The Average-Frequency Term in Acoustic Beat Phenomena

Jim Kelley, Peter M. Austin

May 2026

Abstract

Introductory explanations of acoustic beats often begin with a correct trigonometric identity: the sum of two equal-amplitude sinusoidal tones can be rewritten as the product of a slowly varying envelope term and a carrier term at the arithmetic mean frequency. The identity is useful, but it is often interpreted too literally. This note argues for a narrower and safer statement: in ideal linear superposition, two pure tones at f_1 and f_2 do not generate an additional Fourier spectral component at $f_{\text{avg}} = (f_1 + f_2)/2$. The average-frequency term is a carrier in one algebraic representation of the time-domain waveform, not evidence that a new independent tone has appeared in the spectrum.

The same point can be made without treating sine and cosine as the main story. Each tone may instead be pictured as a rotating pointer. The sound pressure is the horizontal shadow of that pointer, and the beat appears as the two pointers repeatedly line up and drift apart. No average-frequency oscillator has to be introduced.

The distinction matters in teaching. The signed modulation term in the product identity has frequency $|f_1 - f_2|/2$, while the audible beat rate, counted as loudness maxima, is normally $|f_1 - f_2|$. In the pointer picture, the audible beat is the rate at which the two tones re-align, while the half-rate modulation term is a signed-envelope convention. Treating those two quantities as identical invites avoidable confusion. Likewise, finite-window spectrum analysis can show time-varying magnitudes or blurred ridges because of window length, leakage, phase, and resolution; that behavior should not be mistaken for physical variation in the source amplitudes. Auditory perception is also a separate question: a listener may hear beats, roughness, one unresolved pitch, or two resolved tones depending on frequency spacing, level, timbre, and context.

The proposed correction is pedagogical rather than revolutionary. The standard equation should still be taught, but its status should be stated clearly: it is an identity that reorganizes the same signal. It does not by itself prove the existence or audibility of a newly created average-frequency tone.

Keywords: acoustic beats; superposition; Fourier spectrum; average frequency; rotating pointer model; modulation; spectrogram; auditory perception

1 Introduction

When two nearby pure tones are played together, the combined sound can pulse in loudness. This familiar effect is usually called beating. A standard classroom explanation writes the sum of two sinusoids as a product of a slow term and a fast term. The fast term is then described as the average frequency, while the slow term is described as the beat.

The algebra is correct. The difficulty appears when the product form is treated as if it proves that a new physical spectral component has been created at the arithmetic mean frequency. That stronger interpretation is not correct for ideal linear superposition of two pure tones.

The purpose of this note is to separate four ideas that are often compressed into one classroom statement:

1. the algebraic identity,
2. the equivalent rotating-pointer description,
3. the Fourier spectral content,
4. the finite-window display produced by an analyzer, and
5. the pitch or beating that a listener reports.

These ideas are related, but they are not interchangeable.

2 The Equal-Amplitude Identity

Let two equal-amplitude pure tones be

$$y(t) = A\cos(2\pi f_1 t) + A\cos(2\pi f_2 t). \tag{1}$$

The sum-to-product identity gives

$$y(t) = 2A\cos\left(2\pi\frac{f_1 - f_2}{2}t\right)\cos\left(2\pi\frac{f_1 + f_2}{2}t\right). \tag{2}$$

Define

$$f_{\mathrm{avg}} = \frac{f_1 + f_2}{2}, \qquad f_{\mathrm{mod}} = \frac{|f_1 - f_2|}{2}. \tag{3}$$

Then Equation (2) can be written as

$$y(t) = 2A\cos(2\pi f_{\mathrm{mod}} t)\cos(2\pi f_{\mathrm{avg}} t), \tag{4}$$

up to the sign convention for $f_1 - f_2$.

This equation is an identity. It is not a new physical mechanism. It says that the same time-domain waveform can be described as the product of a carrier-like term and a signed envelope-like term.

