ELECTRONOTES 213

Newsletter of the Musical Engineering Group

1016 Hanshaw Road, Ithaca, New York 14850

Volume 22, Number 213 January 2013

REVISITING:

MUSICAL SCALE MATHEMATICS

-by Bernie Hutchins

INTRODUCTION:

Back in the earliest days of *Electronotes* I received a mailing from Sebastian von Hoerner at the National Radio Astronomy Observatory in Green Bank, W. Va. with a paper called "Universal Music" enclosed. It was extremely interesting to me. In the context of one of his main research interests, he speculated on whether or not an extraterrestrial intelligent life who

developed music in some sense might have developed a 12-tone-per-octave scale as humans have. In the process, he was perhaps more to the point saying why we ourselves have 12 tones per octave in our chromatic scale, from which we usually select subsets for such scales a major and minor 7-tone scales, pentatonic scales, and numerous "modes".

I wrote back to him asking where it would be published, or if he was submitting it to me for ***Electronotes***. He was kind enough to offer a version of it to me for ***Electronotes*.** This we published [1] and at least three related papers exist [2-4]. Further I learned that Sebastian was currently at Cornell for the term and had an office that was a full 100 yards from my physics lab. So I met him in person, and heard a seminar presentation of "Universal Music". Yes, Carl Sagan was there – the closest I ever got to actually meeting Carl Sagan!

The importance of "Universal Music" is that I feel it gives not only a possible explanation of why 12 tones are chosen (as providing best approximations to the simple ratios of 3/2, 4/3, 5/3, and 5/4) within a reasonable set of choices of total tones, but a very convincing explanation that has dominated my understanding since then [5]. Further, it was easy to analyze the problem using a computer program with various error criteria [6] and for many years, I gave this problem as a homework exercise in the use of a least squares solution. Until recently, I don't recall thinking of any other explanations, and I still favor the "Universal Music" by a wide margin. Below we will review this, and discuss a fairly recent alternative discussion. But first, I was trying to think of what I had before "Universal Music". It had to be the "circle of fifths".

AROUND THE CIRCLE WITH FIFTHS (or FOURTHS)

The basic idea of the circle of fifth (described for at least 300 years) is that you assume that a perfect musical fifth should be an exact 3/2 ratio. A fifth is five notes of a major scale, like C to G. If we build all notes on fifths (assuming this can be done) what other notes do we get. If you are near a piano, sit down and press the lowest C that is available. Then go up a fifth to G. Another fifth takes you to the D, then to the A, and so on (see Fig. 1). Perhaps if you had to guess, you would have no good idea as to whether or not this was going anywhere useful. Yet we do find we get back to C seven octaves up. So for the piano, it works. If we assume that in addition to a fifth being a perfect 3/2 ratio, an octave is exactly a 2:1 ratio, this high-end C should be at a frequency of 2^7=128, or seven octaves. It's not. It's $(3/2)^{12}$ = 129.74634, or 12 fifth, which comes out a bit sharp to about a 1.36% "error". (Note: In case you were thinking about it, $128^{1/12}$ = 1.4983041 = $2^{7/12}$, the equal tempered fifth – of course!) But, to a very good approximation, <u>we do get 12 tones</u>, based on this alone.

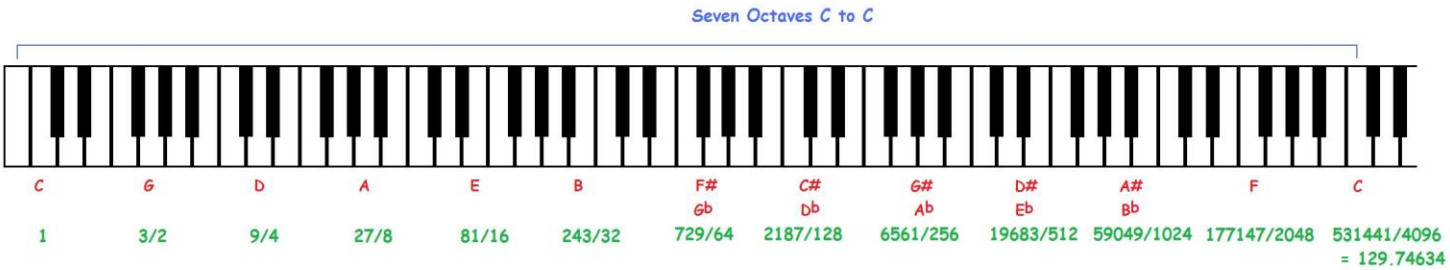


Fig. 1 Circle of Fifths

The reader is invited to try this with a circle of fourths (starting on C and going to F, etc.), assuming a perfect fourth is a 4/3 ratio. You will find this closes back on C after five octaves, but in 12 fourths. The upper C is now at 31.569292 = $(4/3)^{12}$ instead of at 32, so it is bit flat (about 1.35%) this time, instead of 1.36% sharp the way the circle of fifths was. So again, we get all 12 notes. (Likewise, $32^{1/12} = 2^{5/12}$ =1.3348399, the equal tempered fourth). (For reference, recall that a half-tone, like C to C#, is about 6%.) Would it work with other perfect ratios? What about a perfect third = 5/4? Nope, that gives you only three notes (C, E, and G#). What about a perfect sixth? Nope - that gives four notes (C, A,F#, and D#).

We may well have been very comfortable with this explanation. But, if there is a preference for those other consonant ratios, how well are we doing on them?

THE THIRD, FOURTH, FIFTH, and SIXTH : JUST INTONATION

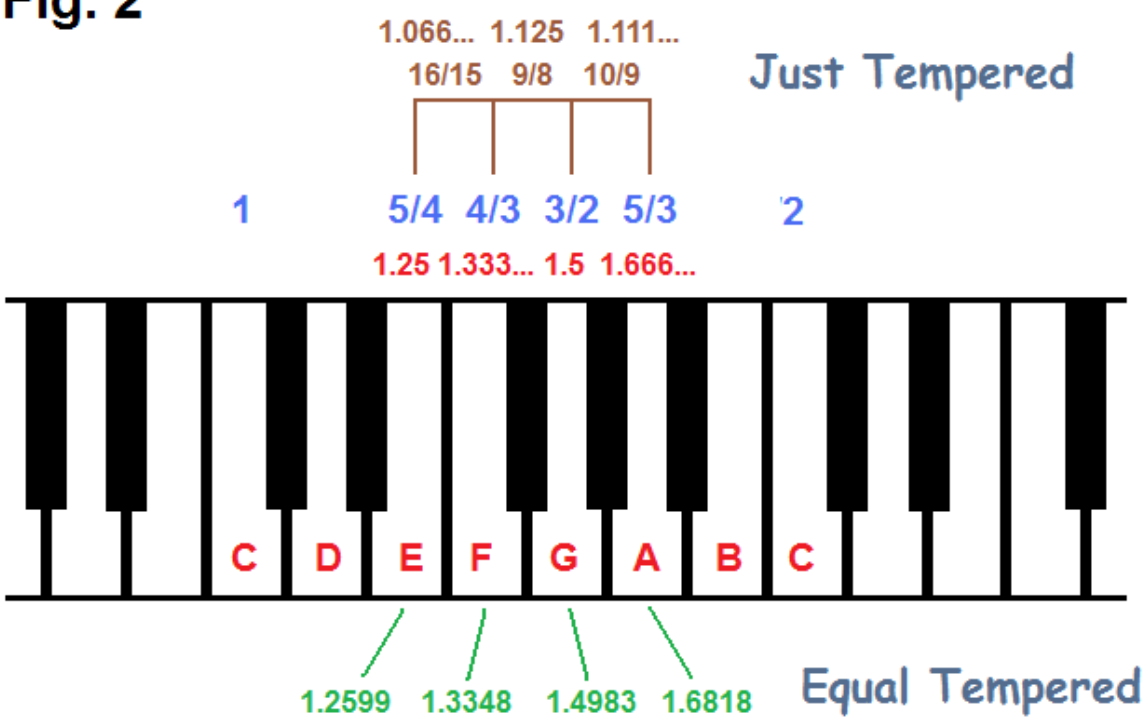
Having decided to look at the circle of fifths, and decided that the fifth should be the ratio of 3/2, what other small integer ratios should we examine? First it may not be obvious that we are even concerned with ratios of small integers. If asked, we might parrot back the idea that exact ratios of small integers are “consonant” while ratios close to small integers (thus ratios of large integers, or truly irrational ones) “beat” and are “dissonant”. This is more complicated than that (see discussion later in this issue). When adding up two sine waves, they may and do beat as an observable (by ear and on a scope) amplitude variation, often very annoying. The same is not generally true of such waveforms as sawtooth, square, or pulse, - those with sudden jumps. We have primary and secondary (subjective) beating. So the exact details matter. Likewise the notion as to whether or not a pair of tones is dissonant is a matter of individual experience and culture. Many people (such as myself) find dissonance just another (welcome) “tone color” that is available. So let’s take the desire to have our scale tones in small integer ratios as an axiom. Note that we do want ratios that are within one octave, between 1 and 2. Our already chosen ratios of 3/2 and 4/3 are of this type we require.

Clearly we might look at continuing from 3/2 to 4/3 and then to 5/4. Indeed, this we will associate with the position of a third (a “major” third), like E of a C major scale. In the opposite direction, we might consider 5/3, what we will take to be the sixth (like A of a C major scale). So now we have a third, a fourth, a fifth, and a sixth. This (with the root) will turn out to be five of the seven tones of what we need for a major scale.

Nothing prevents us from choosing the number of tones and the scale tone pitches any way we like. But even as we agree to this freedom, in choosing simple ratios, we can’t help mentioning a major scale as though it was obviously <u>already there</u> (that why we mentioned C, E, F, G, and A). So as a solid point of reference here, perhaps we do need to mention what is usually done. This is the 12-tone equal-tempered scale. In this development (available for perhaps 200-300 years or so), the 12 tones within an octave are all equally spaced, at the same RATIO. This ratio is the “twelfth root of two” which is about 1.0594637… so that:

$$[\sqrt[12]{2}\,]^{12} = 2 \qquad (1)$$

Fig. 2



$$\sqrt[12]{2} = 1.0594631\ldots \qquad [\sqrt[12]{2}\,]^2 = 1.122462\ldots$$

The tones generated with equal temperament (see below) are not small integer ratios. Indeed we are looking to choose a musically transposable scale that we can use that approximates the actual tones (that we might ideally prefer) to a very satisfactory degree. In what we do below, we are talking about small integer ratios, what is usually called “just tuning” but we will look to the equal-tempered tuning ratios as a very familiar and useful reference for comparison.

CIRCLE OF FIFTHS

Tone	Circle	First Oct.	Ordered	Eq. Temp	Just	C Scale
0	1.0000	1.0000	1.0000	1.0000	1	C
1	1.5000	1.5000	1.0679	1.0595		
2	2.2500	1.1250	1.1250	1.1225		
3	3.3750	1.6875	1.2014	1.1892		
4	5.0625	1.2656	1.2656	1.2599	5/4=1.25	E
5	7.5938	1.8984	1.3515	1.3348	4/3=1.333...	F
6	11.3906	1.4238	1.4238	1.4142		
7	17.0859	1.0679	1.5000	1.4983	3/2=1.5	G
8	25.6289	1.6018	1.6018	1.5874		
9	38.4434	1.2014	1.6875	1.6818	5/3=1.666...	A
10	57.6650	1.8020	1.8020	1.7818		
11	86.4976	1.3515	1.8984	1.8877		
12	129.7463	2.0273	2.0273	2.0000		

CIRCLE OF FOURTHS

Tone	Circle	First Oct.	Ordered	Eq. Temp	Just	C Scale
0	1.0000	1.0000	1.0000	1.0000	1	C
1	1.3333	1.3333	1.0535	1.0595		
2	1.7778	1.7778	1.1099	1.1225		
3	2.3704	1.1852	1.1852	1.1892		
4	3.1605	1.5802	1.2486	1.2599	5/4=1.25	E
5	4.2140	1.0535	1.3333	1.3348	4/3=1.333...	F
6	5.6187	1.4047	1.4047	1.4142		
7	7.4915	1.8729	1.4798	1.4983	3/2=1.5	G
8	9.9887	1.2486	1.5802	1.5874		
9	13.3183	1.6648	1.6648	1.6818	5/3=1.666...	A
10	17.7577	1.1099	1.7778	1.7818		
11	23.6770	1.4798	1.8729	1.8877		
12	31.5693	1.9731	1.9731	2.0000		

Fig. 2 and the table on page 5 just above contain a goodly amount of information. Our overall goal is to explain how the four tones E, F, G, and A are positioned, and to set a background for the two additional tones we need (D and B, as well as the "black keys" to some degree). In the just tuning version, we have decided to give the four initial tones ratios of 5/4, 4/3, 3/2, and 5/3 respectively. Integer ratios are lovely, but we <u>also</u> need to express them as decimals. For example, we will be discovering the ratios 9/8 and 10/9 – which one is larger – right off the top of your head!

Fig. 2 shows a keyboard, with the notes of a C major scale marked in red. Above the keyboard in blue we show the <u>chosen</u> ratios 5/4, 4/3, 3/2, and 5/3, with their decimal equivalents below the blue numbers, in red. Below the keyboard, in green, are the decimals for the corresponding equal-tempered scale. These are the numbers:

$$[\sqrt[12]{2}\,]^{n} \qquad (2)$$

for n = 4, 5, 7, and 9. We note the very impressive agreement with the decimal versions of the simple ratios (green and red decimals). This is kind of the whole message. This good agreement is a bit of luck. It is not luck that we chose a total number of tones that worked best (for a relatively small number of tones total), but it does seem to have worked out better than we had much right to suppose could happen.

Next we look at the dissection of the small integer ratios (brown numbers) shown above the keyboard. For example, we had 5/4 and 4/3. So in going from E to F the frequency increased by 16/15= 1.0666… . This is a musical "half-step" in this case. When we go from F to G, a musical "full-step" the ratio, from 4/3 to 3/2, is 9/8 or 1.125. Appropriately enough, a full-step seems to be about twice half-step. The third jump from G to A is from 3/2 to 5/3, which multiplies by 10/9 or 1.111… in this case. So it's similar, but not the same. Again a full-step is about twice the half-step. In summary here, we find one choice for a half step (16/15) and two possible choices for a full step (9/8 and 10/9). It is very useful as well to note the equal tempered ratios (black numbers below the keyboard) for a half-step, and for a full step. Note that the equal tempered ratio for a full step (1.1224621…) is between 10/9 (smaller) and 9/8 (larger). So everyone is more or less in "ballpark" agreement.

We could have laid out this scenario by trying different numbers of total tones per octave. Indeed we have done this before [6], and will look at this again shortly here. However, we have also reviewed the circle of fifths and the circle of fourths as a hint about the choice of 12 tones. The table on page 5 shows the data for these two choices - the upper half being for fifths, and the lower half for fourths. There are seven columns in the table. The first column is just the number of the tone from n=0 to n=12. Column 5 shows the equal-tempered ratios, including the four cases from Fig. 2, green numbers. Columns 6 and 7 repeat the information for the four small integer ratio cases from Fig. 2.

So now back to column 2. This is the frequency ratios for the circles. So it is $(3/2)^n$ for the upper portion, and $(4/3)^n$ for the lower portion. These quickly exceed 2, so are returned to the first octave (column 3) by dividing by the appropriate power of 2. The third column is thus a bit scrambled. But we have now located 12 different tones, and these can be put in ascending order, column 4.

We directly compare the ordered columns (column 4) to the equal-tempered ratios of column 5, and the four just-tempered choices of column 6. Note that the circles provide sequences of tones that are similar to the equal-tempered methods. The circle of fifths is of course perfect for the G, and the circle of fourths is perfect for the F, by definition. But all and all, not a bad set of numbers supporting the notion that 12 tones is “right” and pretty much an “accident” of a preference for small integer ratios, and our desire to have total tones in a general range (finite resolution of frequencies by the ear, and finite length hands!).

So that’s really only four notes (five with the C included), and we wanted two more for a major scale, and seven more for the full chromatic scale. What we have to work with is the suggestion that a half-step might be 16/15, and a full-step 9/8 and/or 10/9. How does this help?

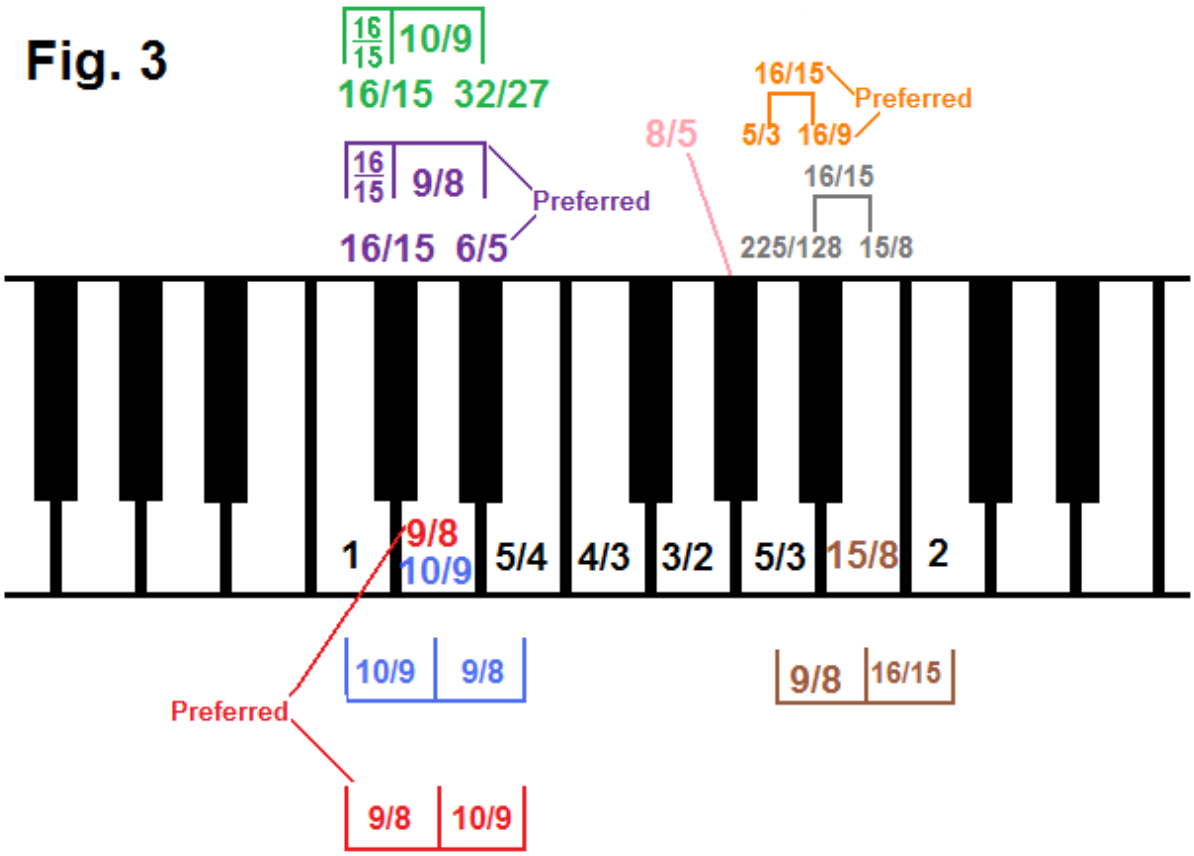


Fig. 3 shows some obvious manipulations. Let's start by figuring out how we can get a note B between A (5/3) and the octave C (2). We have a full step (A to B) and a half step (B to C) to get from A to C. Can we build this from our proposed ratios? Yes, if we place the B at a full-step (chosen as the 9/8 option) above 5/3, there is exactly 16/15 more needed to reach C. So a ratio of 15/8 (approaching 2) for B seems locked in (brown numbers in Fig. 3).

The choice for setting D is not so clear. But note that we need two full steps to get to E, and that the suggested ratios (10/9) and (9/8), in either order, produce 5/4, the needed value for E (blue or red numbers in Fig. 2). If we had to make a choice, we have that:

$$(10/9) = 1.1111\ldots$$

$$+ 1\%$$

$$(2^{2/12}) = 1.122462\ldots$$

$$+ 0.226\%$$

$$(9/8) = 1.125$$

so the choice of 9/8 first (red in Fig. 3) is closer to the equal-tempered result.

So that would get us pretty well to a major scale. It is perhaps surprising that if we go through all our choices, we have:

$$(9/8)(10/9)(16/15)(9/8)(10/9)(9/8)(16/15) = 2$$

This scale only works in C major if we insist on the just tuning. But we are mainly interested in how this leads us to <u>the choice of the total number of tones</u> for the equal-tempered version which permits all keys and modulations (in the musical sense) between keys.

But we do want to consider a couple more tones here. Two are ones needed for the minor scale. This means that we also need to know the ratios for E^b and B^b. Once we get to the "black keys" things get more complicated as we have more possibilities. For the moment, let's assume we still insist on just using (16/15), (10/9), and (9/8) as possible ratios. This gives the possibilities of Fig. 3 above the keyboard. To reach the E^b from C, two possibilities are shown (purple and green), for which we likely prefer the purple, placing a 9/8 full tone between $C^\#$ and E^b rather than the 10/9 fill tone, as this gives us the ratio of the lower integers 6/5, and a nice continuation of the 3/2, 4/3, 5/4 downward sequence. E^b chosen as 6/5 is also 16/15 up from the 9/8 selected for D. For the B^b, we have options of going up from A or down from B as shown (orange and grey). The lowest integers are in the ratio that puts B^b at 16/9 (orange). Although the details are not shown, A^b is best obtained (similarly) as 16/15 up from G = 3/2, at 8/5. $F^\#$ is not clear. These choices of E^b and B^b can be compared to equal temperament, for E^b as folows:

(32/27) = 1.1851852….

