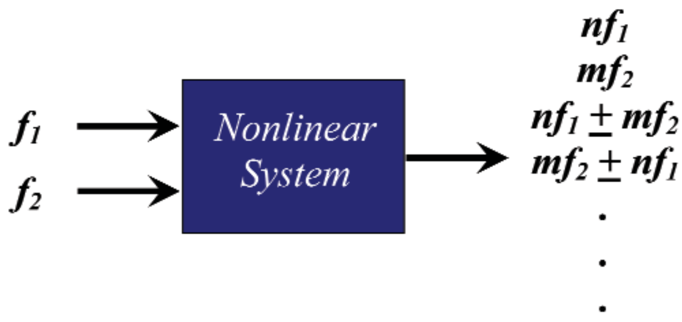


*Fig. 6. If two sinusoidal stimuli are applied to a nonlinear system, the output consists of harmonics of the two input signals, as well as all the possible sum and difference frequencies combinations.*

that $f_1 =$ 40,000 Hz and $f_2 =$ 40,400 Hz, both tones inaudible to humans. As the wave propagates, signals will appear at the harmonics of 40,000 Hz, e.g., 80,000 Hz, 120,000 Hz,…; at the harmonics of 40,400 Hz, e.g., 80,800 Hz, 121,200 Hz, …; and at the sum and difference frequencies, e.g., 400 Hz, 80,400 Hz, …; in other words a quite complicated spectrum. The 400-Hz term is audible.

The absorption coefficient of sound in fluids is, very approximately, proportional to $f^2$. Therefore, the two primary signals as well as their harmonics and sum frequencies are absorbed within a relatively short distance from the source compared to the 400-Hz difference frequency. As a result, only the difference frequency is heard far from the source.

There are certainly easier and more direct ways to generate a 400-Hz tone. In addition, the conversion efficiency from the two primary signals to the difference frequency is very low. So what is the advantage of generating a signal this way? The primary advantage is directivity. The width of a sound beam radiated directly from a source of radius $a$ is proportional to $a/\lambda$, where $\lambda$ is the wavelength of the sound. However, it can be shown that the width of this parametrically generated beam is proportional to $L_a/\lambda$, where $L_a$ is the absorption length of the primary beams. At 40 kHz, $L_a$ is approximately 7 m. Therefore, the beam width of the parametrically generated beam is many times narrower than one generated with a conventional loudspeaker. This property has given rise to the term "audio spotlight."[5] Potential applications include museum displays and public address systems. Although reproducing high-fidelity sound is technically challenging, the listening experience is, nevertheless, quite unique.

A range of physical phenomena precluded by linear acoustic theory are manifest at high amplitudes. The goal of this article has been to introduce a few of the most basic concepts in nonlinear acoustics, point out how these concepts differ from linear acoustic theory, and discuss practical consequences of them. In so doing, it is hoped that readers are 1) aware of a branch of acoustics with which they may have been previously unfamiliar, and 2) better able to explore this intriguing field further on their own.**AT**

## References

1. M.F. Hamilton and D.T. Blackstock, eds., *Nonlinear Acoustics*, Academic Press, San Diego (1998).

2. R.T. Beyer, *Nonlinear Acoustics*, Acoustical Society of America, Melville (1994).

3. K. Naugolnykh and L. Ostrovsky, *Nonlinear Wave Processes in Acoustics*, Cambridge University Press, Cambridge (1998).

4. R.T. Beyer, "The parameter B/A," Ch. 2 in Ref 1.

5. "Audio spotlight" is a registered trademark of Holosonic Research

Anthony Atchley received a B.A. in Physics from the University of the South, an M.S. in Physics from the New Mexico Institute of Mining and Technology, and a Ph.D. in Physics from the University of Mississippi. From 1985—1997, he was a faculty member in the Physics Department of the Naval Postgraduate School. He has been awarded the ASA's F.V. Hunt Postdoctoral Fellowship and the R. Bruce Lindsay Award, and is a Fellow of the Society. He has conducted research in a broad range of areas including measurements of jet noise from both military and commercial aircraft, sonic booms, nonlinear acoustics, thermoacoustic heat transport, optical imaging of sound fields, and ultrasonics. Since 1997, he has served as Head of Penn State's Graduate Program in Acoustics and is currently President-Elect of the ASA.