3 Beat Rate Versus Signed Modulation

The signed modulation term in Equation (4) has frequency

$$f_{\mathrm{mod}} = \frac{|f_1 - f_2|}{2}. \tag{5}$$

However, loudness maxima occur twice per cycle of the signed cosine envelope, because both the positive and negative peaks correspond to large magnitude. The usual beat rate is therefore

$$f_{\mathrm{b}} = |f_1 - f_2|. \tag{6}$$

This factor-of-two distinction is one of the easiest places for a misleading explanation to enter. A careful teaching version should say: the signed envelope term oscillates at half the difference frequency, while the perceived loudness beat rate is normally the full difference frequency.

4 A Plain Recursive Picture

There is another way to picture the same beat without beginning with the product identity. Imagine each pure tone as a pointer moving around a circle. A higher tone is a pointer that goes around faster; a lower tone is a pointer that goes around more slowly.

At each time step n, write the pointer for tone i as

$$u_{i,n} = \begin{pmatrix} x_{i,n} \\ y_{i,n} \end{pmatrix}. \tag{7}$$

The two numbers $x_{i,n}$ and $y_{i,n}$ simply say where the pointer is on the circle. The update from one step to the next rotates the pointer:

$$u_{i,n+1} = R_i u_{i,n}. \tag{8}$$

The important feature is not the particular notation for R_i. The important feature is that the pointer keeps the same length as it turns:

$$\|u_{i,n+1}\| = \|u_{i,n}\|. \tag{9}$$

So this update does not make the tone grow or shrink. It only advances its phase.

The sound pressure contribution can be read as the horizontal shadow of the pointer:

$$p_{i,n} = A_i x_{i,n}. \tag{10}$$

For two tones, the measured pressure is just the two shadows added together:

$$p_n = A_1 x_{1,n} + A_2 x_{2,n}. \tag{11}$$

This picture is useful because it keeps the two source tones separate. There is one rotating pointer for f_1, and one rotating pointer for f_2. The beat comes from how well the two pointers line up as they go around.

That line-up can be measured by the dot product

$$u_{1,n} \cdot u_{2,n} = x_{1,n} x_{2,n} + y_{1,n} y_{2,n}. \tag{12}$$

When the dot product is large and positive, the pointers are mostly pointing the same way. When it is negative, they are mostly opposed. When it is near zero, they are roughly at right angles.

The combined size of the two-pointer state is

$$B_n^2 = \|A_1 u_{1,n} + A_2 u_{2,n}\|^2. \tag{13}$$

For unit-length pointers, this expands to

$$B_n^2 = A_1^2 + A_2^2 + 2A_1 A_2 (u_{1,n} \cdot u_{2,n}). \tag{14}$$

This is the key point in plain terms: the beat is the changing line-up between the two original tones. We do not need to add a third pointer at the average frequency. The two original pointers are enough.

Optional technical form

The pointer update may be written explicitly as a norm-preserving rational rotation,

$$R(a_i) = \frac{1}{1+a_i^2}\begin{pmatrix} 1-a_i^2 & -2a_i \\ 2a_i & 1-a_i^2 \end{pmatrix}. \tag{15}$$

This matrix satisfies

$$R(a_i)^T R(a_i) = I, \qquad \|u_{i,n+1}\|^2 = \|u_{i,n}\|^2. \tag{16}$$

Thus it advances phase without changing the pointer length.

For two source tones, the two updates remain separate:

$$u_{1,n+1} = R(a_1)u_{1,n}, \qquad u_{2,n+1} = R(a_2)u_{2,n}. \tag{17}$$

The combined pointer state is

$$U_n = A_1 u_{1,n} + A_2 u_{2,n}. \tag{18}$$

Its squared size is

$$B_n^2 = \|A_1 u_{1,n} + A_2 u_{2,n}\|^2. \tag{19}$$

Expanding gives

$$B_n^2 = A_1^2\|u_{1,n}\|^2 + A_2^2\|u_{2,n}\|^2 + 2A_1A_2\, u_{1,n}\cdot u_{2,n}. \tag{20}$$

For unit-length pointers, this reduces to Equation (14). This is the formal version of the same classroom picture: the beat term is the changing dot product between the two original tones.

5 How The Beat Rate Appears

The recursive picture also explains the factor-of-two issue in a more physical way. The audible beat is controlled by how often the two pointers come back into the same relative alignment.