+ 0.34%

($2^{3/12}$) = 1.1892071… (equal tempered E^b)

+ 0.9%

(6/5) = 1.2

and for B^b:

(225/128) = 1.7578125

+ 1.1%

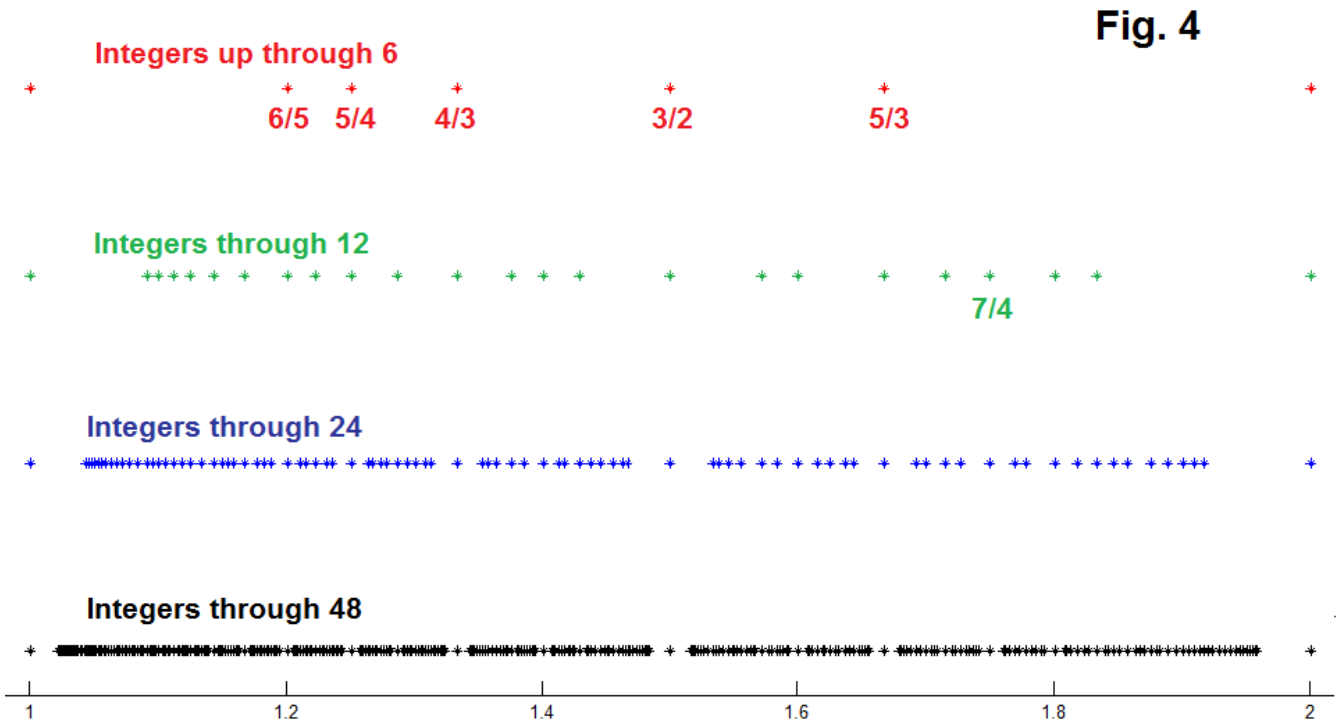
(16/9) = 1.7777777…

+ 0.23%

($2^{10/12}$) = 1.7817974…. (equal tempered B^b)

Another possibility for B^b (not shown in Fig. 3) would be 9/5 = 1.8, which is 9/10 <u>down</u> from 2, just using one of our full tone ratios. This would involve a new semitone ratio, 27/25 = 1.08 up from A.

The reader may well ask "Who made the silly rule of only using 16/15, 10/9, and 9/8?" Indeed this is artificial. In fact, the only real "rule" of just intonation seems to be that we want small integer ratios. Nothing speaks directly to the density and uniformity or spacing, or even the total number of tones in any scale, except as a user may wish for a tone not provided. It is interesting to consider what the distribution of small integer ratios might look like. Fig. 4 shows the possible ratios <u>that are between 1 and 2</u> for choices of a maximum integer of 6, 12, 24, and 48. Note that the choice of a maximum integer of 6 yields just five choices, the four we used initially (Fig. 2) plus the 6/5 we subsequently added (for E^b). It is further curious that if you look at the cases below these, <u>they</u> of course remain, but they seem to "repel" nearby competition. Well, not exactly that. For example, note that 5/4 (1.25) is present, but in the red case, 7/4 is not – because we didn't allow a 7 yet. By the green case, 7/4 pops up, as does, for example, 7/5 (at 1.4), and these continue downward. Getting close to one small integer ratio would mean that one of the integers would have to change slightly. For example, if we had 3/2 = 1.5 and wanted another ratio very close to it, that might be represented by 3.01/2 or 301/200, and integers this large are not allowed yet, and won't appear for a while.



It is clear at this point that we have plenty of rational ratios to choose from if we wanted to just have just intonation. But the advantages of **small** integer ratios become rapidly unavailable if we try too hard to arrive at a particular decimal fraction. Just out of curiosity, what would happen if we tried to represent an equal tempered scales as ratios? This is easy to look at with Matlab's ***rats*** function. The ***rats*** function takes any decimal number and attempts to find, within some precision that can be controlled (but not be made infinite!), a rational approximation. The sequence $2^{n/12}$ for n= 0 to 12 becomes:

1 196/185 55/49 44/37 349/277 295/221 239/169 442/295 227/143 3002/1785 1527/857 185/98 2

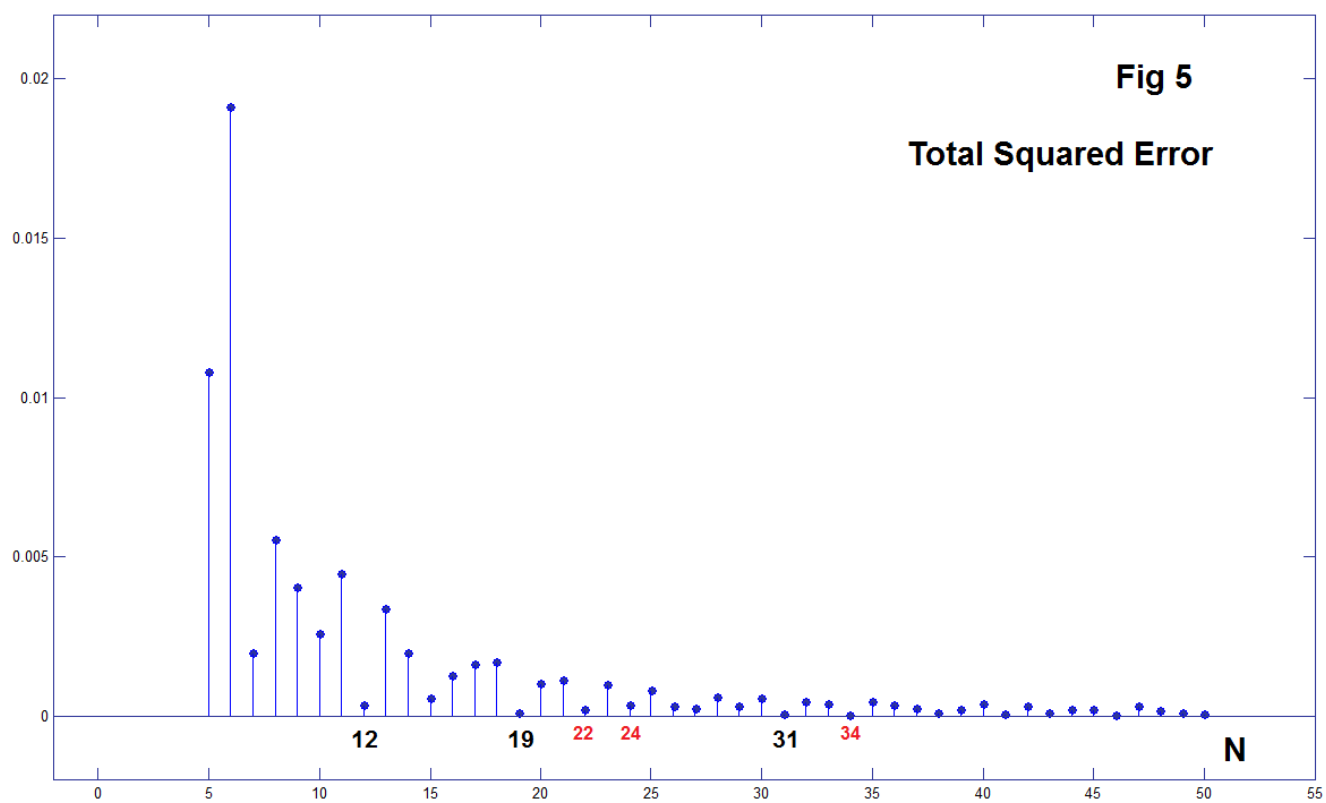
Just in case anyone supposed this would work.

OTHER EQUAL-TEMPERED SCALE POSSIBILITIES

It is not obvious (to me at least) why an equal tempered scale of 12 tones should work as well as it does. I guess some would wonder why it doesn't work better. But the obvious ploy is to start with the assumption that all tones should be placed at equal ratios and we want good approximations to the preferred small integer ratios. What number of tones should we choose? We look for small error overall, in some sense, perhaps mathematically derived, or perhaps just fit by hand (as in von Horner's "Universal Music" as he first presented it here [1]).

I have worked on this problem many times, and as mentioned, assigned it as a homework problem many times (students love real examples of course). Further, it is a good example of an engineering solution – a best answer given specified resources.

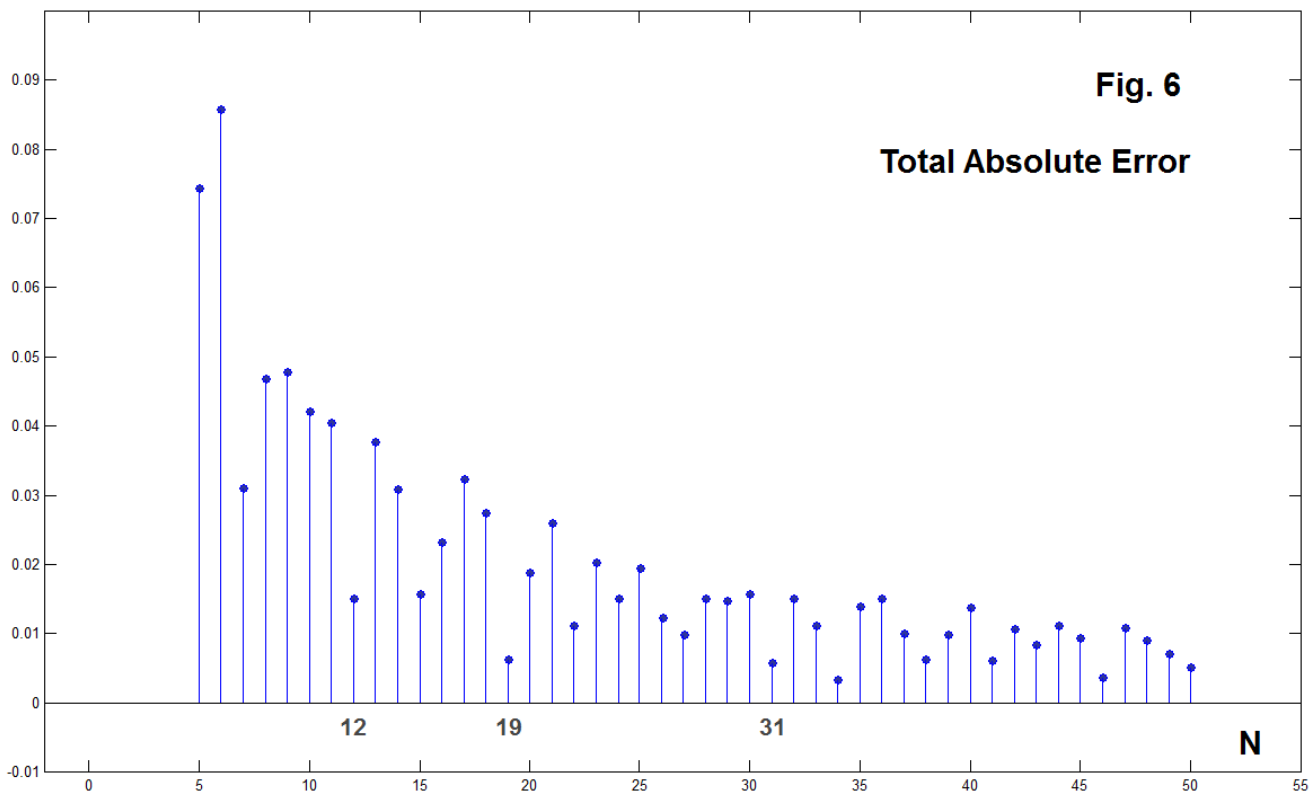
The problem is set up by first specifying the ratios 5/4, 4/3, 3/2, and 5/3 as the goal. The proposed solution is to approximate these by selecting the closest ratios among those available for some N-tone equal tempered scale: $2^{n/N}$ where n runs from 1 to N. This is done by a computer search. It is easy to locate the equal-tempered tone that is closest to the desired ratio. Then we can use some measure of the fit for a particular N. First and foremost the total squared error is obviously one choice (Fig. 5). Other choices would be the total absolute error, and the maximum absolute error. Here we will also look at the fit to the individual ratios one at a time for additional insight. Some Matlab program we have used are posted [12]. We have chosen to look at scales from 5 tones to 50 tones. Scales of fewer than 5 tones give very large errors and are not of great interest. (This being distinct from the notion of selecting 5 tones from among the larger set of, for example 12 tones, which is clearly much used, as is obviously, the selection of 7 tones from among the 12.)



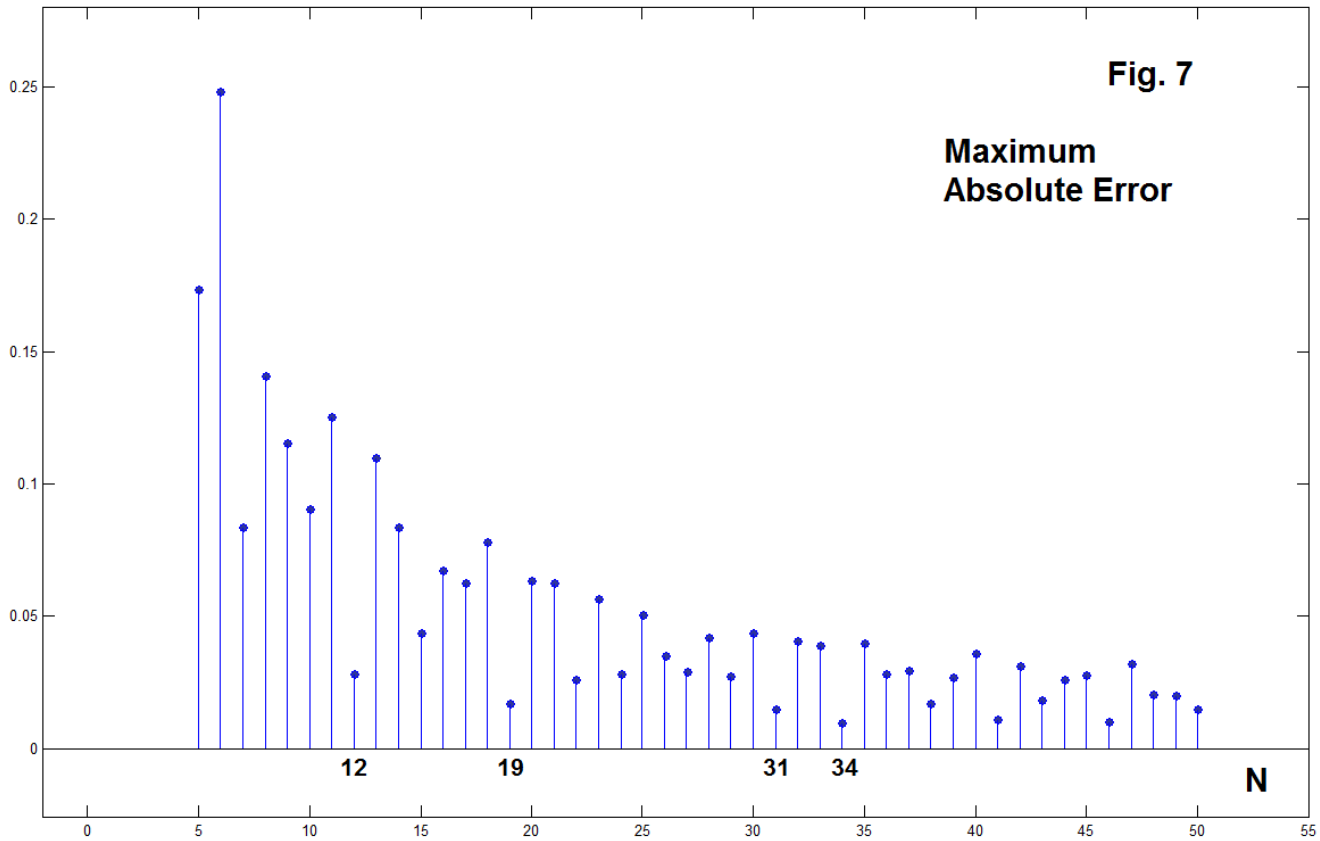
From Fig. 5, we see the famous dip in the error at N=12, and better fits for such larger values of N as 19 and 31. We note as well the general downward trend in the error as N increases – exactly what we expect as our choices get closer together with increasing N. Another point that needs to be made is that this differs slightly from our previous presentation [4] because that presentation normalized the error to the value of the ratio. It makes little difference.

Some additional remarks can be made. It should be clear that whenever we double the size of N, we get back exactly the tones for the original N plus those in between. Thus doubling the size of N can't be any worse, and may be better. Note that if we have a fairly low error for N, then it is probably unlikely that doubling N will have any improvement. For example, for N=12, we have good fits to the four ratios, and no improvement by using N=24. On the other hand, N=11 is a poor fit while N=22 is quite good (better than N=12), so we know that this doubling resulted in additional candidates of value. The change from N=17 to N=34 is clearly another example of this improvement.

The second error criterion we will consider (Fig. 6) is the total absolute error. Instead of squaring the errors (thus emphasizing the larger ones) we add up the error magnitudes on all four ratios. The general result is very similar to the squared error result. We see the same



features at N=12, 19, and 31 (although N=15 and 34 also look good). The third error test is the use of maximum absolute error, and this is shown on Fig. 7. The overall result is similar to Fig. 5 and Fig. 6. To better understand the difference between Fig. 7 and Fig. 6, suppose we compare this to an exam that has four problems to solve. A student fails if the error is too high – naturally enough. Fig. 6 would be the case where we use all four problems to determine the grade (as we normally grade exams). Fig. 7 would be the case where we grade the entire exam based on the one problem with the most error!



Figures 5-7 do not indicate where the error is. This is perhaps fair enough because we are in a kind of pass/fail mode. Music is famous for being sensitive to isolated “clinkers”. The audience dozes until the performer hits one wrong note! If you ever tried to play a piano which had 98.86364% of its keys functioning properly (one bad key!) this is the same idea. Thus above we were looking at <u>an acceptable approximation to all four ratios</u>.

But we can take a look at the error on the individual ratios. We will call these Fig. 8E, Fig. 8F, Fig. 8G, and Fig. 8A corresponding to the C major scale. Fig 8E is immediately attractive as it shows some interesting structure. However, this much structure is not evident in the corresponding three figures that follow. We can take a look at these individually to see