If the two source tones have frequencies f_1 and f_2, then one pointer gains on the other at the difference rate

$$|f_1 - f_2|. \tag{21}$$

So the time between repeated alignments is

$$T_b = \frac{1}{|f_1 - f_2|}. \tag{22}$$

The corresponding beat rate is

$$f_\mathrm{b} = \frac{1}{T_b} = |f_1 - f_2|. \tag{23}$$

This is the rate at which the loudness swells repeat. For example, tones at 440 Hz and 444 Hz produce four loudness swells per second, so the beat rate is 4 Hz.

The signed envelope in the product formula is slightly different. Its positive and negative peaks both sound loud, because loudness follows size rather than sign. Therefore the signed envelope completes one full signed cycle in twice the beat interval:

$$T_\mathrm{signed} = 2T_b. \tag{24}$$

That gives

$$f_\mathrm{mod} = \frac{1}{T_\mathrm{signed}} = \frac{f_\mathrm{b}}{2}. \tag{25}$$

In short: the beat rate is the rate at which the two tones re-align; the signed modulation frequency is half that rate because it keeps track of sign as well as size.

Optional technical form

The same idea can be written as a relative update between the two pointers:

$$Q = R(a_1)^{-1}R(a_2). \tag{26}$$

The overlap between the two pointers is

$$m_n = u_{1,n} \cdot u_{2,n}. \tag{27}$$

For planar rotations of this kind, this may be written as

$$m_n = u_{1,0} \cdot Q^n u_{2,0}. \tag{28}$$

The audible beat is governed by the return of this relative alignment, not by a newly created average-frequency component. Let N_b be the smallest step count for which

$$Q^{N_b} \approx I. \tag{29}$$

For time step Δt,

$$f_{\text{b}} = \frac{1}{N_b \Delta t}. \tag{30}$$

The signed envelope returns to the same signed state after twice that interval:

$$N_{\text{signed}} = 2N_b. \tag{31}$$

Therefore

$$f_{\text{mod}} = \frac{1}{N_{\text{signed}}\Delta t} = \frac{1}{2N_b\Delta t} = \frac{f_{\text{b}}}{2}. \tag{32}$$

This is the formal version of the same teaching point: audible beats follow repeated relative alignment, while the signed envelope convention gives half that rate.

6 More Than Two Tones

The same picture scales to a more complicated sound. A guitar note, for example, is not usually just one pure tone. It contains a fundamental and a set of partials. In the pointer picture, each component gets its own pointer:

$$p_n = \sum_i A_i x_{i,n}. \tag{33}$$

The measured pressure is still just the sum of the horizontal shadows.

The combined pointer state can be written as

$$U_n = \sum_i A_i u_{i,n}. \tag{34}$$

Its squared size is

$$B_n^2 = \left\| \sum_i A_i u_{i,n} \right\|^2. \tag{35}$$

For unit-length pointers, this becomes

$$B_n^2 = \sum_i A_i^2 + 2\sum_{i<j} A_i A_j (u_{i,n} \cdot u_{j,n}). \tag{36}$$

The first sum gives the separate strengths of the components. The second sum gives the pairwise line-ups between components. Those pairwise line-ups are where beating, roughness, and slow amplitude changes can appear.

This avoids an unnecessary multiplication of imaginary causes. A complex sound does not require a new average-frequency oscillator between every pair of tones. It requires only the original components and their changing relationships.

Optional technical form

For a chain of tones, each component has its own update:

$$u_{i,n+1} = R(a_i)u_{i,n}, \qquad p_n = \sum_i A_i x_{i,n}, \qquad U_n = \sum_i A_i u_{i,n}. \tag{37}$$

The squared envelope is

$$B_n^2 = \left\| \sum_i A_i u_{i,n} \right\|^2. \tag{38}$$

Expanding it gives

$$B_n^2 = \sum_i A_i^2 \|u_{i,n}\|^2 + 2 \sum_{i<j} A_i A_j \, u_{i,n} \cdot u_{j,n}. \tag{39}$$

For unit-length pointers, this reduces to Equation (36). Again, the extra structure comes from pairwise overlaps among the original components, not from newly invented average-frequency tones.

7 What The Fourier Spectrum Contains

For the ideal signal in Equation (1), the Fourier spectrum contains components at the two original source frequencies, f_1 and f_2. In the ideal linear case, there is no additional independent line at f_{avg}.

The product form can make the average-frequency term look like a new object, but it is a coordinate choice for describing the same waveform. It is similar to changing basis: the description changes, while the signal being described does not acquire an extra source frequency.

For that reason, the safe claim is:

> Linear superposition of two ideal pure tones does not generate an additional Fourier spectral component at the arithmetic mean frequency.

This is different from saying that the average frequency "does not exist." It exists as a useful algebraic quantity. What does not follow from the identity is the existence of a separate spectral line or separately generated audible tone at that frequency.

8 Claim Status

Claim	Status
The sum can be rewritten as carrier times envelope.	Algebraic identity.
The same beat can be written as changing line-up between rotating pointers.	Equivalent representation; no average-frequency oscillator is required.
The ideal Fourier spectrum contains f_1 and f_2.	Standard consequence of linear superposition.
No independent line at f_{avg} appears in the ideal two-tone spectrum.	Spectral claim, testable by controlled analysis.
Listeners always do, or never do, perceive a central pitch.	Psychoacoustic claim requiring a separate experiment.
Textbooks should avoid saying the identity proves audibility of f_{avg}.	Pedagogical recommendation.

9 Finite-Window Analysis

A real spectrum analyzer, FFT display, or short-time Fourier transform does not measure an infinite-duration ideal signal. It measures a windowed portion of the signal. The displayed result therefore depends on sampling rate, window function, window length, bin spacing, phase, leakage, and averaging mode.

A finite-window spectrogram may show ridges that blur together, separate, or vary in apparent intensity. That display can be useful, but it should not be read as direct proof that the physical source amplitudes are changing. The safer interpretation is:

> Windowed analysis can make the component magnitudes appear time-dependent because the measurement window interacts with phase, leakage, and resolution. This is a property of the analysis display, not necessarily a change in the two generated tones.

10 Auditory Perception Is A Separate Question

The absence of a Fourier spectral line at f_{avg} does not by itself settle what a listener hears. The auditory system is frequency-selective, nonlinear in important ways, and sensitive to level, spacing, phase locking, masking, roughness, and musical context. When two tones are close together, a listener may report a pulsing tone, roughness, one unresolved pitch, two resolved pitches, or some mixture of these.

The point of this note is narrower. The product-form equation alone should not be used as proof that the ear receives or hears a newly generated average-frequency tone. Perception must be tested as perception.

11 Suggested Numerical Demonstration

A concise demonstration can make the distinction visible without overstating it. Use

$$x[n] = \cos\left(2\pi f_1 \frac{n}{f_s}\right) + \cos\left(2\pi f_2 \frac{n}{f_s}\right), \tag{40}$$

with

$$f_s = 48000 \text{ Hz}, \qquad f_1 = 440 \text{ Hz}, \qquad f_2 = 444 \text{ Hz}. \tag{41}$$

Then

$$f_{\text{avg}} = 442 \text{ Hz}, \qquad f_{\text{mod}} = 2 \text{ Hz}, \qquad f_{\text{b}} = 4 \text{ Hz}. \tag{42}$$