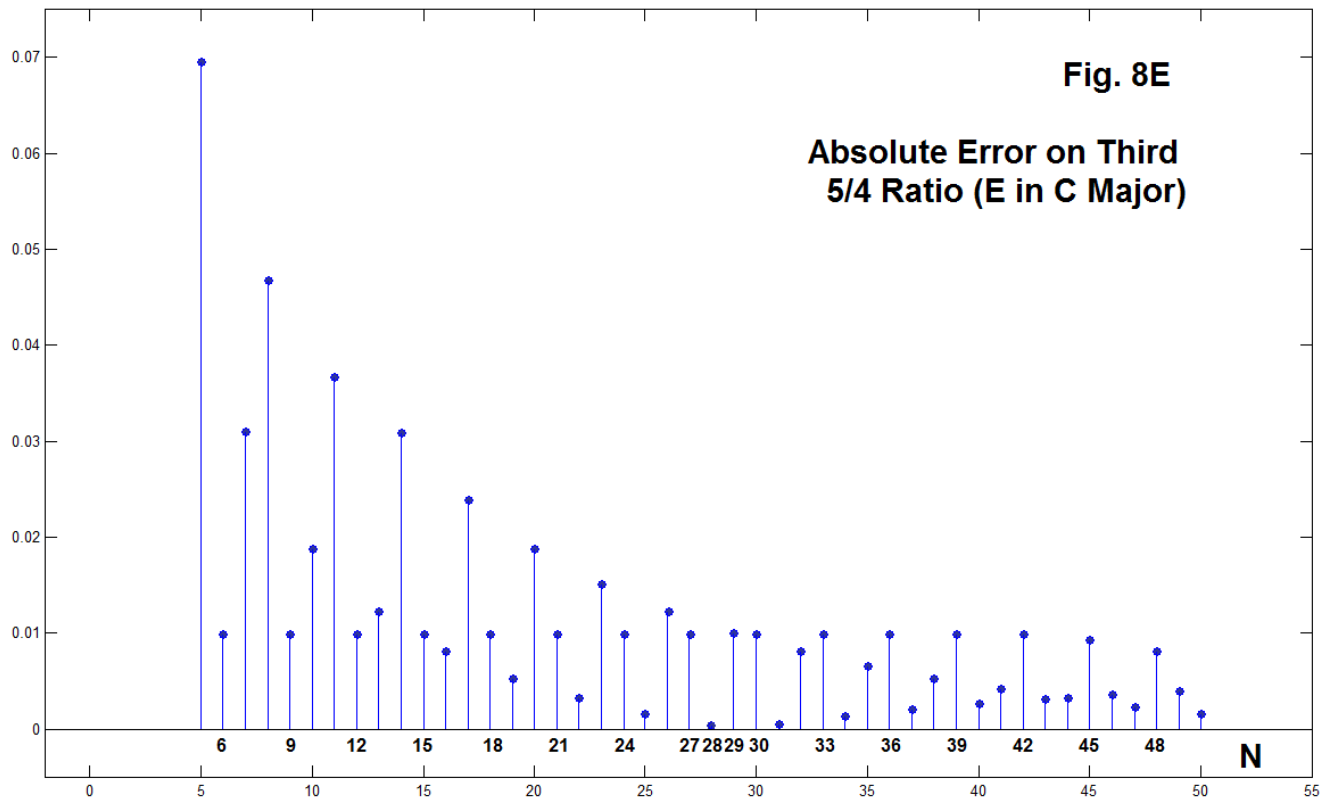


Fig. 8E
Absolute Error on Third
5/4 Ratio (E in C Major)
N

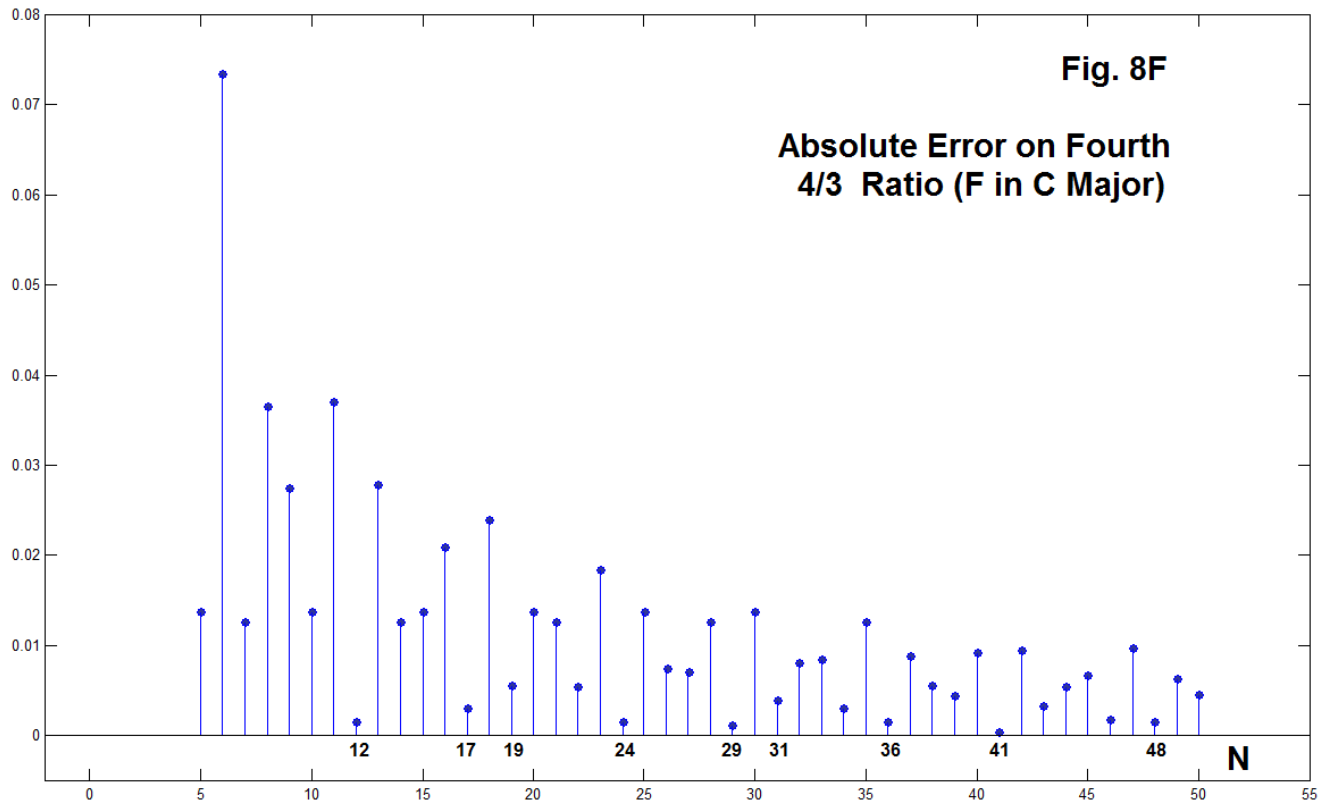


Fig. 8F
Absolute Error on Fourth
4/3 Ratio (F in C Major)
N

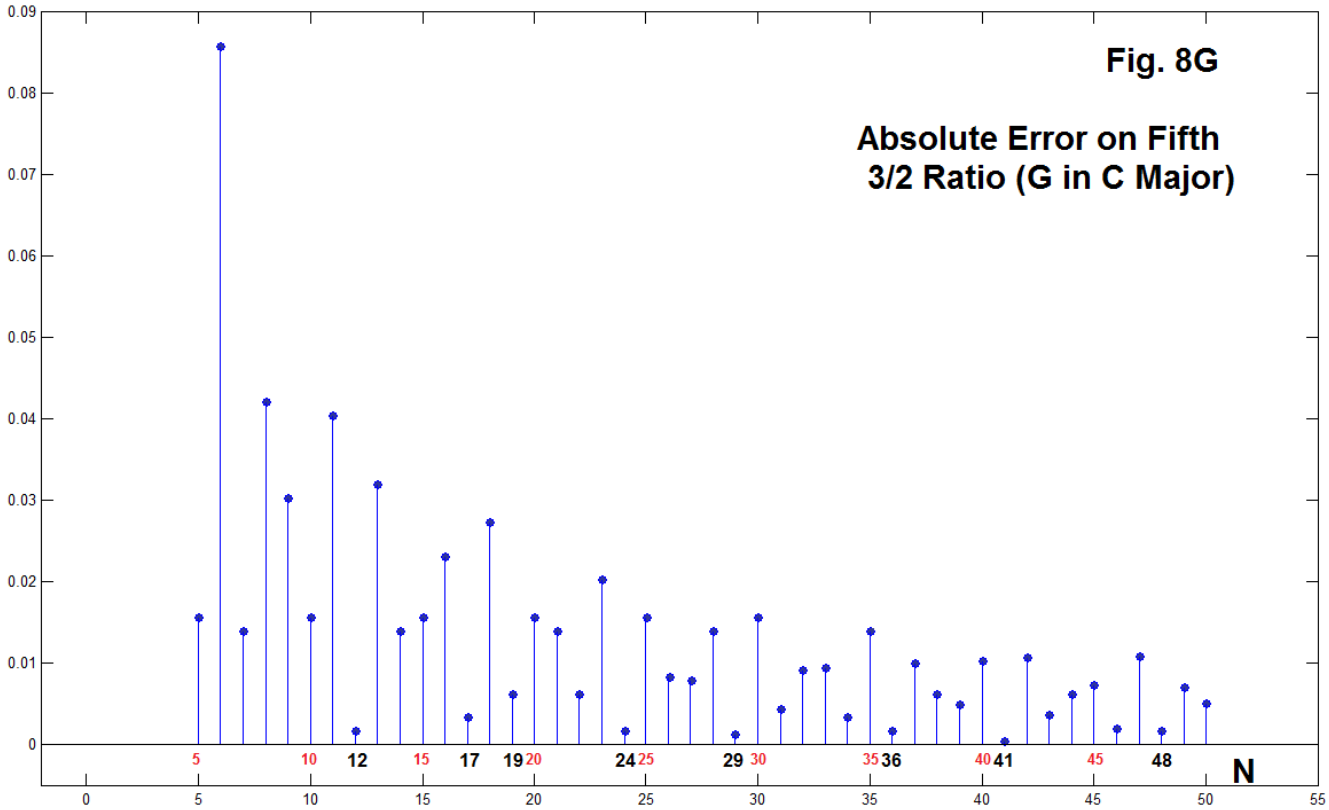


Fig. 8G
Absolute Error on Fifth
3/2 Ratio (G in C Major)
N

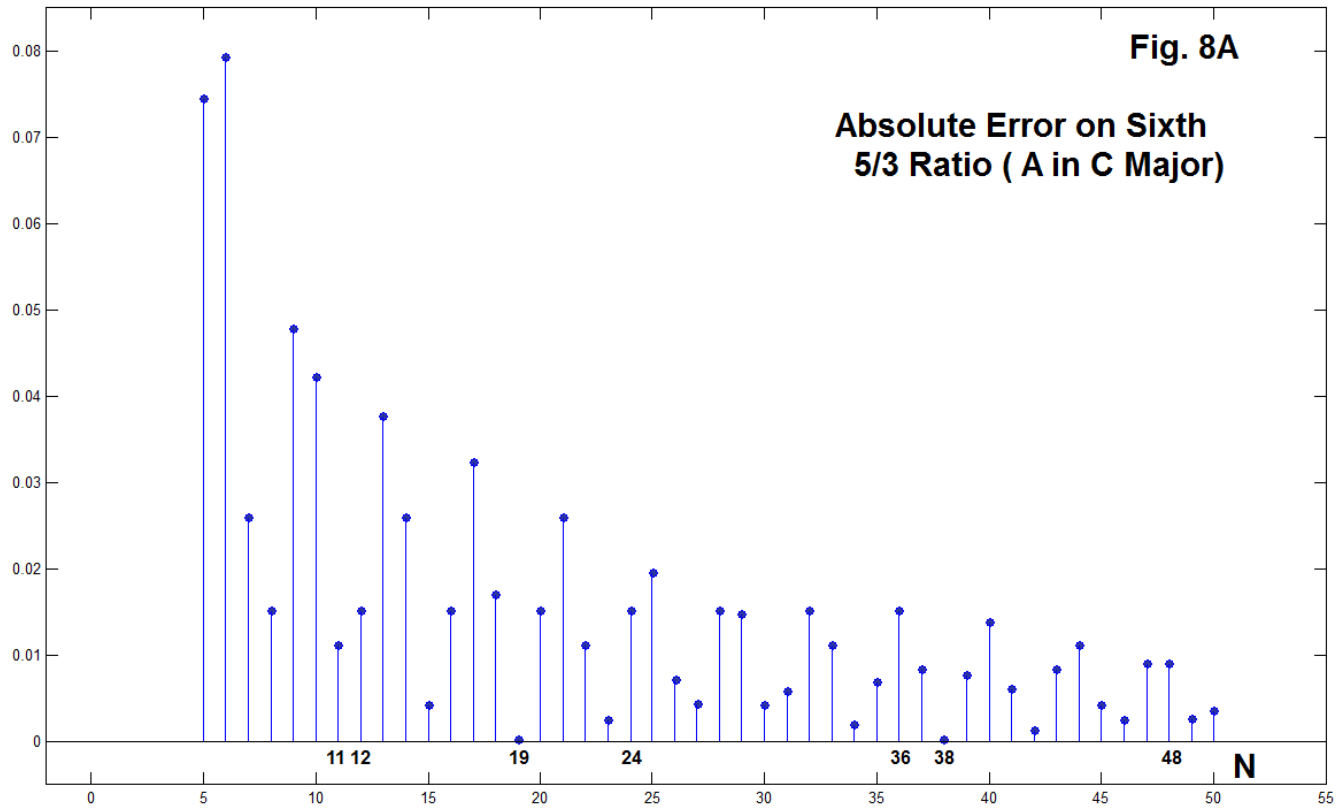


Fig. 8A
Absolute Error on Sixth
5/3 Ratio (A in C Major)
N

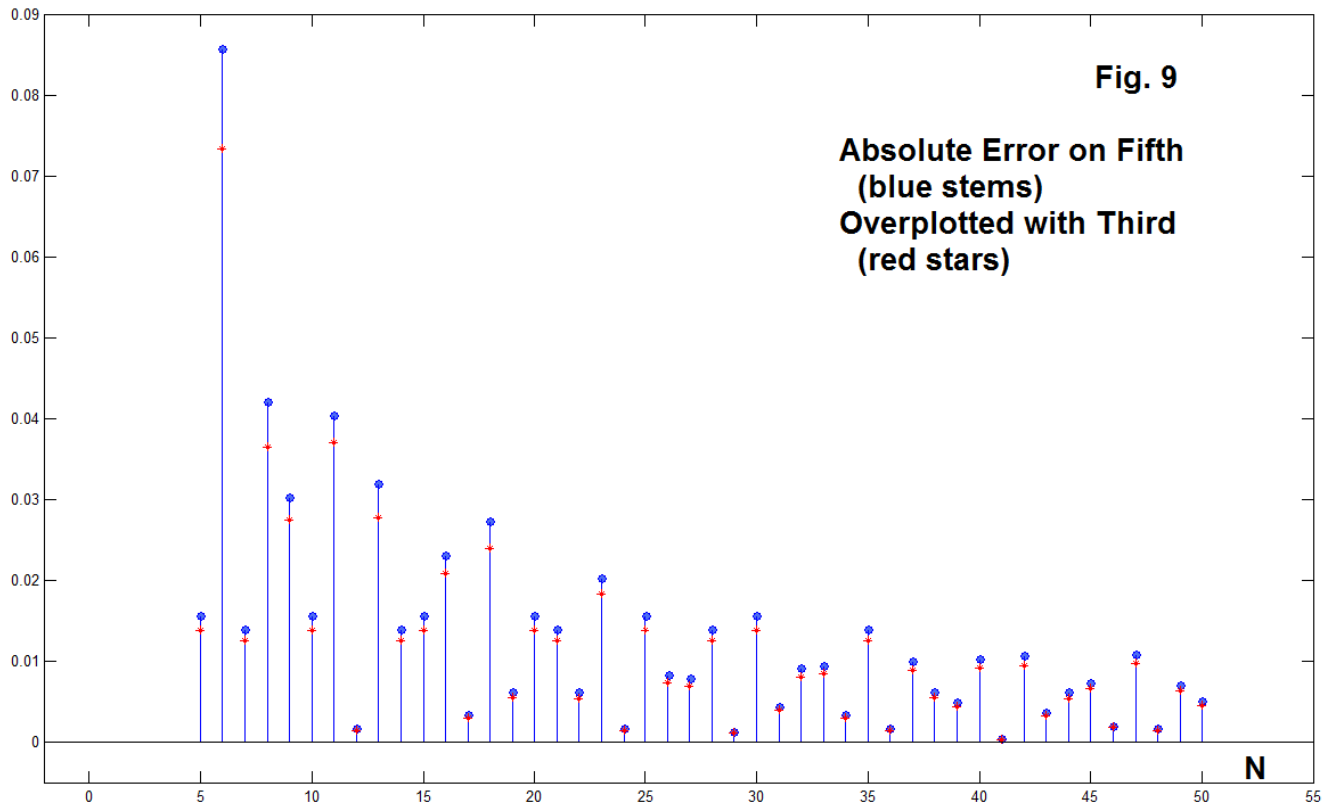
what we can learn – mostly that the error at the individual ratio level jumps around quite a bit. Some scale lengths that we have not heard from yet will show outstanding performance for some particular ratios (we expect this).

With regard to Fig. 8E, the third of the scale, or E on the C-scale, ideally this is the ratio of 5/4 = 1.25. In the 12-tone equal tempered scale, it is $2^{4/12}$ = 1.2599. Clearly this is the result we get from $2^{2/6}$, $2^{3/9}$, $2^{4/12}$, $2^{5/15}$ $2^{6/18}$, and so on. Thus we see the same error for E starting at N=3, and at all multiples of 3. Except at 48! And why do we have the same apparent error at N=29, not a generally recognized multiple of 3! The error at N=29 is almost exactly the same as for N=12, etc., by accident. The choices are for 1.2400 and 1.2700, so 1.24 is the closest to 1.25. But the error is essentially the same magnitude, but of different sign, as the recurring 1.2599. Note: The fit at N=28 is an outstanding $2^{9/28}$ = 1.2495673.

What is going on at N=48, where the error is less, is that at this point, the 1.2599 choice is no longer the best. The best choice is 1.2419 ($2^{15/48}$) that has become available and is better than $2^{16/48}$. Actually, going back to N=45, the 1.2599 error was already abandoned (in favor of $2^{14/45}$ = 1.2407) but we can't as easily tell that from the plot. For N=42, $2^{13/42}$ = 1.2393, so this is the largest N such that the 1.2599 is still in charge. This case, where a particular solution for smaller N is replaced by a smaller error solution for a larger N is of course expected. We give the details here because the ease with which the computer program does this is to be appreciated.

The equal-tempered scale for 12 tones is well known to have a sharp major third, and this is reflected in the error. The fit to the sixth (Fig. 8A) is of about the same as to the third for N=12, but the fits to the fourth (Fig. 8F) and the fifth (Fig. 8G) are much better. For the fourth (Fig. 8F) we were looking for 1.3333…. and the N=12 value is 1.3348. This choice persists at N=24, 36, and 48. Note that N=29 is somewhat lower ($2^{12/29}$ = 1.3322) and N=41 seems outstanding ($2^{17/41}$ = 1.3330). For lower values of N, note that N=12 is better that N=19, and here N=17 is also better than N=19. All these results we think of as pretty much "accidental".

From Fig. 8G, the fifth, we see a graph that is very very similar to Fig. 8F for the fourth. In fact, in examining this to comment on these four graphs of Fig. 8, I was convinced I had accidently repeated one, and looked more carefully to see which one I had used! But they are not identical (see Fig. 9) - the similarity seems to be the result of their being symmetrically placed about the diminished fifth. That is, one (F) is a half tone below $F^{\#}$, and the other (G) is a half tone above $F^{\#}$. In the equal tempered scale, $F^{\#}$ is $2^{6/12}$ or the square root of two, the exact midpoint of the scale. So this is <u>certainly true for the equal tempered scale</u>, and the geometric mean is: $\sqrt{2^{5/12}2^{7/12}} = \sqrt{2}$. But <u>could it be true if we put in the small integer ratios</u> instead of the equal-tempered values? <u>Well</u>, <u>it is</u>: $\sqrt{\left(\frac{4}{3}\right)\left(\frac{3}{2}\right)} = \sqrt{2}$. A bit of an unexpected result.



Accordingly, there is not much new in Fig. 8G not mentioned in Fig. 8F. This leaves us to consider Fig. 8A, the sixth. Note that this result does not have the same sort of symmetry we had with the fourth and the fifth as $\sqrt{\left(\frac{5}{4}\right)\left(\frac{3}{5}\right)} = \sqrt{3}/2$. We do see N=11 beating N=12, and an outstanding result for N=19 since $2^{14/19}$ = 1.6665.

OTHER POSSIBLE CAUSES ?

As I have said above, the small integer ratio as a guiding mechanism for a 12 tone chromatic scales (as supported into antiquity by the circle of fifths) has dominated my thinking with regard to scales. So "Universal Music" [1] from 1974, which predated even the ***Musical Engineer's Handbook*** (1975), which gave a similar presentation [5], and the error minimization exercises (at least from about 1992) [6] tell me what I need to know. While we probably don't need an excuse to "revisit" as issue after periods for 20 – 37 years (!) there was a secondary reason for writing this up now.

Back in December 2007, when the January 2008 issue of ***Discover*** magazine arrived, I noted an item [7] which reported on the findings of Dale Purves and his group in Neuroscience at Duke. One does not expect a popularized science magazine to give many details, particularly in a short item, or necessarily even to get the general ideas right, but they serve us by providing leads. Here they said "Musical Scales Mimic the Sound of Language" and this was interesting, although I assumed they got it backward. Although reported correctly, the reason it is backward in the actual research (or just wrong) is that, as is often the case, cause and effect are confused. See Appendix A for details.

The ear, the brain, psychoacoustics, and the physics of musical instruments (including the human voice which is of course primarily for human speech) constrain what is possible, and guide what actually evolves, at least long term. This "basis" drives both speech and music, which most likely co-evolved, toward different ends. Certainly the human vocal apparatus was the first human musical instrument, but for thousands of years (Pythagoras studied plucked strings about 600 BC) humans have produced mechanical musical instruments (plucked and bowed strings, wind-driven pipes, drum membranes) which advanced musical traditions and cultures. These could be produced and replicated with some precision.

Voiced speech is produced by vocal-cord-regulated pulses of air, with their harmonic structure, subsequently filtered by an acoustical filter (the vocal tract) which has certain (relatively weak) resonances. Human beings – specifically (but far from exclusively) their vocal tracts – are not made and replicated with precision. Formants are all over the place varying with gender, overall size of individual humans, and almost with abandon. There are really only two measurable formants in any voiced sound (principally vowels) and not only are the frequencies of these widely spread, but the ratios varies widely. A plot of f_1 against f_2 (the two lowest formants) in a plane shows large, often partly overlapping, "balloons" ("Paterson/ Barney Vowel Charts") corresponding to certain vowel sounds. If we want to say we have 12 tones in a scale, as opposed to perhaps 11 or 13, then we have to know our confirming data to 1 part in 12, and we do not. Formants are simply not well-defined. Well, perhaps not so bad considering that we use much of the same "apparatus" for breathing and for eating, and we borrow these for making sounds. The point is that any suggestion that measured positions for formants are reliable (even for vowels) is very unlikely.

[As an aside, we note that formants are not necessary for language (obviously, you are reading this) or even for speech. Speech progresses at a tiny bit rate (perhaps 50 bits/second) and is massively redundant in all respects. As basically a fascinating aside, to illustrate that formants are not necessary, even acoustically, I mention James Gleick's wonderful book ***The Information*** [8] which describes "talking drums". These are the canonic two-tone drums which transmit messages through a dense jungle. Some sort of code? No – they talk. The user's language is tonal and uses two tones, the same as the drums. Words they use for ordinary speech require the proper tonal sequence (pitch) within the word in

addition to vowels (formants) and consonants. The drums only transmit the tones! Is this all that is required? Certainly not, the tonal sequence could correspond to any number of words, and giving just the tones would be highly ambiguous ”utterance”. Context helps a bit – if you have a pretty good idea what is about to be said, you hear it much better. But with the drums, it is redundancy in the sense of a very round-about (and much longer) set of possibilities, some standard, some invented on the spot, and it is for the human brain to juggles the possibilities for what must make sense. I find this amazing. Especially as we would generally ascribe drum communications to “primitive people”. Too good not to share?]

So, with ***Discover*** acting to provide leads, I could find the actual paper, Ross *et al* [9]. I do not really understand this paper. In some sense, it is reminiscent of the circle-of-fifths back-casting its frequencies into a single octave. Here it is casting formants to an octave. In approaching a new paper, it is bad to start with certain expectations (but try not to do this – impossible) but this paper is just not convincing. No signal there!