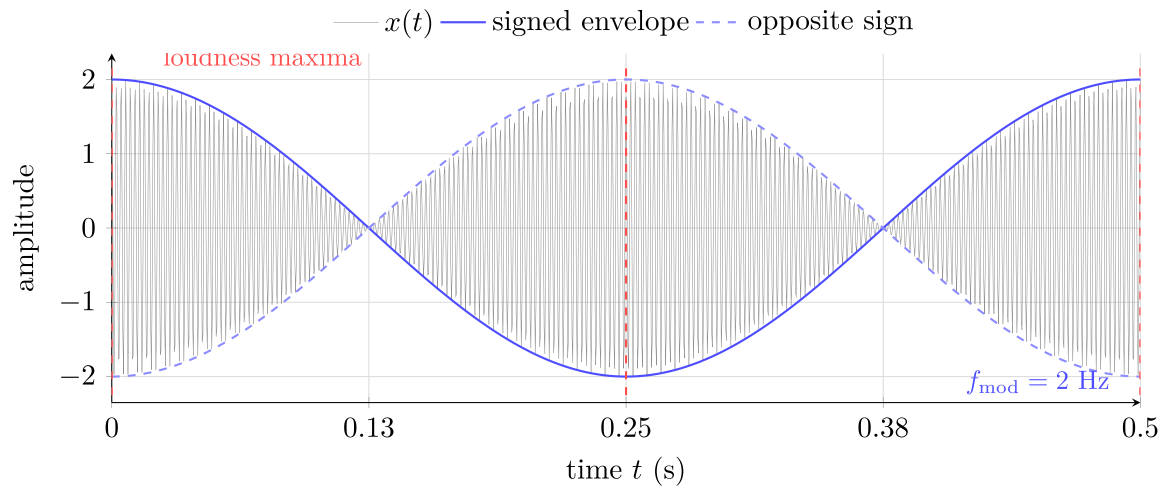

Figure 1: Time-domain beat waveform for $f_1 = 440$ Hz and $f_2 = 444$ Hz. The signed envelope completes two cycles per second, while loudness maxima occur four times per second.

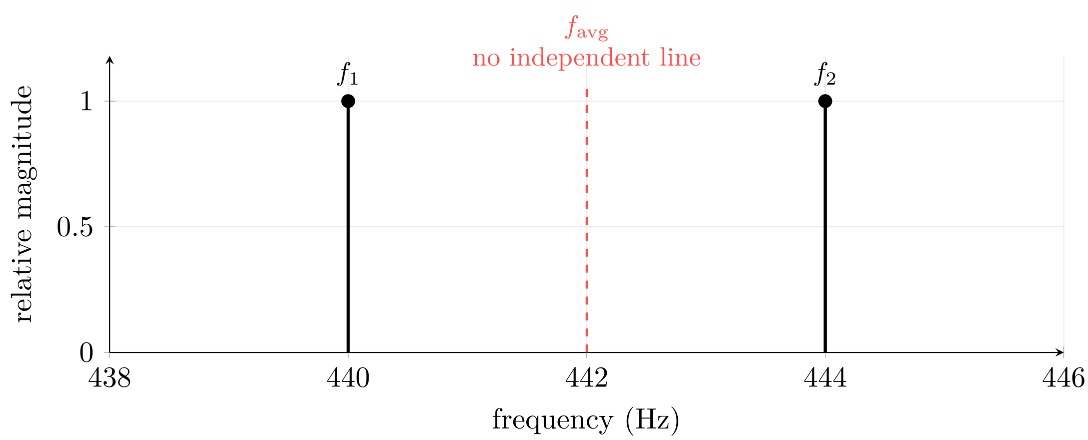

Figure 2: Long-window Fourier spectrum of the same ideal two-tone signal. The expected spectral lines are at the original component frequencies. The dashed midpoint marks $f_{\text{avg}} = 442$ Hz, where no independent line is expected in the ideal linear case.

For a direct comparison with the conventional sinusoidal step, choose each pointer step so that tone i advances by $2\pi f_i / f_s$ radians per sample. In a trig-free computational version, the step can instead be calibrated by counting how many samples it takes the pointer to return to its starting direction.

A useful failure condition should also be stated. If a controlled, low-distortion, linear two-tone setup produced a stable spectral line at f_{avg} above leakage and noise after windowing and distortion controls, that would be evidence for a real nonlinear process or measurement artifact to

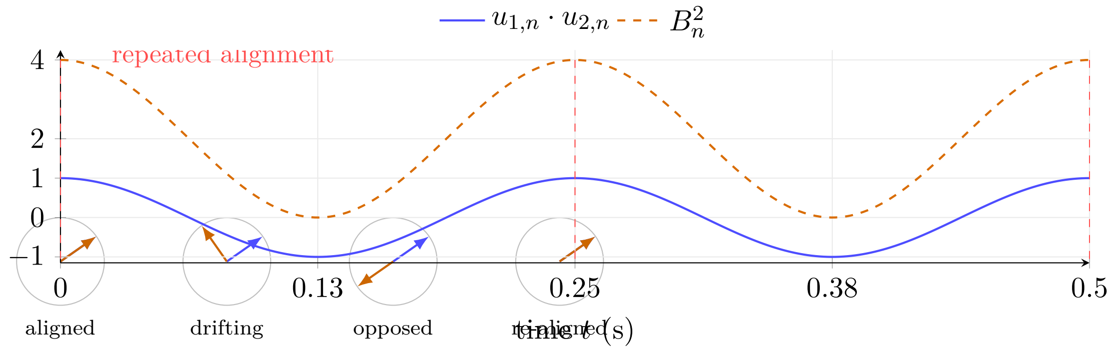

Figure 3: Rotating-pointer view of the same beat. The two source tones remain distinct; the beat appears as the changing line-up between their pointers.

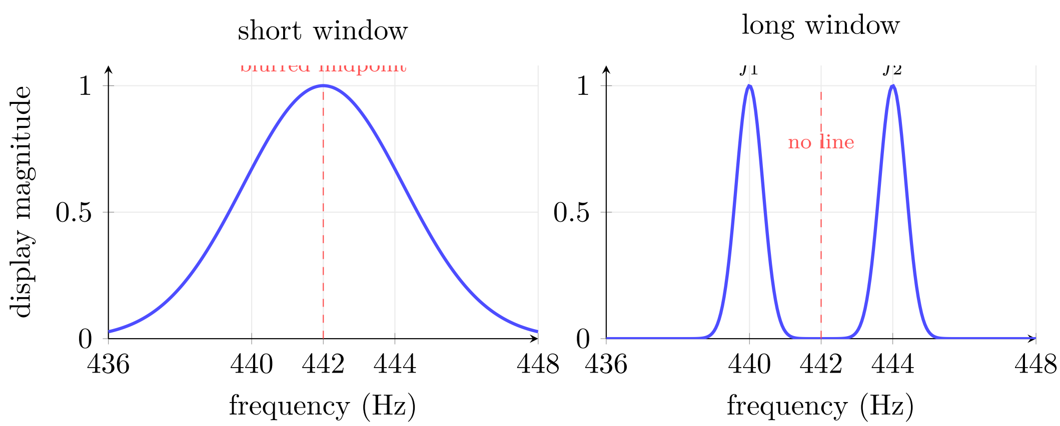

Figure 4: Short-time Fourier displays depend on window length and resolution. A short window may blur the two tones, while a longer window can resolve the two components more clearly.

investigate. It would not follow merely from the trigonometric identity or from the rotating-pointer overlap model.

12 Teaching Recommendation

The common classroom sentence should be revised from:

> Two nearby tones produce a tone at the average frequency that varies in loudness at the beat frequency.

to:

> The sum of two equal-amplitude nearby tones can be rewritten as a carrier-like term at the average frequency multiplied by a signed envelope term. In ideal linear superposition, the Fourier spectrum still contains the original two tones, not a newly generated spectral line at the average frequency.

The same point can be taught with the rotating-pointer picture:

> Each tone remains its own rotating pointer. The beat is the changing line-up between the pointers, and the audible beat rate is the rate at which that line-up repeats.

This keeps the useful identity while removing the misleading physical interpretation.

13 Conclusion

The product-form beat equation is valid and worth teaching. The problem is not the algebra, but the interpretation sometimes attached to it. The average-frequency term is a useful carrier in a time-domain representation. It is not, by itself, evidence that linear superposition has generated a new spectral component or that a listener must hear a separate average-frequency tone.

The rotating-pointer reading reaches the same conclusion from another direction. It keeps the original tones separate and locates the beat in their changing line-up. The average-frequency carrier is therefore not needed as a physical oscillator in order to explain the beat.

The clean pedagogical distinction is therefore:

> The algebraic carrier exists in the representation; the source tones remain distinct; the ideal Fourier components remain at the original frequencies; the perceptual report must be studied separately.

References

[1] J. Wolfe, "Beats," Physclips, University of New South Wales. https://www.animations.physics.unsw.edu.au/jw/beats.htm

[2] F. J. Harris, "On the use of windows for harmonic analysis with the discrete Fourier transform," *Proceedings of the IEEE*, vol. 66, no. 1, pp. 51–83, 1978.

[3] J. Krizman and N. Kraus, "Evolving perspectives on the sources of the frequency-following response," *Nature Communications*, vol. 10, 2019. https://www.nature.com/articles/s41467-019-13003-w

[4] B. C. J. Moore, *An Introduction to the Psychology of Hearing*. Brill, 2012.