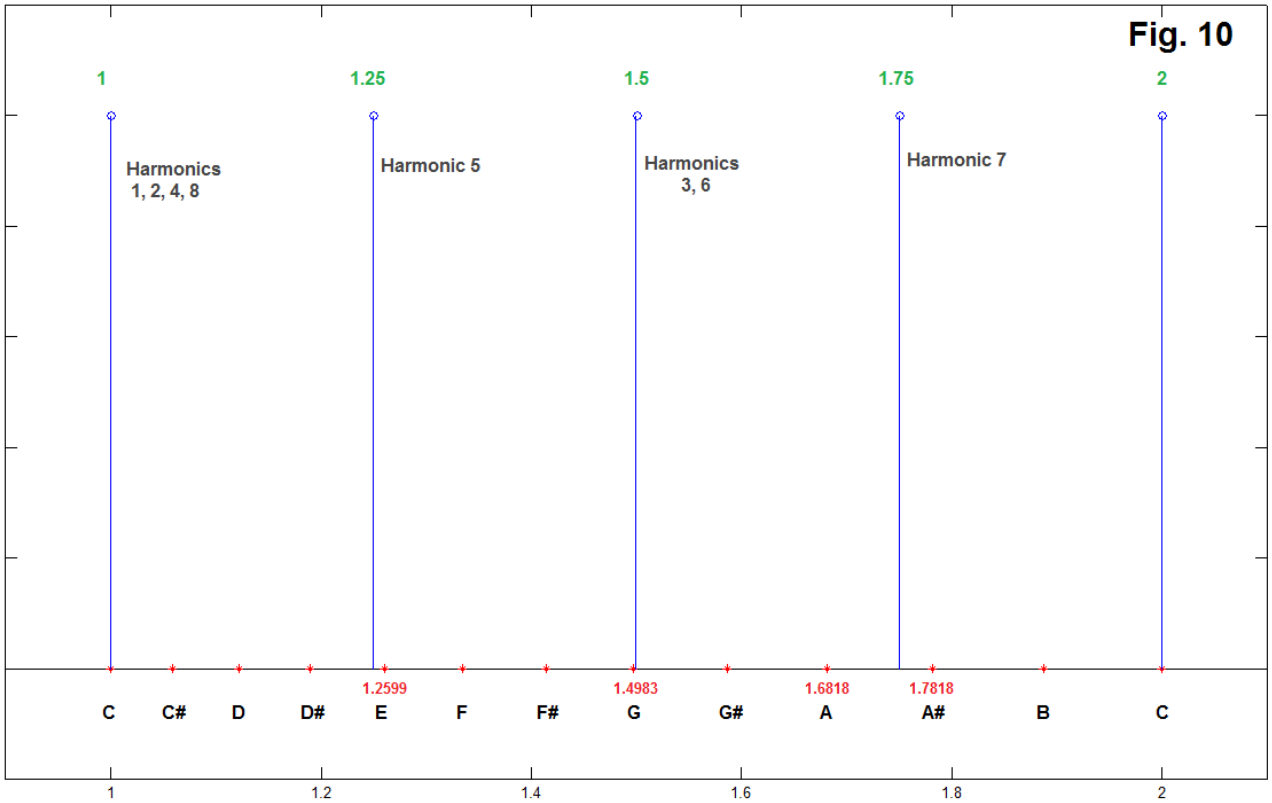
So about Dec. 11, 2007 I emailed Pervus to ask if he was aware of the idea of just fitting the fundamental low-integer ratios and thereafter filling in obvious gaps. He replied quickly that he was not, and I replied with attachments of [1] and [6], to which he replied with thanks, and the suggestion that they took it seriously and would look it over shortly. This would not be the first time, by any means, that I had noticed neurobiologists and perceptual psychologists missing something quite apparent that engineer’s just kind of had as part of their regular tool-kit. I heard nothing further.

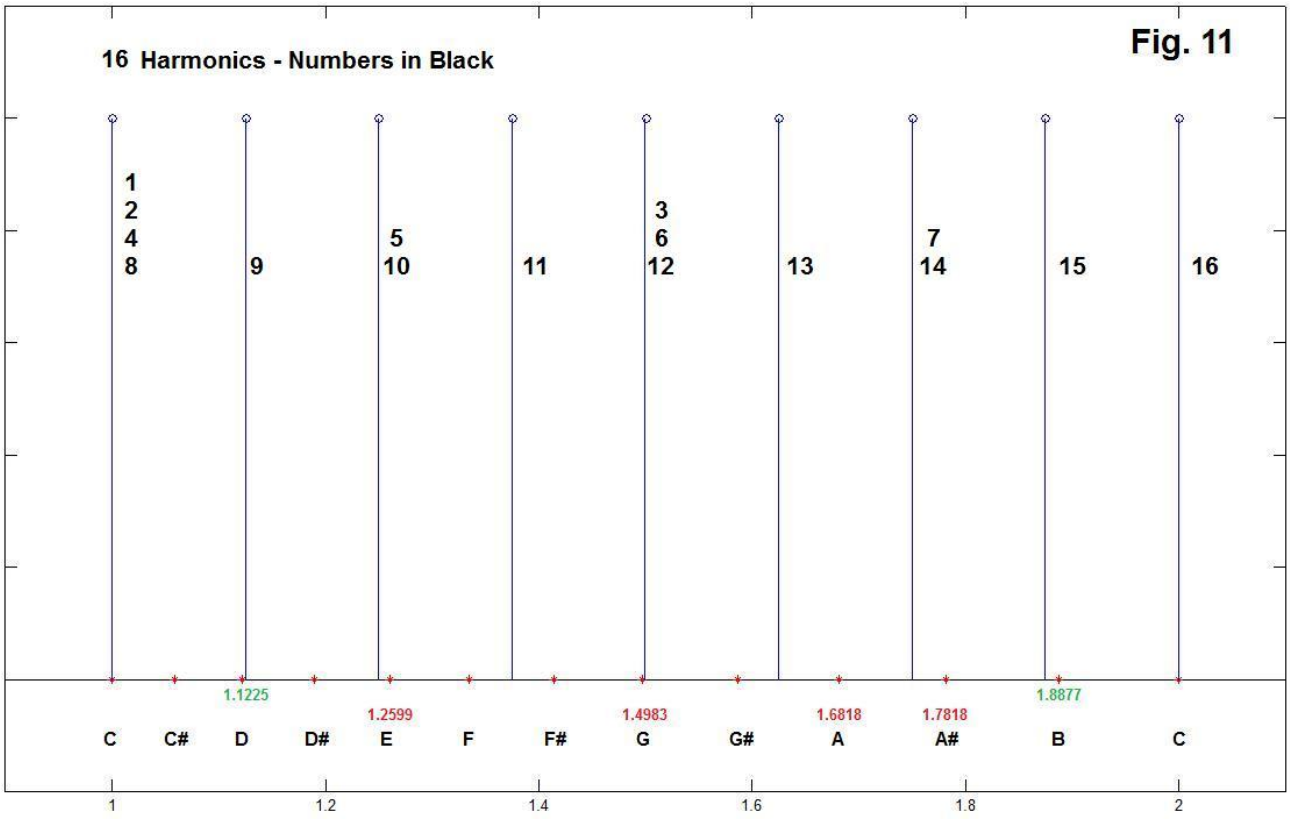
Subsequently however, the same lab at Duke produced a paper of which I have just become aware, Gill and Purves [10]. This paper is well-presented and has interesting information about a variety of scales. Here they are concerned with selecting scales that resemble harmonic sequences, and not speech formants (although they do allude to speech as a source of the <u>harmonic training</u>). The <u>exact criterion used here</u> to determine similarity to a harmonic series may be new, but the basic idea of selecting scale tones from a harmonic series is known and known to be limited to perhaps at most 6 or 7 harmonics, yielding only a fifth and a major third. One problem is that as we go to higher and higher octaves, the number of harmonics within that octave double, while the number of scale tones remains the same. If you look to high enough octaves, you will eventually find harmonics in any position you desire.

Here is an example: Let’s consider a series of harmonics from 1 to 8. Thus the harmonic numbers are 1,2,…8 and we will simply take the frequencies to be these same integers. So for the frequency range 1 to 2, the first octave, we have only the fundamental. We will of course take the frequencies that are multiples of 2 to belong to the endpoints of the first octave. Thus the second octave has harmonics (and frequencies) 2 and 3. The third octave has 4, 5, 6, and 7. The fourth octave begins with 8.

Next we divide the second octave by 2, the third octave by 4, and the fourth octave by 8. This puts all frequencies back into the first octave. Finally, we arrange the results in ascending order (there are repeats of course). The table just below shows the results:

Harmonic Number	Frequency	Back-Casted to First Octave	Sorted	Possible Scale Tone
1	1.0000	1.0000	1.0000	C
2	2.0000	2.0000	1.2500	E
3	3.0000	1.5000	1.5000	G
4	4.0000	2.0000	1.5000	G
5	5.0000	1.2500	1.7500	?
6	6.0000	1.5000	2.0000	C
7	7.0000	1.7500	2.0000	C
8	8.0000	2.0000	2.0000	C




16 Harmonics - Numbers in Black
Fig. 11
1
2
4
8
9
5
10
11
3
6
12
13
7
14
15
16
1.1225
1.2599
1.4983
1.6818
1.7818
1.8877
C
C#
D
D#
E
F
F#
G
G#
A
A#
B
C
1
1.2
1.4
1.6
1.8
2

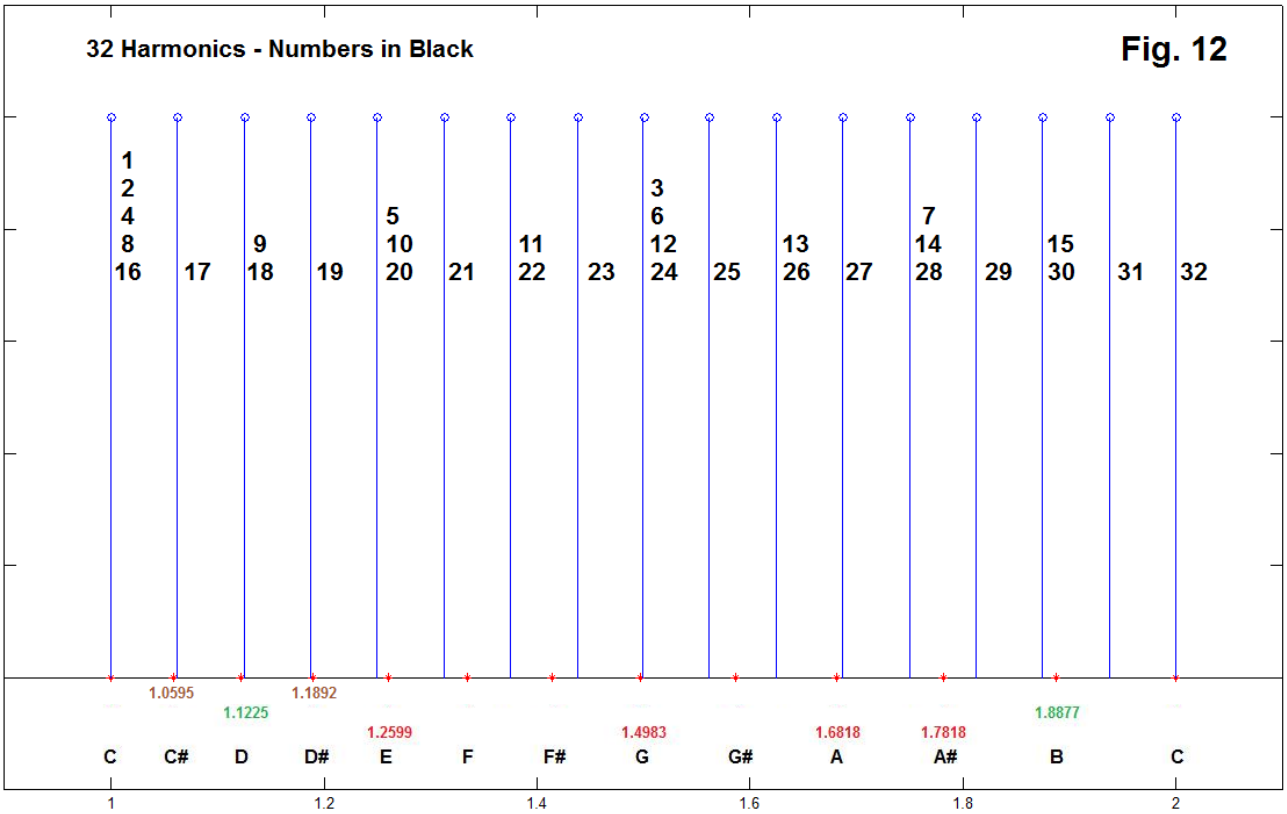

32 Harmonics - Numbers in Black
Fig. 12
1
2
4
8
16
17
9
18
19
5
10
20
21
11
22
23
3
6
12
24
25
13
26
27
7
14
28
29
15
30
31
32
1.0595
1.1892
1.1225
1.2599
1.4983
1.6818
1.7818
1.8877
C
C#
D
D#
E
F
F#
G
G#
A
A#
B
C
1
1.2
1.4
1.6
1.8
2

Fig. 10 does show evidence for a C major chord (C-E-G), but not much for a scale. In particular the 7th harmonic (back-casting to 1.75) is not contributing a tone. Fig.10 plots these results against the 12-tone equal-tempered scale positions (red stars). Fig. 11 shows a similar result where we use 16 harmonics. Note that this looks very promising for the lowest tones (C, C#, D, and D#), but fails completely for the upper portion. Fig. 12 shows 32 harmonics, making the point that you can eventually get better and better approximations if you use enough harmonics (although most are not used at all). And of course, a 32nd harmonic of something like A=440 Hz is 14080 Hz and gets near the top of the hearing range.

The use of a harmonic series to argue for a particular scale is problematic. The argument that "conspecific vocalizations" (human speech) trains us to favor harmonic relationships puts the argument further afield. In the new paper [10], they do not mention the earlier paper [9]. And while they do (now) mention von Hoerner – they dismiss the ideas:

> *" A third approach has used error minimization algorithms to predict scale structures under the assumption of competing preferences for small integer ratios and equal intervals between successive scale tones [15,16]. This method can account for the structure of the equal-tempered 12-tone chromatic scale but cannot account for any of the five to seven-tone scales commonly used to make music. Moreover, no basis was provided for the underlying assumptions."*

(The two references, 15 and 16, are exactly [2] and [3] here.) I have to disagree with everything said after "but" in the quote above.

First of all, the 12 tone determination by the error minimization gives the possible <u>choices of 12</u> from which we may then chose <u>subsets</u> for smaller sized scales. And, we cannot justify too much precision with regard to choosing available tones, or claiming that tones are significantly and practically different. Charles Taylor summed it up very very well [11]. Indeed, in forming a major scale we had no trouble (above) to choose 5 tones from among 12 on the basis of low integer ratios, and the additional 2 for "filling up holes". As for total tones, 7, a relatively small number, as compared to all 12 or perhaps 19 or 31 etc., we have only to put on our engineer's hat and say "FINITE RESOURCES". This relates as well to the comment in the quote about "underlying assumptions".

So if we agree that we have 12 tones reasonably well selected, starting from the overall goal of approximating certain low integer ratios, what else is necessary to select subgroups of these for various scales of perhaps 7 notes, 5 notes, etc.? One thing would be to look at

possible combinations. Finding the number of combinations of N things (N possible tones) taken r (r tones in scale) at a time is the classic formula:

$$C = \frac{N!}{(N-r)!\,r!}$$

which is 792 for N=12 and r=7 (or for r=5 for that matter). That's a lot of scales, mathematically, even before we consider what we need and an afford. I do not know how to reduce this number properly here. One thing is that we would certainly not consider starting on a different note to be a different scale. And we don't get a different scale if we overrun the octave. Perhaps more importantly, we expect a relatively uniform distribution. Scales in common use (major, minor, the various "modes" and pentatonic scales) tend to be constructed of only full steps and half steps (very occasionally 1.5 steps) with no clustering of half steps beyond two. A scale consisting of tones taken serially from 1 to 7 of all 12 is not used in practice. So the number is perhaps 20 different scales (?), most of which are very seldom used. And the reductions suggested just above? Well it is almost certainly a matter of esthetics much most likely than of neuroscience. It is a matter of tradition, and as suggested above, of resources.

REFERENCES:

[1] von Hoerner, S., "Universal Music", ***Electronotes***, Vol. 5, No. 35, Feb. 15, 1974, pp 2-4, posted at link below:

http://electronotes.netfirms.com/UniversalMusic.PDF

[2] von Hoerner, S., "Universal Music?", ***Psychology of Music***, Vol. 2, No. 2 (October 1974) pp 18-28.

[3] von Hoerner, S., "The Definition of Major Scales for Chromatic Scales of 12, 19, and 31 Divisions Per Octave." ***Psychology of Music***, Vol. 4, No. 1, (April 1976), pp 12-23

[4] von Hoerner, S., "What is Music All About?", ***The Observer***, Vol. 15, No. 1, Feb. 1974

http://www.nrao.edu/archives/NRAO/observerV15N1.pdf

(This seems to be a newsletter of the Green Banks Observatory. A very good easy to read description by von Hoerner. You may be distracted by all the other interesting stuff there from vegetable gardening to absolute pitch in small children.)

[5] Hutchins, B., “Musical Scales” in ***Musical Engineer’s Handbook*** (1975) pp 1e(3)-1e(5), posted at:

http://electronotes.netfirms.com/Scales-MEH.PDF

[6] Hutchins, B., “ A Least-Square and Min-Max Error Estimation of the Musical Scale Question,” ***Electronotes***, Vol. 17, No. 179, June 1992

[7] S. Kruglinski, “Musical Scales Mimic the Sound of Language” from ***Discover***, January 2008, a short description available at:

http://discovermagazine.com/2008/jan/musical-scales-mimic-sound-of-language#.UL6zY9na8dA

[8] Gleick, J. “Drums That Talk (When a Code Is Not a Code)” Chapter 1 of ***The Information***, Pantheon (2011).

[9] Ross, D., & J. Choi, D. Purves, “Musical Intervals in Speech” , ***PANS***, Vol. 104, No. 23, June 2007

http://www.pnas.org/content/104/23/9852.full

[10] Gill, K.Z. and D. Purves “A Biological Rationale for Musical Scales” ***PLoS ONE*** 4: e8144. doi:10.1371/ journal.pone.0008144 (2009)

http://www.purveslab.net/publications/gill_purves_2009.pdf

[11] Taylor, C., Exploring Music – The Science and Technology of Tones and Tunes, Inst. of Physics (1992). Taylor wrote:

Some of the differences are very small indeed and, while I understand the theoretical reasons for the choice of some of them, I find it difficult to believe that anyone but the most highly trained musicians could really distinguish between some of them. I suspect too that all tuners would not be capable of setting particular instruments in a particular scale without the aid of electronics. How it was done in the time of Helmholtz I find it difficult to imagine.

The first point to get clear is that music comes first and scales later. Scales can be compared with the grammar of a language; it is perfectly possible to speak a language for the whole of one's life without ever understanding its grammar. And, in the same way, musicians, especially those in the folk tradition, can write, play and sing musical compositions without being consciously aware of the scale structure involved.

[12] http://electronotes.netfirms.com/ScalePrograms.txt

* * * * * * * * *

APPENDIX A: COMMENTS ON ROSS *ET AL* [9]

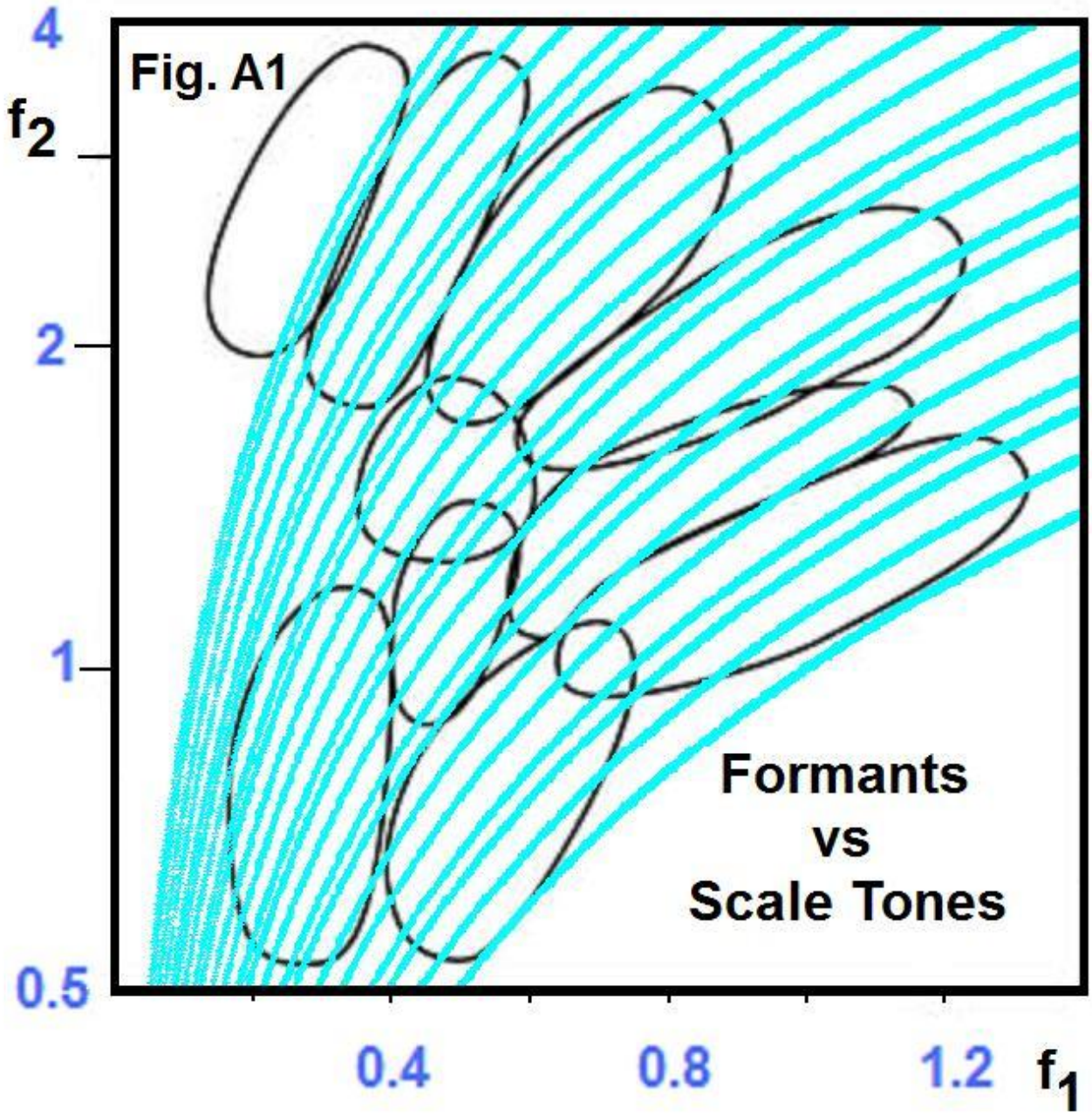
The two main claims of Ross *et al* [9] are that (1) there is no existing understanding of why a 12 tone chromatic scale is in general use and (2) that the spacing of formants in human speech provides such an explanation. From the overall discussion above, it should be clear that good satisfactory arguments for the origins of a 12-tone musical scale are long-standing and relatively abundant. In this appendix we show why vocal formant placement as an explanation for musical scales is unlikely.

(a) CAUSE AND EFFECT

It is not inconceivable that vocal formants could occur at scale tones. It is perhaps even likely. Both have a common basis in psychoacoustics, etc., and the same hearing mechanism could well have a preference for small frequency ratios in both cases. This would mean both are the effects of a common cause, not one the cause of the other, in either order. Considering that singing almost certainly represents the origins of human music, the *de facto* emergence of scales might have been a very early occurrence. Today we do generally sing or are otherwise accompanied by harmonies (chords) that are scale tones. Formants that are scale tones might be serving like actual chords in a rudimentary way. This is speculation only.

(b) FORMANTS ARE TOO POORLY DEFINED TO DETERMINE SCALE TONES

Much is known about speech formants. The work of Petersen and Barney is classic [see Peterson, G. and H. Barney, "Control Methods Used in a Study of the Vowels," ***JASA***, Vol. 24, No. 2, pp 175-184 (Mar. 1952)]. Fig. A1 below shows a version of their "balloon" graph which is overplotted here with blue curves representing the frequency ratios of scale tones over three full octaves. (Note that these are straight lines, constants ratios of f_2 to f_1, radially outward, appearing as curves because of the log scale on the y-axis.) The curve that is most lower right is a 1:1 ratio. The next curve above it is a 9/8 ratio, and so on for three octaves (8:1) of a major scale. Under the curves we have 10 vowels represented by the range of formant ratios of f_1 to f_2. We have removed the indicators of the actual vowels to avoid clutter – they are not essential here. These balloon regions were obtained by Peterson and Barney by measuring formants for a range of speakers for the particular vowels. Note that there is considerable variability in both the f_1 and f_2 directions, with some overlap. Typically a chosen vowel overlaps 5 to 10 of the scale tone curves. The variations in frequency are fully expected. The essential point here is the fact that the balloons underlie a wide range of ratios.



It is worth noting that the scale ratio curves do have a trend that is elongated in the same direction (generally considered) as the formant balloons. This is almost certainly due to the fact that here we have acoustic filters (a vocal tract) that varies in length from individual to individual, but less so in "pipe scale" as it is called (longer <u>and</u> wider). The Peterson/Barney work clearly associates formant ratio with particular vowels, although as is evident from Fig. A1, the plane is largely populated and boundaries between different vowels are not absolute. Vowels in English (perhaps, in particular) have been described as existing in a "cloud".

There is no suggestion that I can find in [9] that suggests a relationship between a particular formant and a particular scale tone. Indeed, there likely should not be, except as

we note the similarity of the number of vowels in this case (10) to the number of scale tones (12) and that both are associated with particular ratios, or range of ratios, of two frequencies. So instead of an association of vowels with scale tones, perhaps a larger regard needs to be given to the apparent commonality of the phenomena of scales and the phenomenon of vowels with some overriding aspect of perception (like resolution).

(c) IS IT A NUMERICAL ARTIFACT?

In working with students for so many years, and even spending mandated portions of some design courses on "engineering communications" (issues such as report writing), I have encouraged the retention of elements of a "thinking process" in a written presentation (the report). We of course all tend to neaten up our work and forget the pitfalls we encountered along the way, forgetting that the reader following us, or better still, thinking ahead, is likely to fall into the same. I am bringing this up here to bring forward the point that the distribution of small integer ratios presented in Fig. 4 was done as a curiosity, on the spot, and not with anticipation of how it might relate to a careful study of [9]. Indeed, it was only as I went back later for a careful study of [9] that a noticed the striking similarity of my Fig. 4 to the graphs in [9], for example the one at the link:

http://www.pnas.org/content/104/23/9852/F3.large.jpg

Let's call this online figure Fig. A2, although not printed here. In particular, in both cases, note the concentration of results at the lowest integer ratios (the just tones), and the regions surrounding them on either side where competing choices are "repelled". In the case of Fig. 4, we just chose all integers within ranges (and did not stack up resulting repeats). Looking at the "Methods" of [9], we find the following:

> *"For both the word and monologue data, the nearest harmonic peak to the underlying formant maximum given by Praat was used as an index of the formants: the formant value assigned by rlinear (sic) predictive coding was divided by the fundamental frequency, and the result was rounded to the nearest integer. The ratios of the indices of the first two formants were then calculated as* B/A *where* B *= the formant 2 harmonic index and* A *= formant 1 harmonic index [the data were plotted as* $\log_2$(B/A)*, as is conventional]. Ratios were counted as chromatic if they corresponded to just intonation values for the chromatic scale (see* Discussion) "

The emphasis in red was added.

So what is the range of the formants, and the range of the fundamental. The male voice in the most general terms has a fundamental of say 100 Hz to 150 Hz. As seen in Fig. A1 the first formant runs from 200 Hz to 1400 Hz, and the second from 500 Hz to 4000 Hz. Typically we expect relatively few harmonics (integer multiples of the fundamental) in any formant (perhaps 3 to 5 harmonics). The formant itself is the frequency response of a filter, and formant filters are typically not sharply peaked bandpass features. If we determine the formants from vocalization, we need to "deconvolve" the excitation (pulses of air through the vocal cords) from the filter. This is not easy. In general, we would instead look for spectral peaks to show the frequency response. This relies on seeing the response for the frequencies present (only the individual harmonics). There is not the slightest reason to believe that the peak of a formant's response should be at a harmonic actually present. Accordingly, if the formant is "rounded" there can be considerable uncertainty introduced. This also accounts for the isolation of small integer ratios against an unpopulated background.

So here it appears that the methodology quantizes the formant frequencies to integer multiples of a fundamental (the harmonics). [In a Dec. 10, 2012 email from Purves, he stated that the integer rounding was a problem and that his group had de-emphasized the use of formant positions.] By necessity this rounding forces formant ratios into integer ratios, and generally smaller integers. From that point on, the "universal music" situation is automatic.

APPENDIX B: TOOLS AND DEVICES

Evolution has so often produced mechanisms and relationships so precise and apparently clever that we easily forget that there was no design or intention involved. We may be led down a false path looking for causal relationships when all we really have is contingency and subsequent adaptation. Take for comparison a wrench! Does anyone not suppose that the wrench (the adjustable open-end – Fig. B1) coevolved with nuts and bolts. The wrench involves of course a lever, a screw mechanism, parallel surfaces, and a 30° offset of the planes relative to the handle. A similar earlier device was the "monkey wrench" which was harder to use, but was mostly in the era of square nuts. With the advent of hexagonal nuts and the open end with the 30° offset, it became possible to tighten the nut with as little as 30° of twist of the wrench, by alternately reversing it.

Well, there are lots of variations on this, and many similar clever man-made devices, but we do often notice that there exist very real limits as to when things are getting better, and when they are not, or even if they could change substantially at all. For example, Fig. B1 does show booth square (4 sides) and hex nuts (6 sides), both of which are familiar and usually interchangeable (as I recently observed in putting a muffler back on my chain saw.

Fig. B1 also suggests five unconventional nut shapes, 3, 5, 7, and 12-sided, as well as an irregular shape. We likely don't use these. But if the hex-nut/30°-offset was a good idea, why not, even better, a 12-sided nut and a 15° offset. Well, you protest, that is almost certainly going to slip and shear off the ridges. Six sides is enough. In many cases, four sides was enough. Not only do we not need more, but performance is worse with more. And our needs are flexible. The nut which fell (forever lost) from my saw was hex but all I had (or needed) was a square one. Sometimes things are just "out there".

Is it perhaps the case that scales of 5 or 7 tones are like nuts of 4 or 6 sides? We can't get by with fewer tones very well, but very many more are just too many (rounding off the rims). And some intermediate numbers <u>just don't work</u> at all (like a five-sided nut). We have the tool, the ear/brain processing sound, largely from our deep evolutionary ancestors. The (perhaps) uniquely human notions of speech (one form of language) and music were informed by this in-place machinery that <u>worked naturally</u> (by accident?) with certain formant divisions or frequency divisions (scales) but not others. Interesting, but perhaps too often over-thought.

REVISITING: BEATING

-by Bernie Hutchins

INTRODUCTION:

The audio phenomenon of "beating" is familiar, even to schoolchildren who learn (as they start to play in a band) that two musical instruments that are "out of tune" with each other make an annoying sound. By the time we get to high-school trigonometry and physics, we associate the sum of two sinusoidal waveforms with an amplitude variation. In trig it is the equation:

$$sin(A) + sin(B) = 2\, sin\left(\frac{A+B}{2}\right) \cos\left(\frac{A-B}{2}\right) \qquad (1)$$

and in physics it is just the alternation between constructive and destructive interference. Anything you learned in high-school must be too simple to bother with ever again!

So here are four questions to see if you are really comfortable with beating.

(1) We can parrot back the result that the beating of two sinewaves occurs at a rate equal to the difference frequency. Or is it half the difference frequency – equation (1)? Or is it twice half the difference frequency – two peakings of amplitude for each cycle of $\left(\frac{A-B}{2}\right)$? Factors of two need to be pinned down.

(2) From equation (1) we see four possible frequencies: A, B, (A+B)/2, and (A-B)/2. And isn't it multiplication (not addition) that gives sum and difference frequencies? But then again, equation (1) shows both a sum and a multiply. What is really there?

(3) So far we are talking about a very small difference between A and B. What happens if A and B are not close to each other, <u>but close to a small integer ratio</u>; like A=2 and B=3? Does beating still occur?

(4) What happens in the near-unison case (i.e., A is close to B) if the waveforms we are adding are not sinusoidal (perhaps triangle, square, pulse, sawtooth)? In this case, do we look at things in the frequency domain or the time domain?

Where should we begin? Well, we begin by adding two sine waves of very close frequency together, by equation (1), and by plotting a graph.

TWO CLOSE SINE WAVES

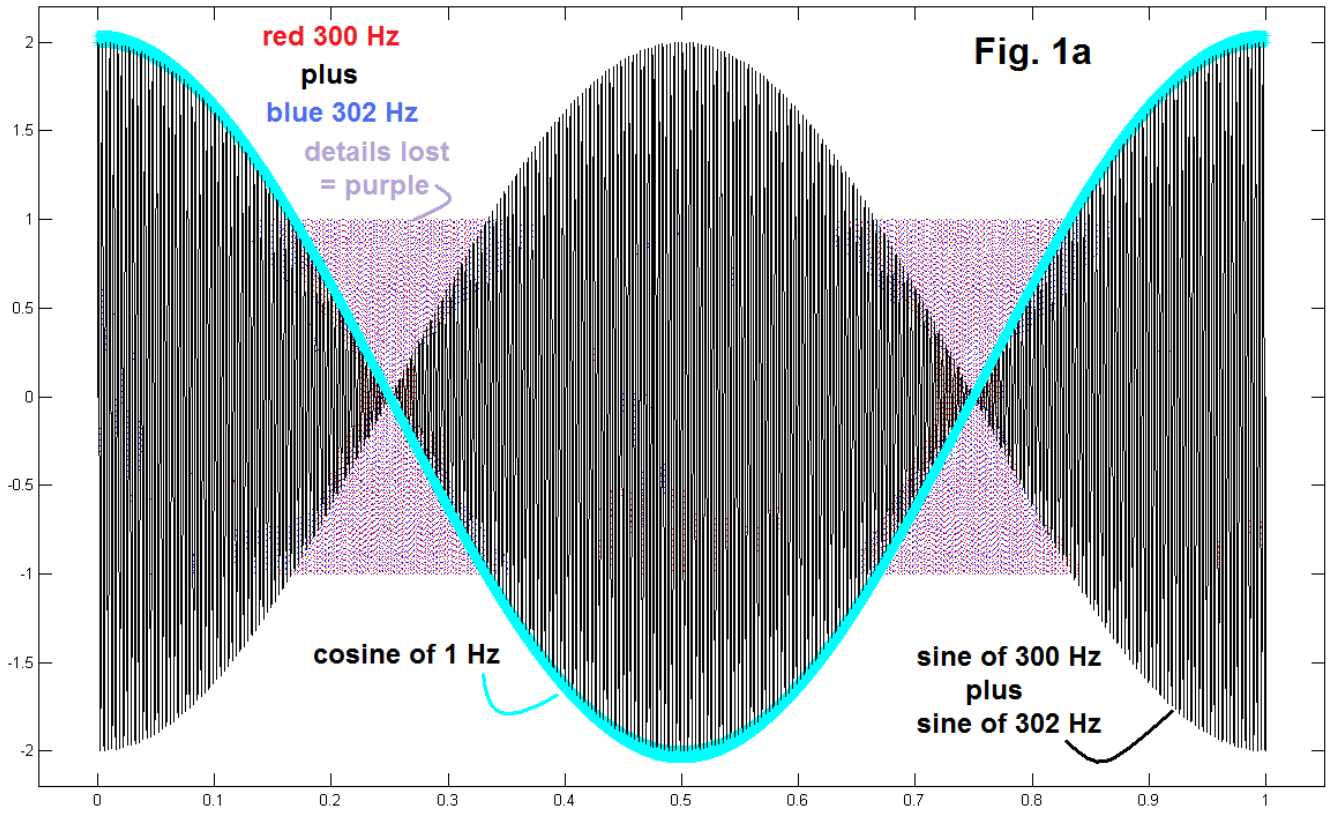
Equation (1) is a familiar trig identity. We most often would see a similar form such as:

$$sin(X) \cdot sin(Y) = (1/2)\,[\,cos(X-Y) - cos(X+Y)] \qquad (2)$$

This equation (and three similar forms of the product side) expresses the "balanced modulator" or "double sideband" result showing that, indeed, multiplying two sinusoidal waveforms together produces sum and difference frequencies. Of course, an equation works both ways, and this is the point of equation (1). The sum of two sine waves is a sinewave of the average frequency multiplied by a sine wave (cosine actually, here) of half the difference frequency. Thinking of the spectra involved, a multiplication results in sum and difference frequencies [equation (2)], <u>or</u> we can say that a spectrum consisting of the sum of two sinusoidal waveforms can be thought of as appropriate balanced modulation [equation (1)]. The form as in equation (2) is likely the most familiar to our readers as the "ring modulator" used in music synthesizers (and in frequency shifters).

Concentrating on equation (1), if we assume that the frequencies corresponding to A and B are very close together, then the $sin\left(\frac{A+B}{2}\right)$ term, the average frequency is very close to being sin(A) or sin(B). If you like, the 2 that begins the equation's right side is just the result of summing two like things. The second term which multiplies (or modulates) this average sinewave (that is very much like either of the originals), is $cos\left(\frac{A-B}{2}\right)$ which has a low number (A-B) which is then even lower, divided by 2. Since cosine is an even function, it does not matter whether A-B is positive or negative, so the order does not matter as we must have in a product. Thus we envision a "normal" type of sine wave modulated by one of quite low frequency. Pretty much the idea of what we see in Fig. 1a, Fig. 1b and Fig. 1c.

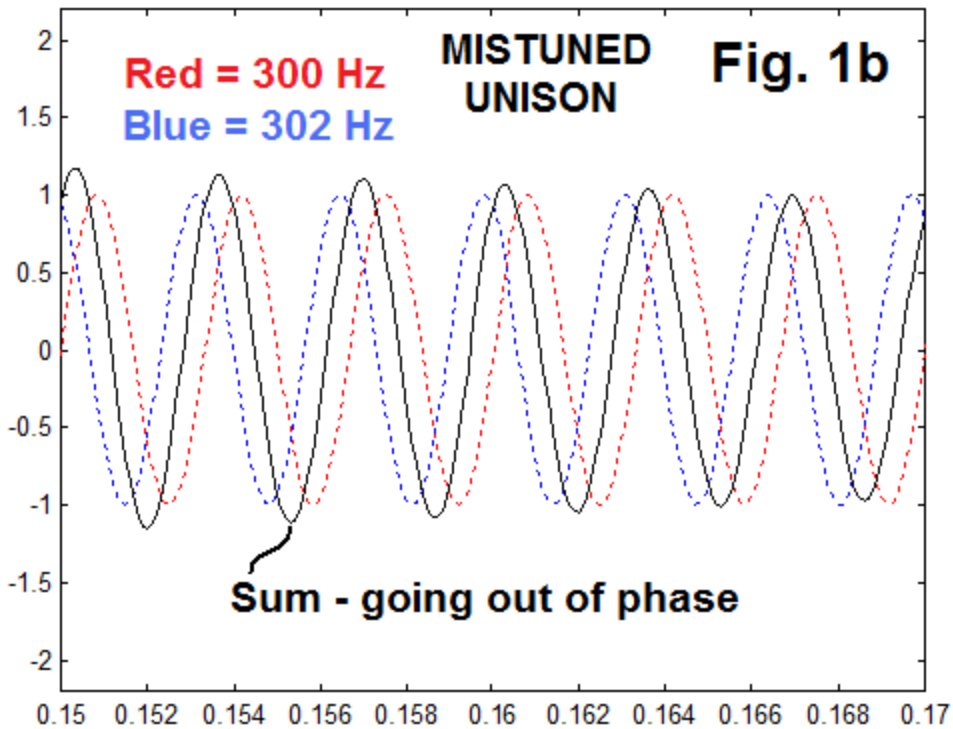
For Fig. 1a, we have chosen two sine waves, one of 300 Hz (red) and the other of 302 Hz (blue), and we display 10,000 samples of both (a full second) so the actual plot is quite purple, and hides behind the black sum for much of the plot. You can't see any of the details, but easily see the two amplitude "lobes" that result. Thus the amplitude envelope of the sum begins high, goes down to zero, back up and down, and then finally up again (and of course repeats on both sides forever). There are two complete amplitude lobes here for the one second, so the amplitude peaks and goes to zero at a rate of 2 Hz, the difference frequency. Yet it is quite true that the modulating waveform is not the difference, but rather half the difference or 1 Hz, the light blue cosine "envelope" that is overplotted. So, we are NOT saying the sum contains the difference frequency (it does NOT). The amplitude beats occur at a rate given by the difference frequency. Tricky.



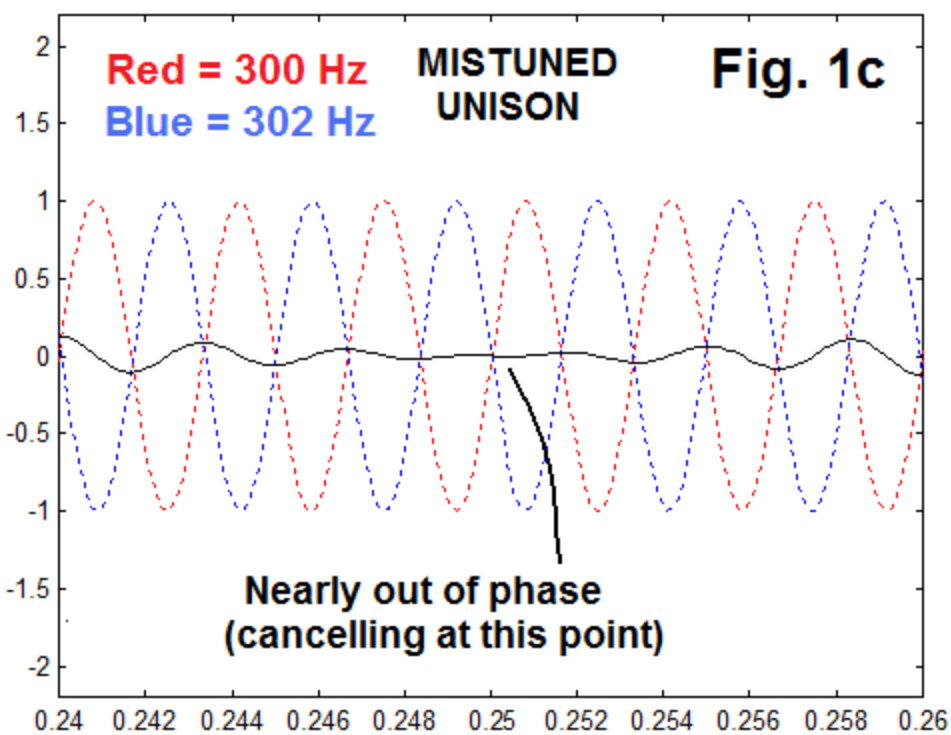
How do we know that the sum does not contain the difference frequency? Because equation (1) says it doesn't! The left-side, a sum which after all defines what a spectrum is, does not contain the difference frequency of 1 Hz. Nor does the equation contain the repetition rate of 2 Hz. Nor does it contain the average frequency of 301 Hz. Clearly if we do a frequency analysis such as a Fourier transform, we get a sum of components that are already there, And equation (1) already does this. That's it. We have 300 Hz, and we have 302 Hz.

Since we can't see the detail of Fig. 1a, two zoom-in figures are also shown. Fig. 1b show the small region around time 0.16 where we see the two sinewaves, which were effectively "in phase" at time 0 going out of phase and starting to cancel. Fig. 1c shows the region around time 0.25, where the two sinewave components actually cancel.

At this point, we can go back and answer questions (1) and (2) without difficulty. Possibly we were not sure of these answers, or had forgotten them. Classic case. In fact, this case can be considered the "control" for questions (3) and (4).



Detail of Fig. 1a shows the two components starting to “drift out of phase” with the sum having decreased from an amplitude of near 2 at time 0 of Fig. 1a to near 1 at this time (centered at 0.16 of Fig. 1a).

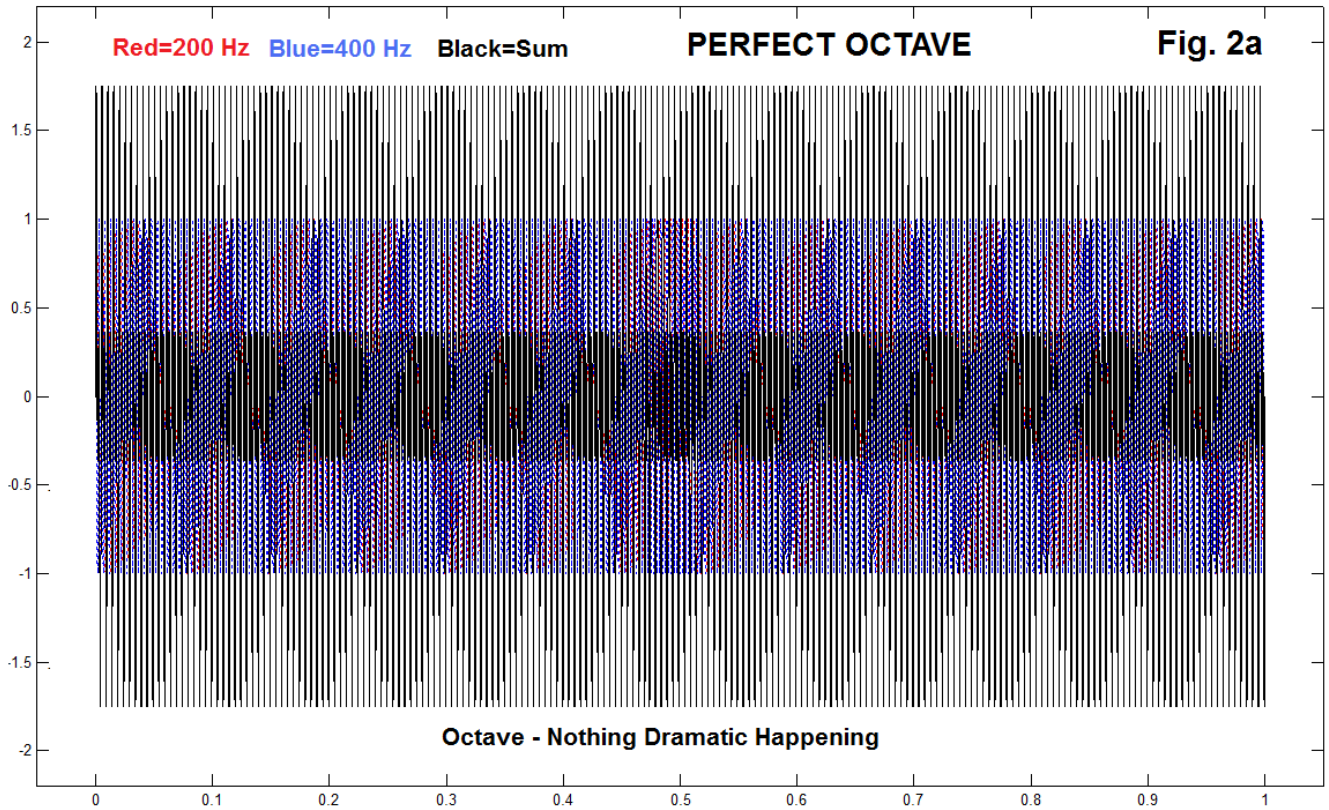


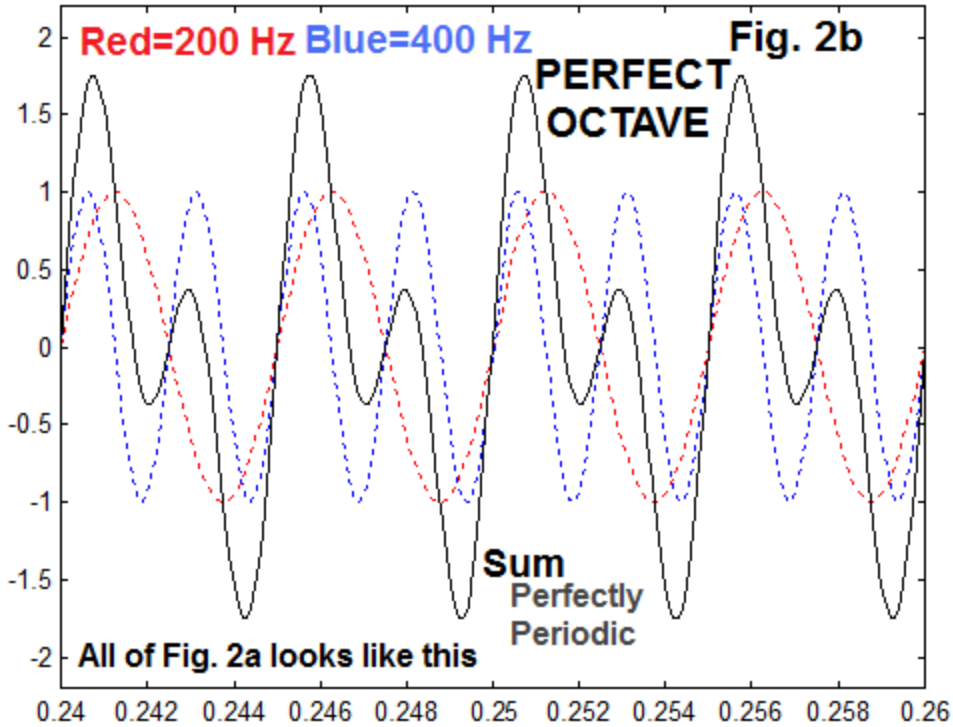
At this point (about time 0.25 of Fig. 1a) the two components are out of phase and cancel, giving the amplitude minimum of Fig. 1a.

BEATING CLOSE TO SMALL INTEGER RATIOS:

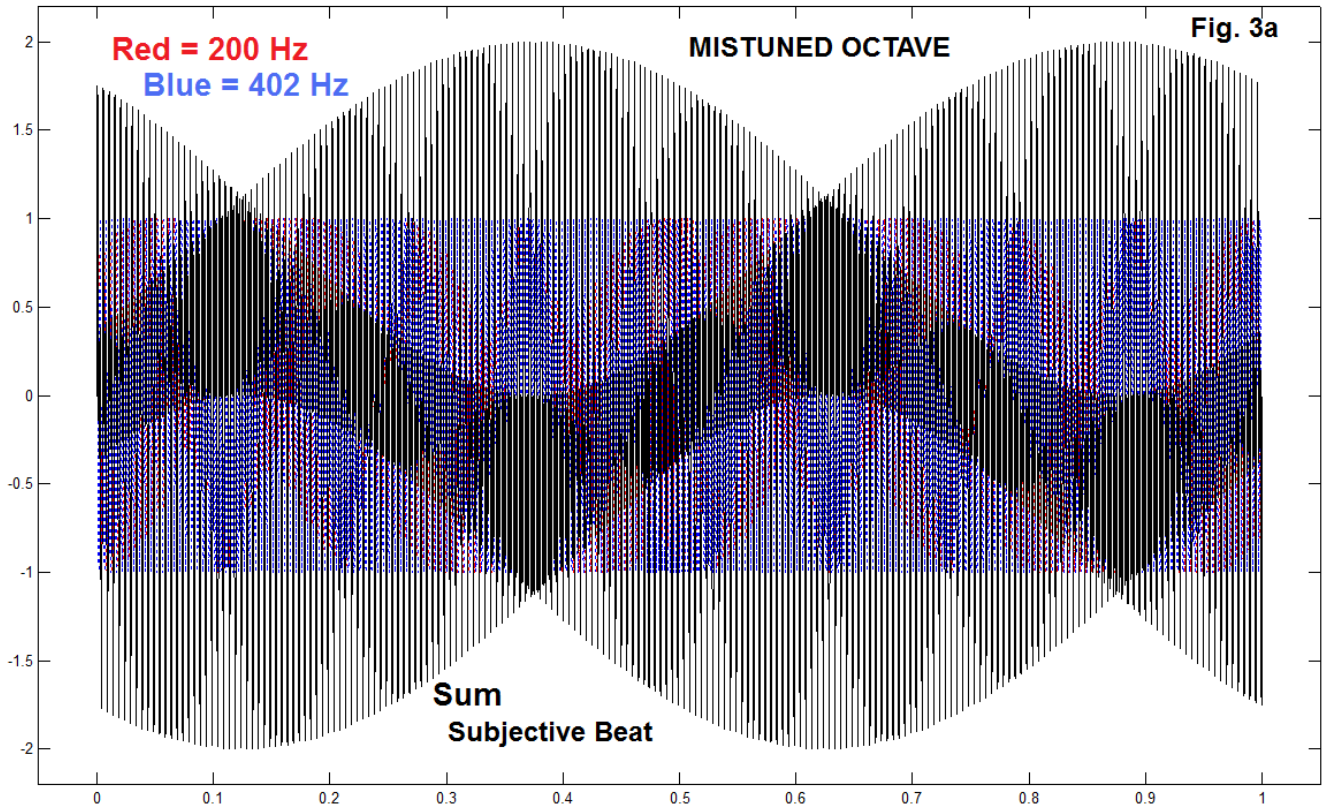
The classic case here is the beating close to "unison" (nearly same frequencies). Soon we will look at the case of complex waveforms, but for the moment, we will stick with sinewaves and will use large ratios. We should perhaps mention that many of the figures here (the ones with the "a" suffix) contain 10,000 points, so the details are lost in the plot. Further, at times these are even difficult to calculate and plot – such things as "screen aliasing" (like a "Moiré pattern") can fool us – we think these are for the most part avoided.

Fig. 2a shows the results of the sum of two waveforms at an octave. The important thing about this figure in comparison to Fig. 1a is that it does not show large amplitude variations – there are no apparent amplitude beats. While it looks from the plot like there are lobes and level changes, what is actually happening is best appreciated by looking at the zoom-in of Fig. 2b. Any zoom-in along the waveform of Fig. 2a will look essentially the same as Fig. 2b. That's all there is – anything else is an artifact of the plotting. Fig. 2a is just a complex waveform consisting of a fundamental and one harmonic (equal amplitude). No beating is occurring. Here we have started with the ratio set to a perfect 2:1 (octave).





Sinewaves at an exact octave ratio as in Fig. 2b at left is <u>perfectly periodic,</u> and represents all of Fig. 2a,



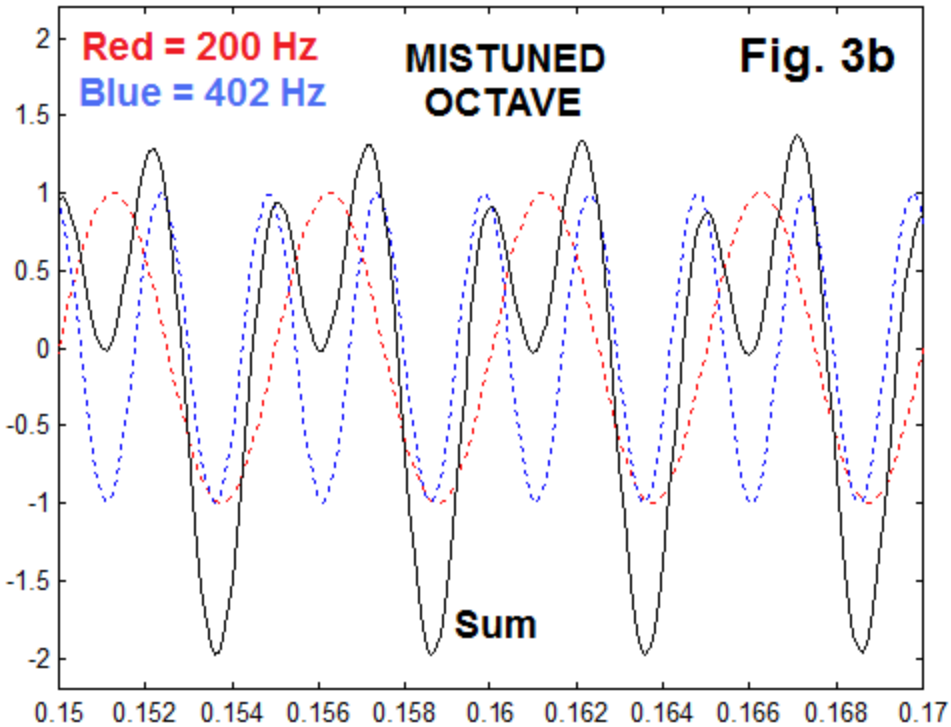
We are now in the position to look at the imperfect octave. Fig 3a shows the case where the upper frequency is moved from 400 Hz to 402 Hz. Once again, this screen/plot is a mess, but what we do see here is that some amplitude variations are occurring, but there is no full cancellation to zero, as there was in Fig. 1a, but rather an overall up/down waver. There are again two complete "events" here. We again get a better feel for what is happening by looking at the zoom-ins, Fig. 3b and Fig. 3c, which we compare to Fig. 2b. While Fig. 2b showed what was going on locally, it is the same for the entirety of Fig. 2a. Fig. 3b and Fig. 3c are snapshots of Fig. 3a. It is hard to see a real change across the limited time width of Fig. 3b and Fig. 3c, but it is there.

One major difference here is seen be going back to equation (1). The frequency corresponding to (A-B)/2 which was 1 Hz in Fig. 1a, the "envelope", is now 101 Hz. Fig. 3d and Fig. 3e e show the situation at about time 0.38 of Fig. 3a, where the overall waver goes positive the most and negative the least. The black curve in both cases is the exact same. In Fig. 3d it is the usual sum, while in Fig. 3e it is the product (hence the light blue and green, solid, curves). The light blue is the same general cosine curve as Fig. 1a, just a much higher frequency.

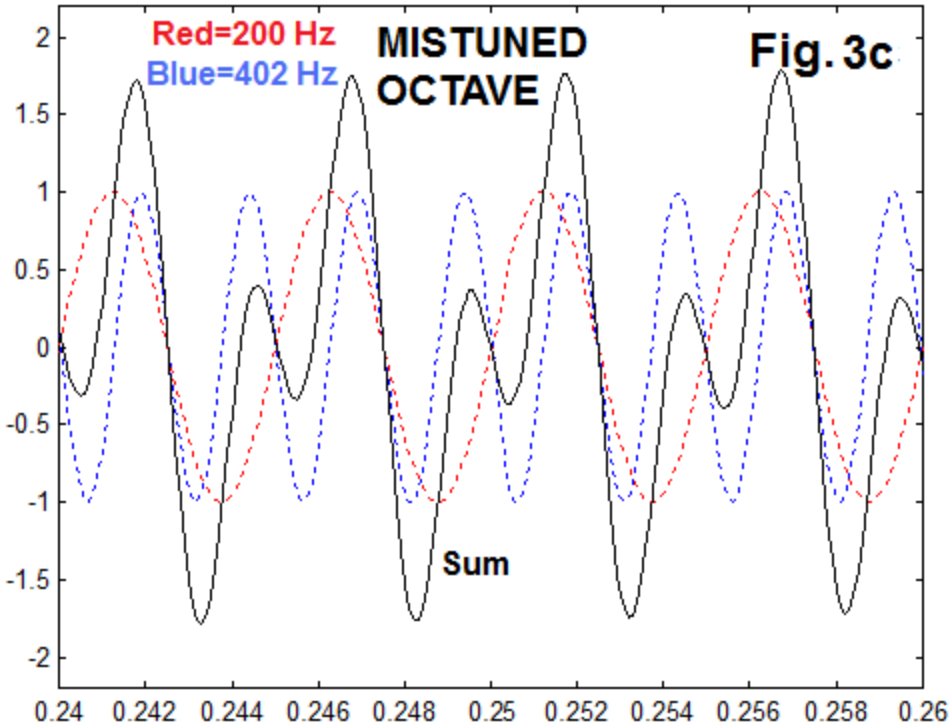
Here we have a "beat" but not one that is full amplitude. These are sometimes called "second-order" or "subjective beats". They are evident to the listener (something is going on twice each second) but it is not by any means as objectionable as the full amplitude variations.

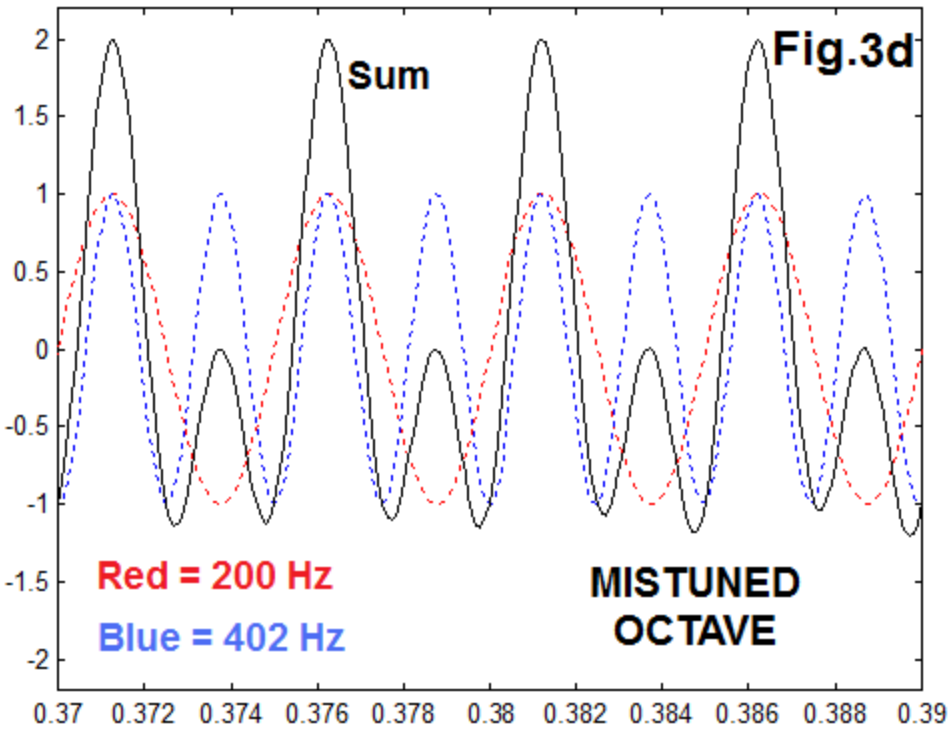
Similar "subjective beats" occur at other small integer ratios, and here as a second example (we could of course consider many more) we look at a 3:2 ratio which is a musical fifth. Fig. 4a shows the case where there is a perfect 3:2 tuning, 300 Hz and 200 Hz. As in Fig. 2a, we see some artificial structure due to the limited resolution of the screen/plot. But clearly there is no apparent amplitude variation. Fig. 4b shows a zoom-in which explains a lot of the artificial plotting "structure" of Fig. 4a. Note that Fig. 4b is typical of any zoom-in that we could have taken from Fig. 4a. It is just a periodic waveform. This is a case of great interest because it has a periodicity of 100 Hz, but no spectral energy at 100 Hz. That is, the famous case of a "missing fundamental". No beating, but an important result.

So we will conclude our look at the mistuned integer ratios by letting the 300 Hz frequency go up to 302 Hz. This result is shown in Fig. 5a, with Fig. 5b and Fig. 5c providing the familiar zoom-ins. As in the case of the mistuned octave, we see in Fig. 5a the usual mess of screen/plot irregularities, but the main point is that there are amplitude variations, but here they are less extensive than in the mistuned octave, and certainly aren't very deep as compared to Fig. 1a. Fig. 5b and Fig. 5c do show that there is a gradual change of wave shape as we move through Fig. 5a, much as we saw with Fig. 3a.

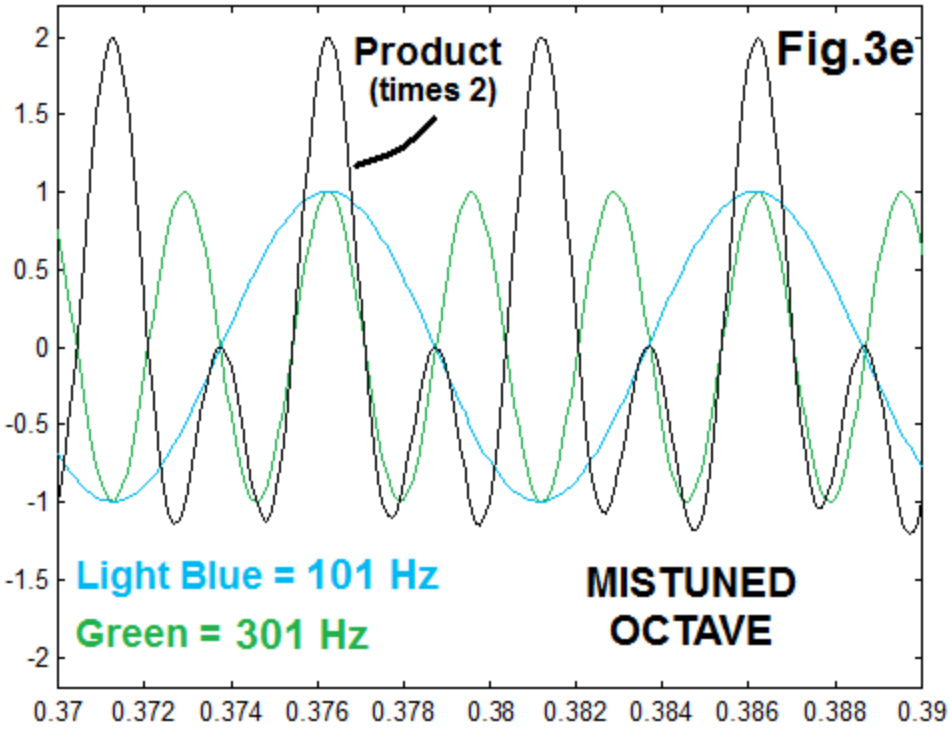


Slow changes across Fig. 3a are better seen by comparing the "snapshots" of Fig. 3b at left and Fig. 3c just below.

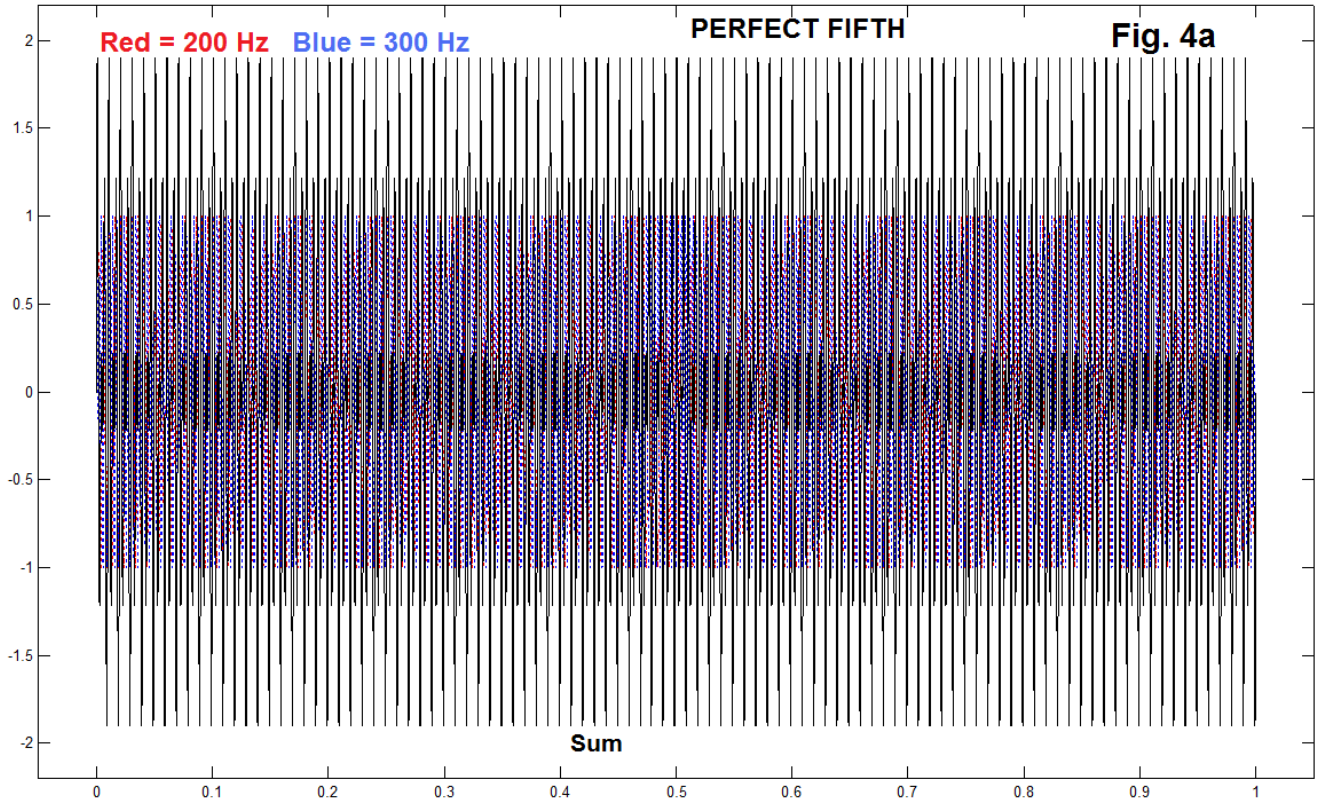


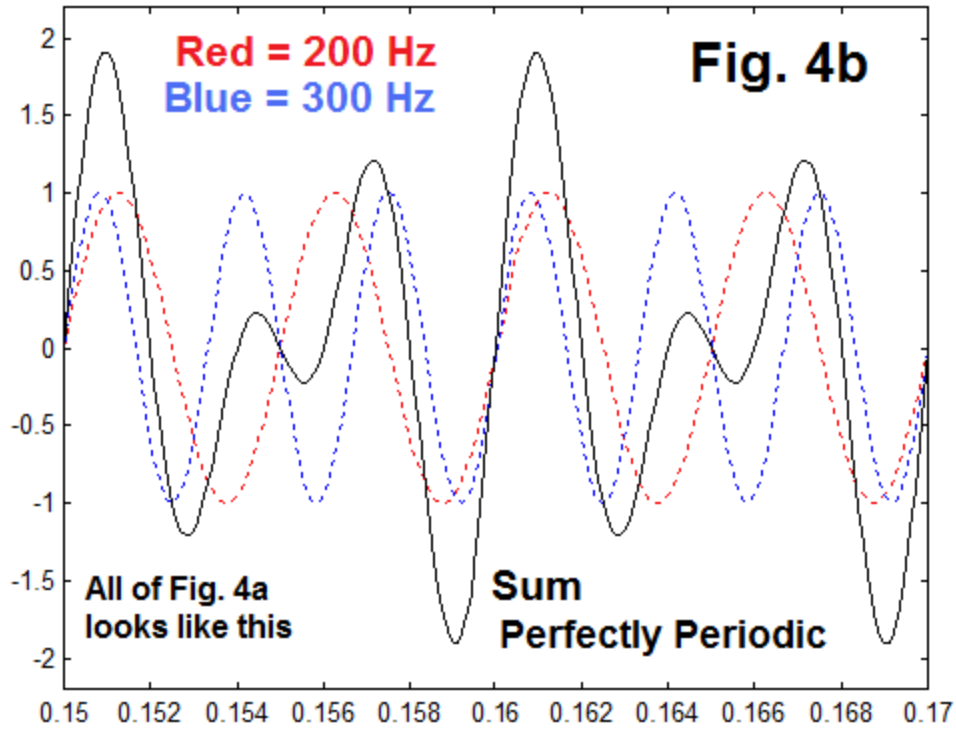


This is the sum form of equation (1)

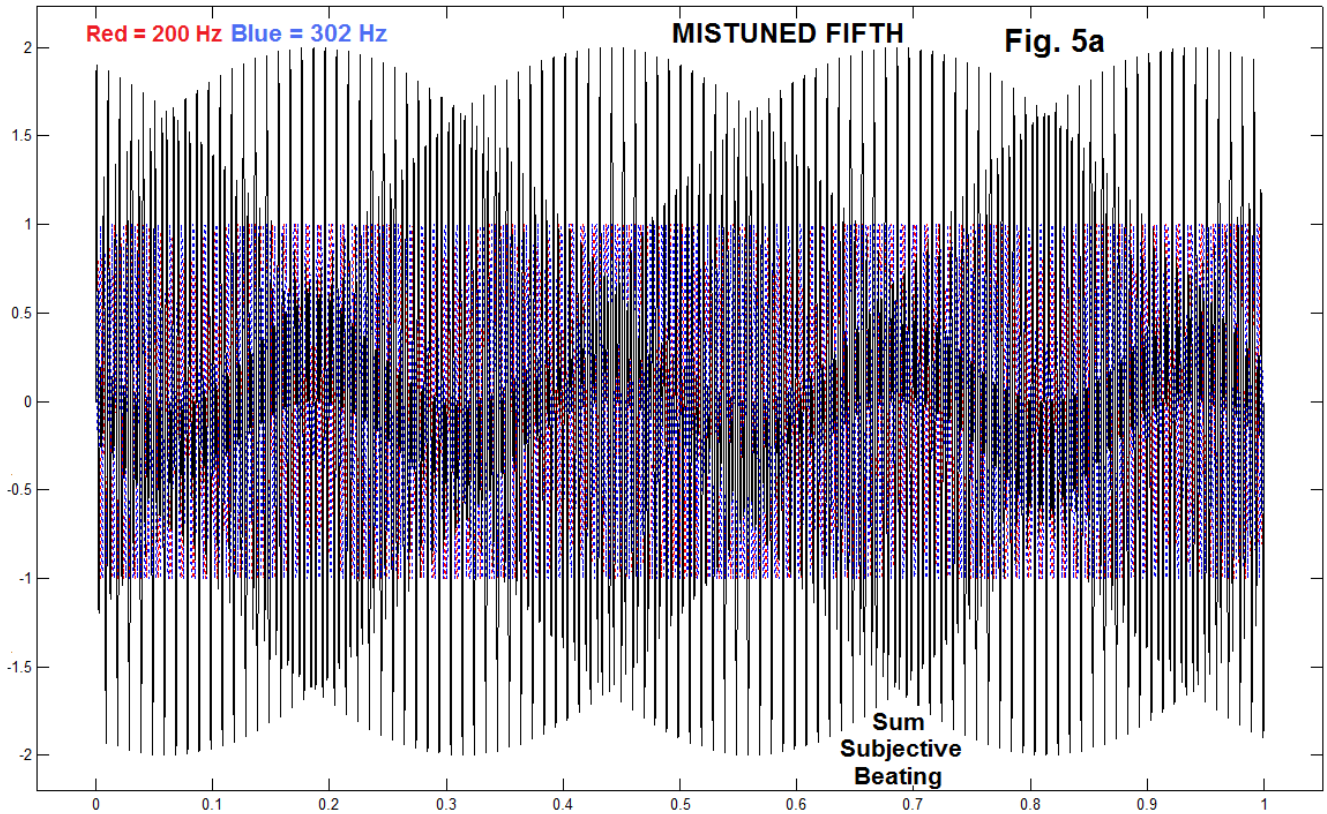


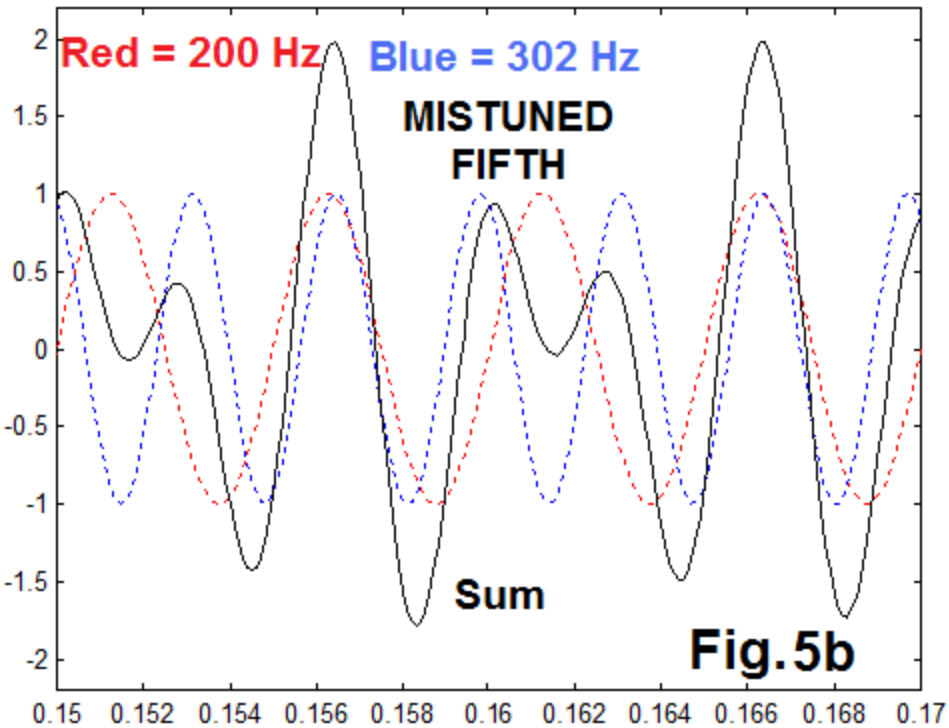
Note that this is the product form of equation (1)



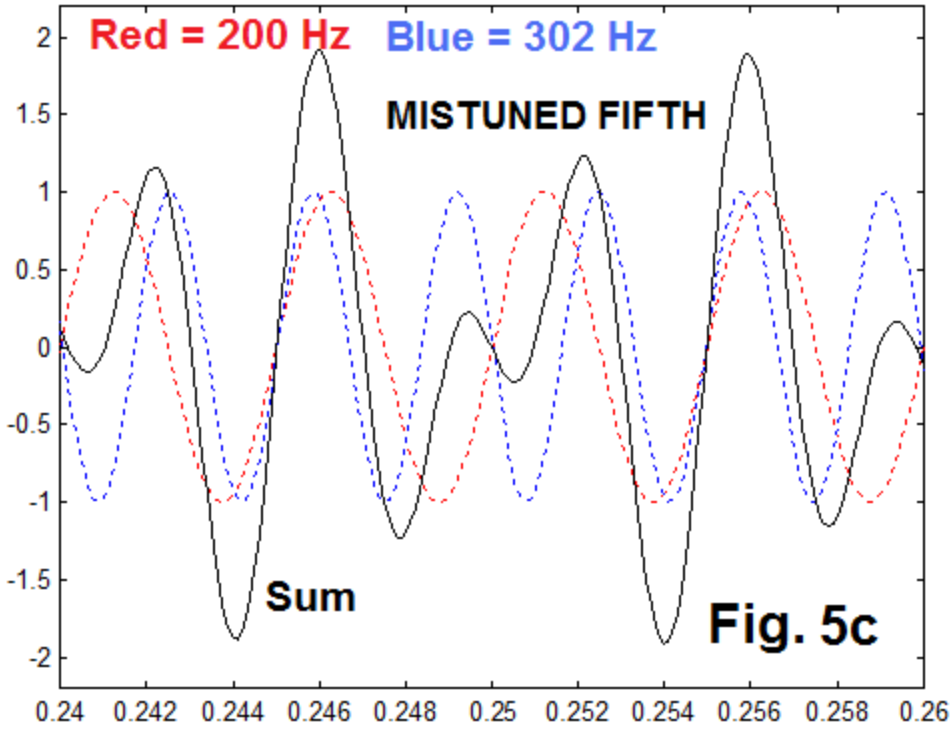


Snapshot of Fig. 4b at left is perfectly periodic and represents the entirety of Fig. 4a.





A snapshot of the imperfectly tuned fifth. A second snapshot is shown in Fig. 5c.



A second snapshot of the imperfectly tuned fifth (compare to Fig. 5b above).

So the answer to question (3) is basically that we get a subjective beating to mistuned ratios of small integers. We are aware of the mistuning, and might well prefer not to have even the subjective beats, but they are nowhere near as annoying as the large, full, amplitude beats of the mistuned unison.

NOW WHAT ABOUT COMPLEX WAVEFORMS:

To some degree you might suppose that we have addressed this issue because the ratios of small integers are elementary complex waveforms. We mean by complex waveforms that they are composed of harmonics as well as (often, but not strictly required), a fundamental. They are represented therefore by a familiar Fourier series. But what we do below is still quite different.

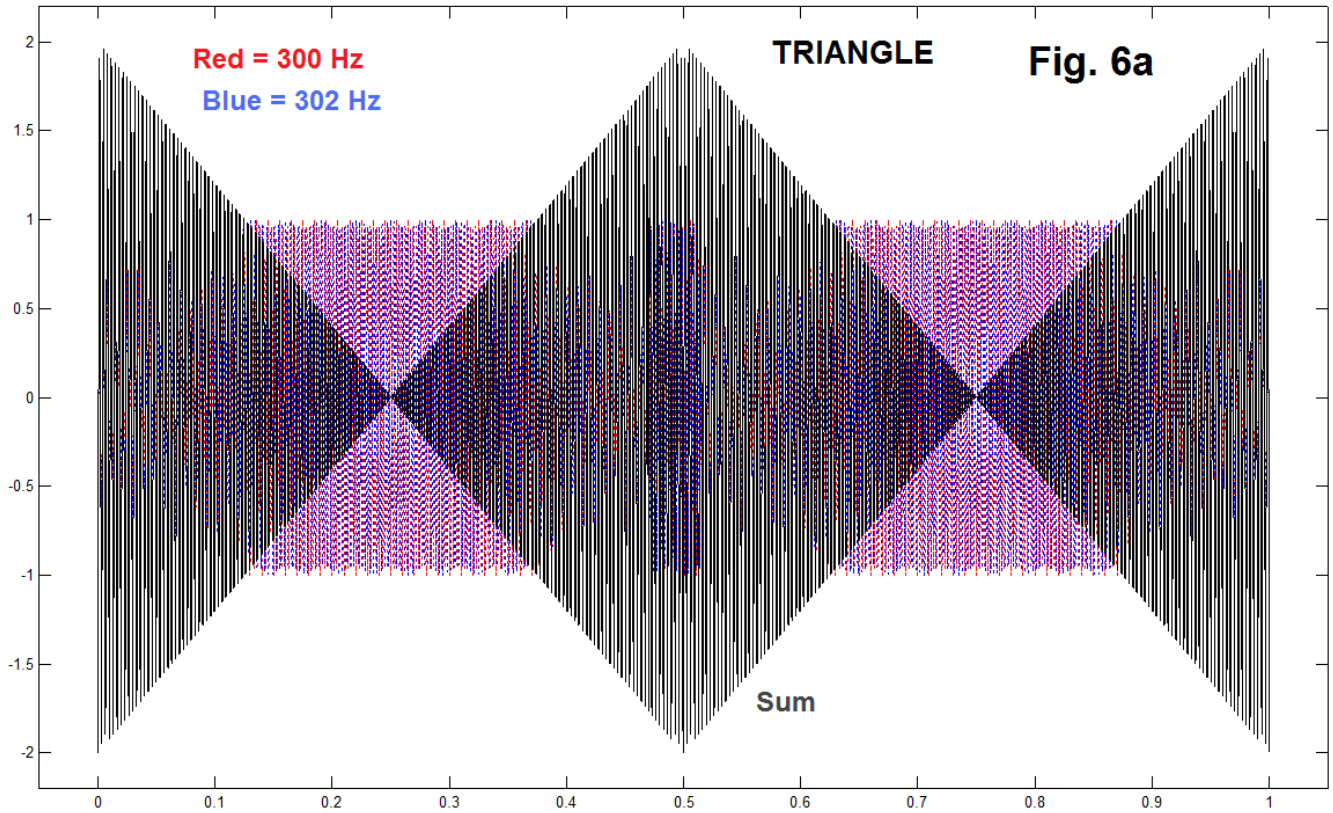
For one thing, we will be talking about a mistuning of the entire waveform, the fundamentals, and any harmonics by a proportional amount. The main difference in our

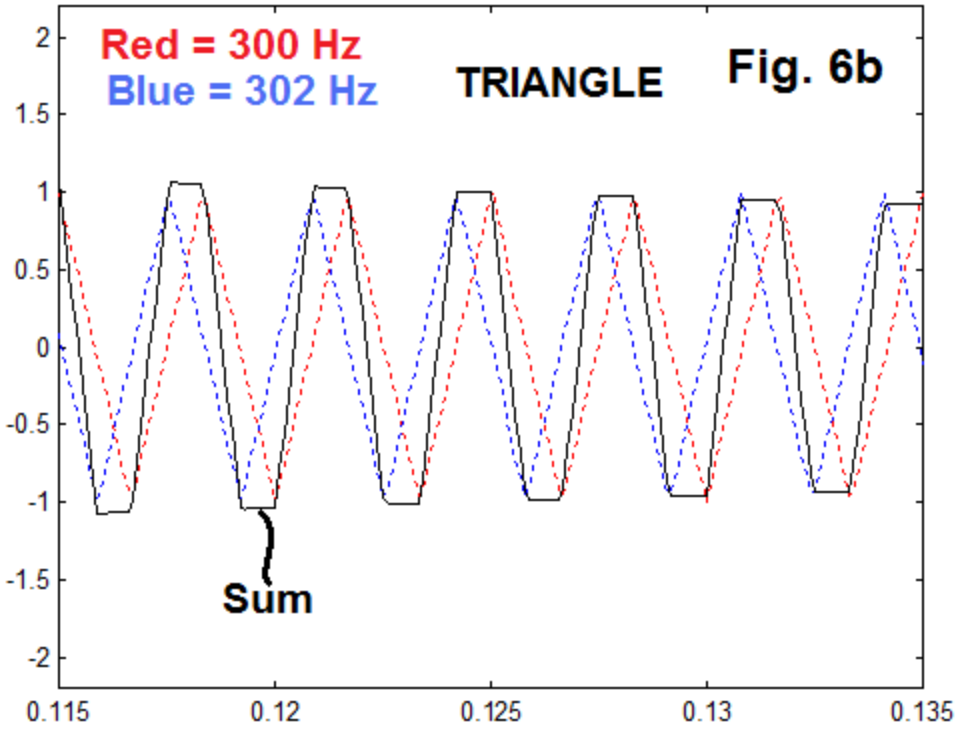
eventual findings (to give away the dénouement!) is that <u>even for mistuned unisons</u>, for waveforms with large discontinuities (like square and saw), <u>annoying large amplitude beats</u> will not be present. This is because harmonics as well as fundamentals beat separately. When one cancels, the others often will not do so at the same time. This has important implications for harmonically rich, real acoustic instruments that may play at unison, as well as at intervals (in harmony). Here we will deal only with the complex waveforms in unison.

A few notes on generating these complex waveforms is appropriate. First, all the same screen-aliasing patterns that we had with sinewaves are still present. Accordingly, the most reliable views of what is really going on are offered by the “b”, “c”, and “d” versions of the figures, while the “a” version is the best overall view of whether or not large amplitude beats are present.

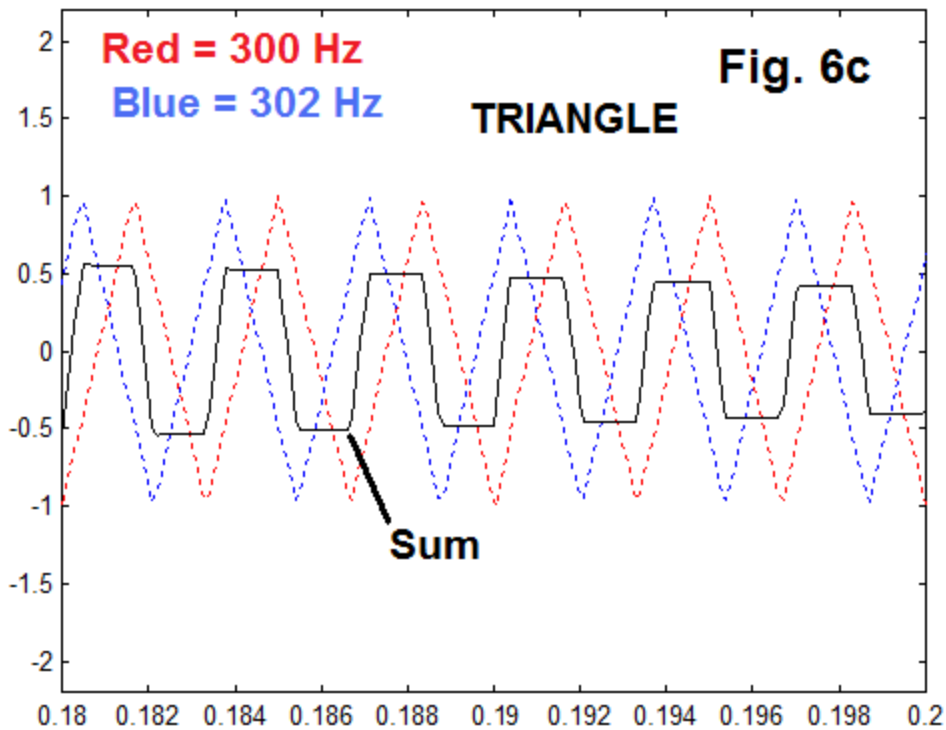
Secondly, we intend to use here examples with triangle, sawtooth, and square waves – those that are very familiar from our synthesizer work. How do we generate these? There are many ways. I am using Matlab here for all these calculations and plots, and there are built-in functions for saw, triangle, and square. For a better overall feeling of control of my investigation, I wrote my own code for these. In doing so, we can draw on a lot of experience with waveshaping, both analog and digital. For example, I get my triangle from a saw by taking the absolute value, doubling it, and subtracting 1, clearly an analog approach. An alternative digital version of getting a triangle (going back to the programmable calculator days!) is to use the built-in sine function and then take an arcsine. Notably, we have to be careful using our own code or checking someone else’s. For example, in writing a sawtooth it seems automatic to increment a staircase starting at perhaps 0 and then when it exceeds 1, reset it to -1, etc. This would seem to works. Or you might just subtract 2 from the overflowed value. That too would seem to work. But if you think about it, and remember that we are trying to generate two waveforms with close, but different frequencies, you have to subtract 2. Otherwise both cases will generally overflow the +1 during the same time step, be reset to -1, and reproduce the previous cycle. They will be locked at the same frequency. In fact, it is probably good to be reminded from time to time that interesting things happen when we do something that “obviously will work exactly as expected”. So we have checked our waveforms carefully.

The simplest of our complex waveforms, in the sense that it is most like the sinewave, is the triangle, and we see from the results (Fig. 6a compared to Fig. 1a) that we get similar deep (full) amplitude variations. Even though the triangle has harmonics, they do all cancel at the same time. This we could show from the Fourier series, but it is probably completely evident in the time domain (Fig. 6d). (In the case of the square wave that follows, which has exactly the same harmonics as the triangle, although stronger, we will still get full cancellation, but a nearly completely disguised amplitude dip, because the amplitude dip is instantaneous and of short duration, and because of the broader spectrum).

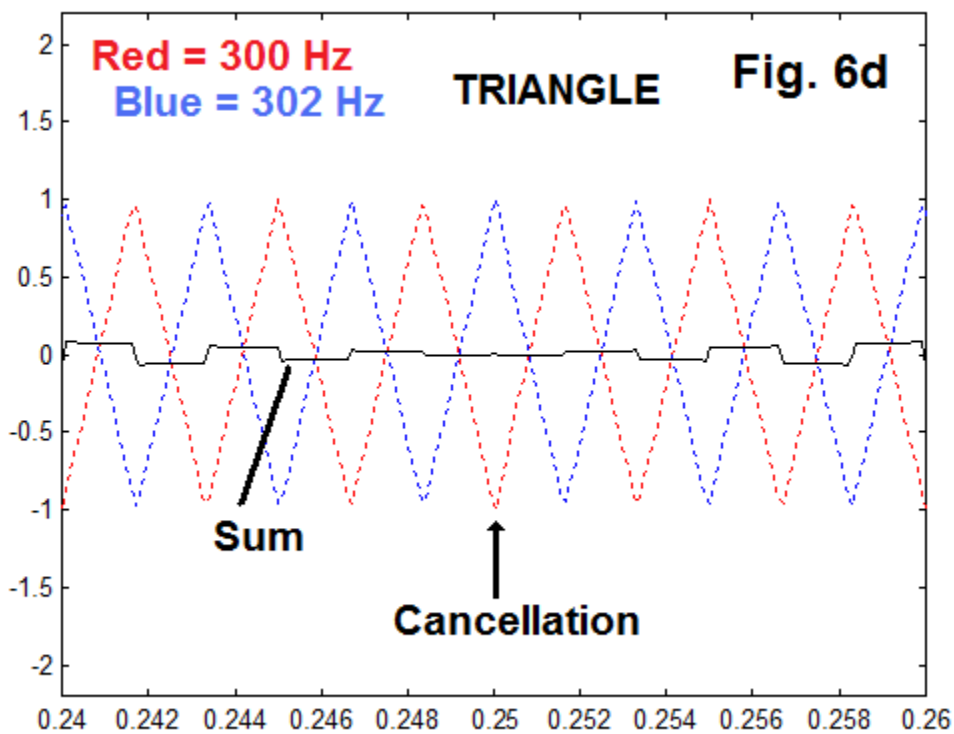




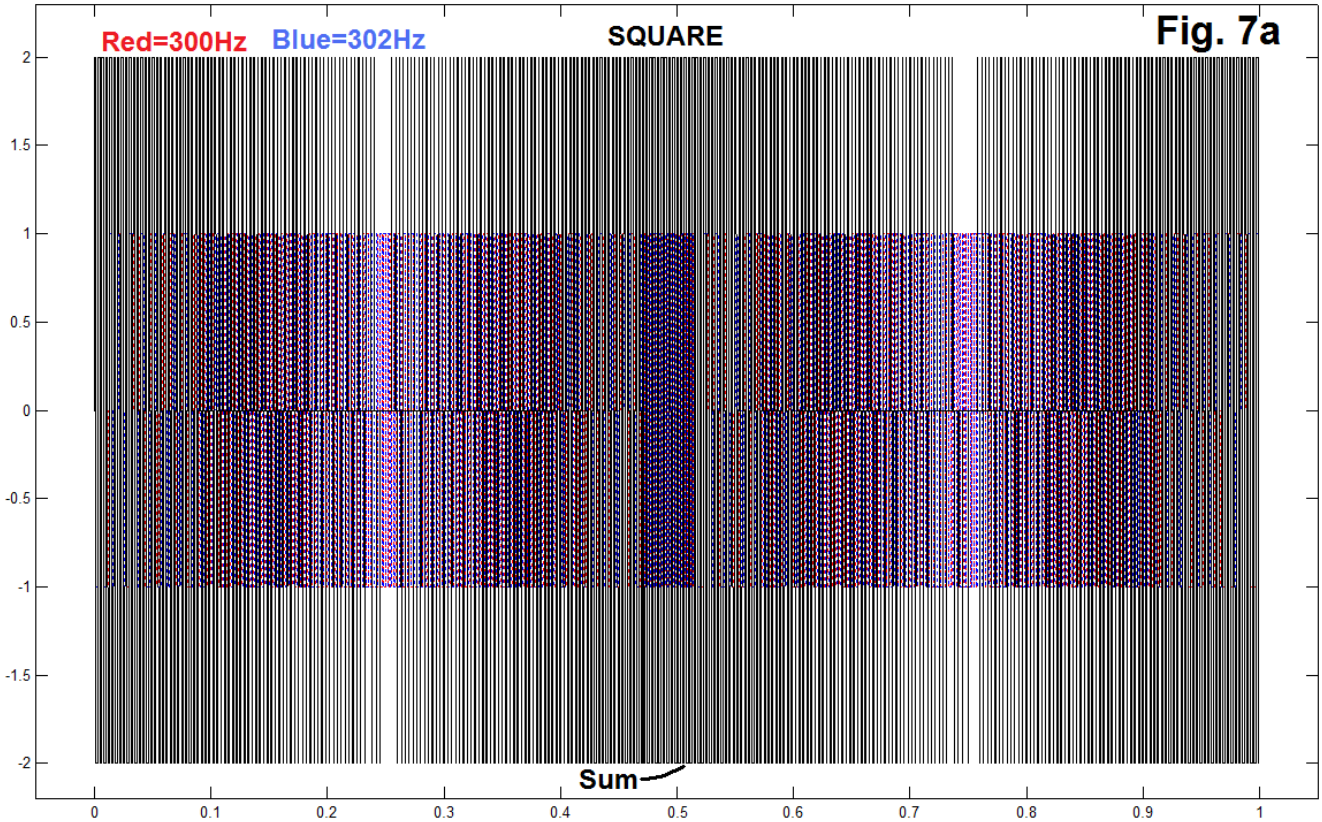
Sum of two triangles has twice the slope and a flat top – looking more like a square wave

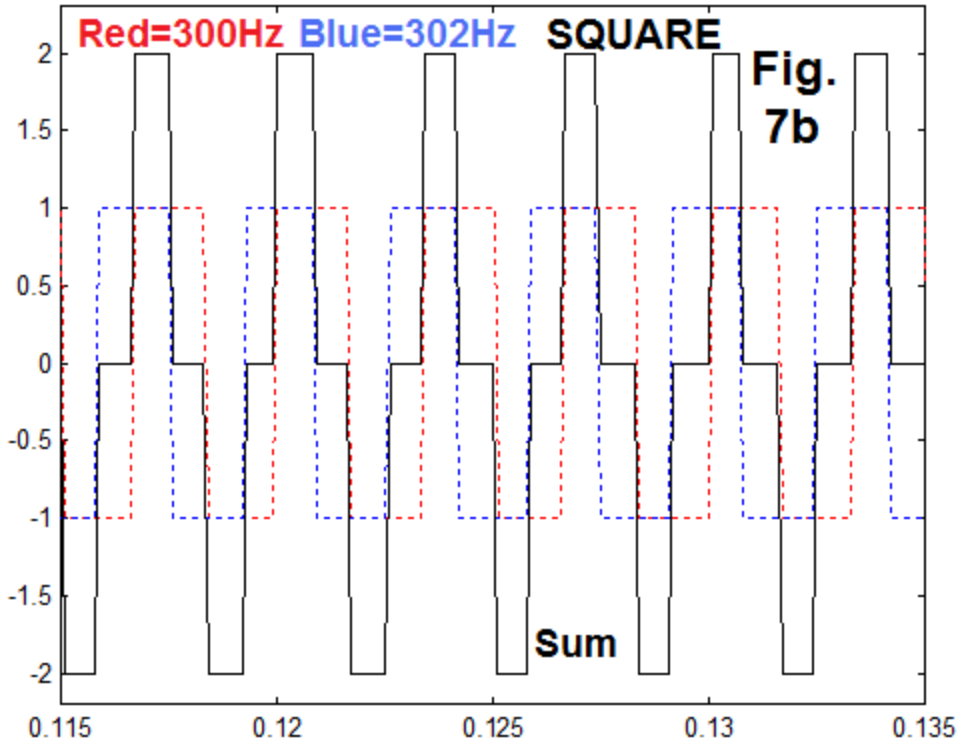


Process continues

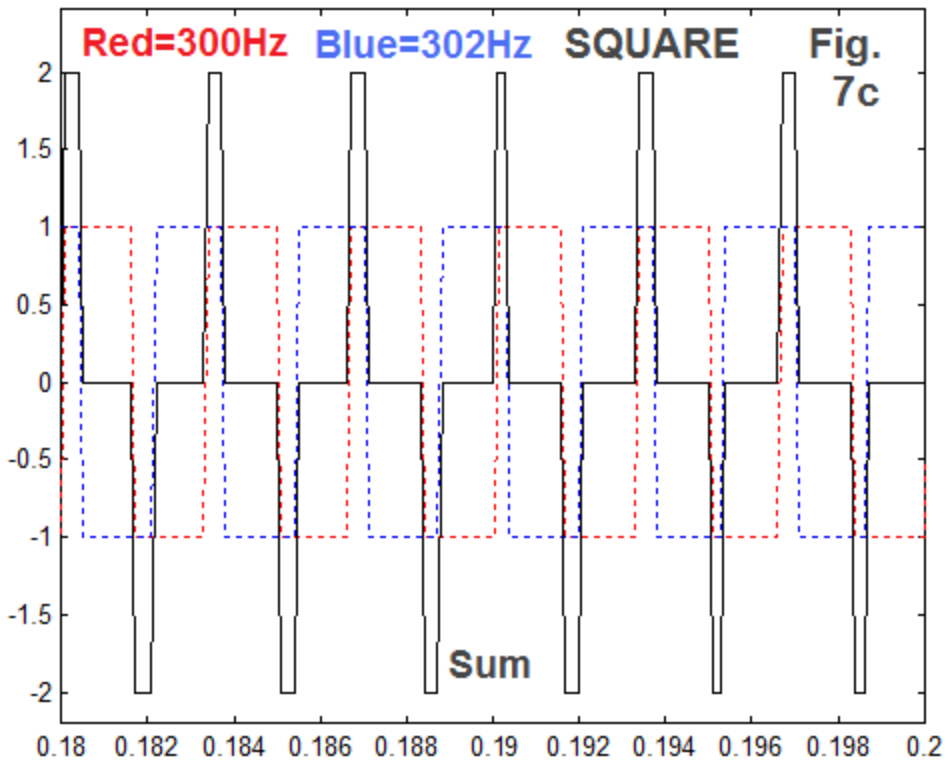


Complete cancellation

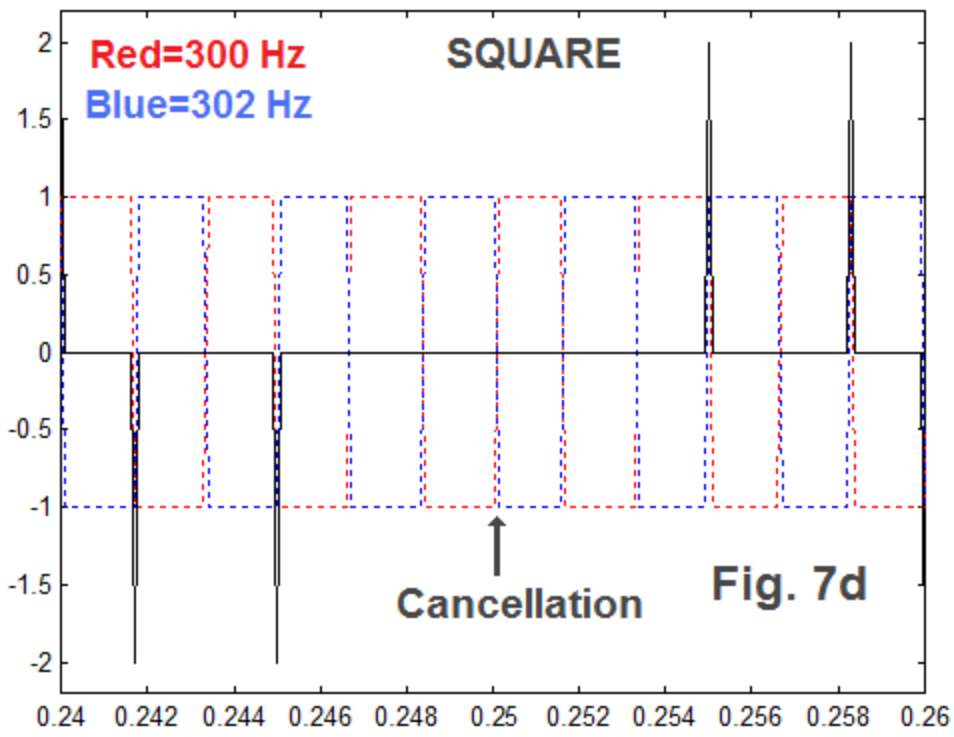




Sum of square waves has only sudden jumps, and only three levels (+2, 0, and -2).



Process continues



Perfect cancellation at the very center, but “spikes” were not gone long. The asymmetry in reappearance is due to roundoff.

It is clear that we are summing two perfectly periodic waveforms, each with its own harmonics that are exact integer multiples of their respective fundamental frequency. Thus the components of what we are adding together can be represented by their Fourier series. Accordingly we could just sum all the sine waves, considering beating in pairs. We mention immediately however that this is much easier to understand by a direct observation in the time domain. The small difference in frequency looks a lot like, although technically not the exact same things, as a gradual phase shift. The waveforms shift apart, and then at some point, they may cancel in whole or in part. This was obvious for the sinewave and its close cousin, the triangle. Less so for the square, but still true. Compare Fig. 1c for the sine, Fig. 6d for the triangle, and Fig. 7d for the square. However, note that we do not hear a pronounced amplitude beat for the square, as we do for the sine and triangle. In the case of the square, the sum disappears by getting thinner rather than by reducing amplitude (compare to the triangle). Moreover, as the sum gets narrower in the case of the square, it's harmonic content rises (narrower pulses), giving it more apparent impact aurally. It is quite distinct from the triangle case. Additional details are written next to the figures.

Although less direct, it is interesting to compare what happens in the time domain to the frequency domain. Note that the Fourier series for the sine, the triangle, and the square are all odd harmonics only.

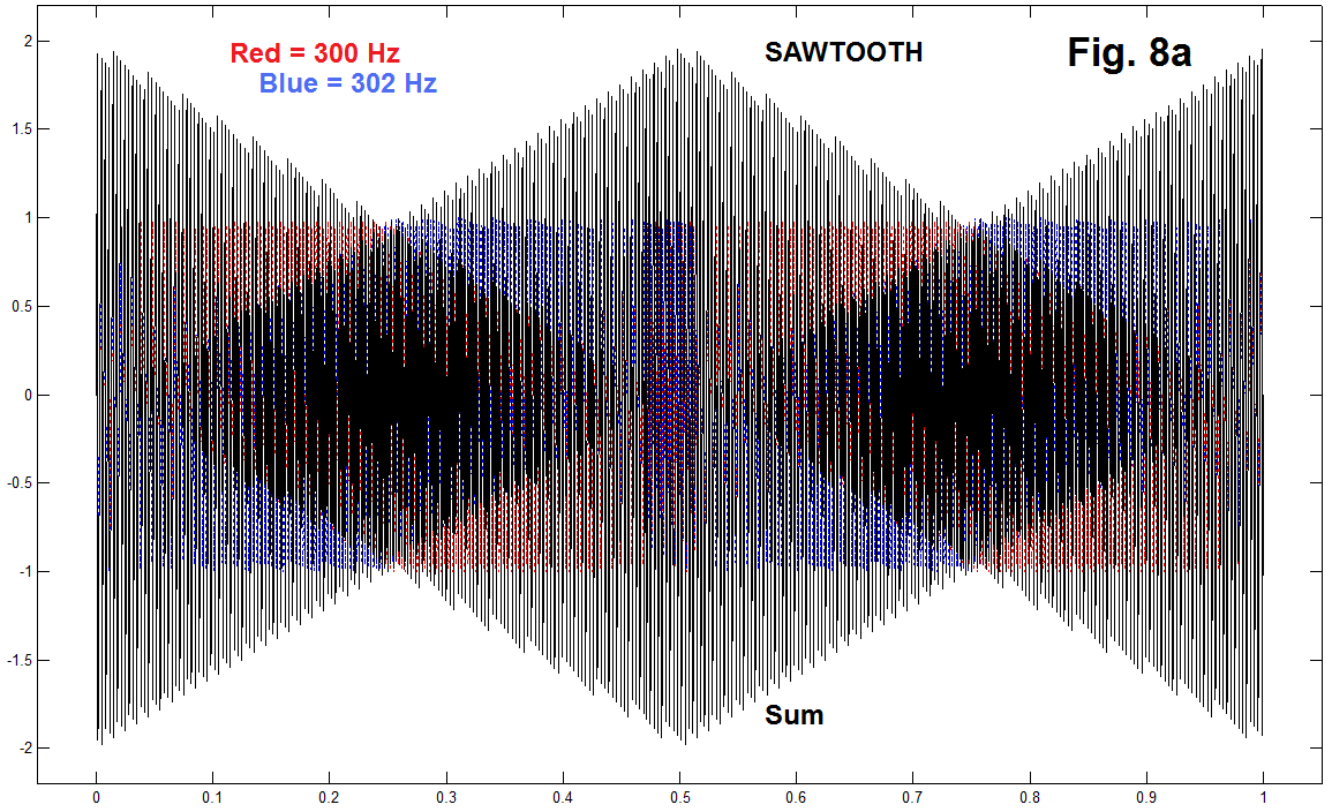
$$x_{sine}(t) = \sin(2\pi t) \tag{3a}$$

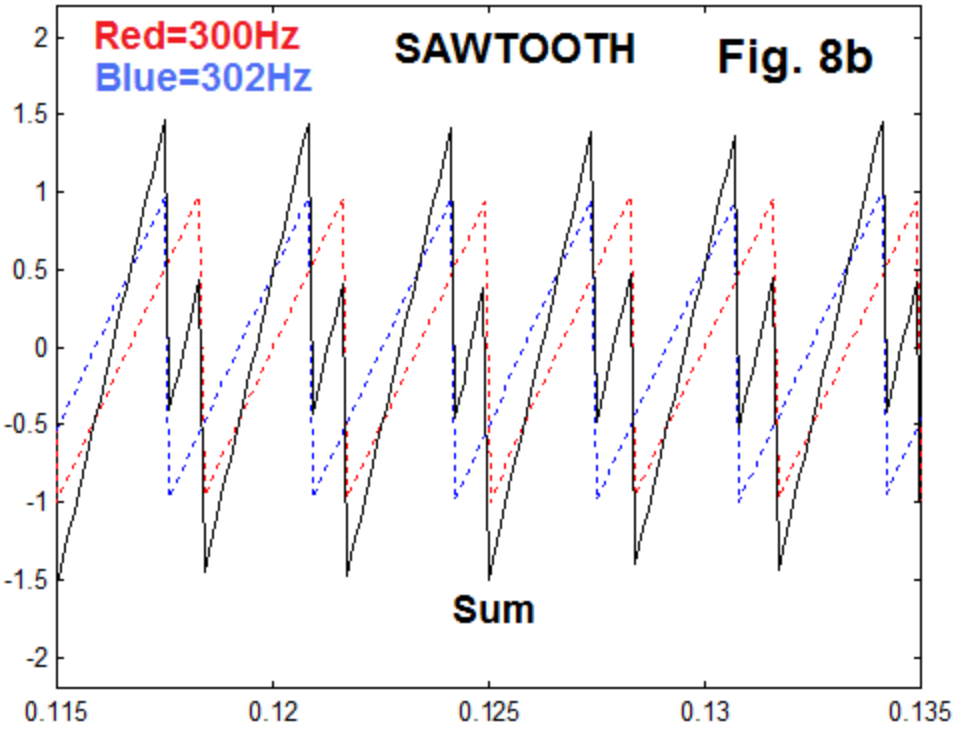
$$x_{triangle}(t) = \left(\frac{\pi}{4}\right)\left[\sin(2\pi t) - \left(\frac{1}{9}\right)\sin(6\pi t) + \left(\frac{1}{25}\right)\sin(10\pi t) - \left(\frac{1}{49}\right)\sin(14\pi t) + \cdots\right] \tag{3b}$$

$$x_{square}(t) = \left(\frac{4}{\pi}\right)\left[\sin(2\pi t) + \left(\frac{1}{3}\right)\sin(6\pi t) + \left(\frac{1}{5}\right)\sin(10\pi t) + \left(\frac{1}{7}\right)\sin(14\pi t) + \cdots\right] \tag{3c}$$

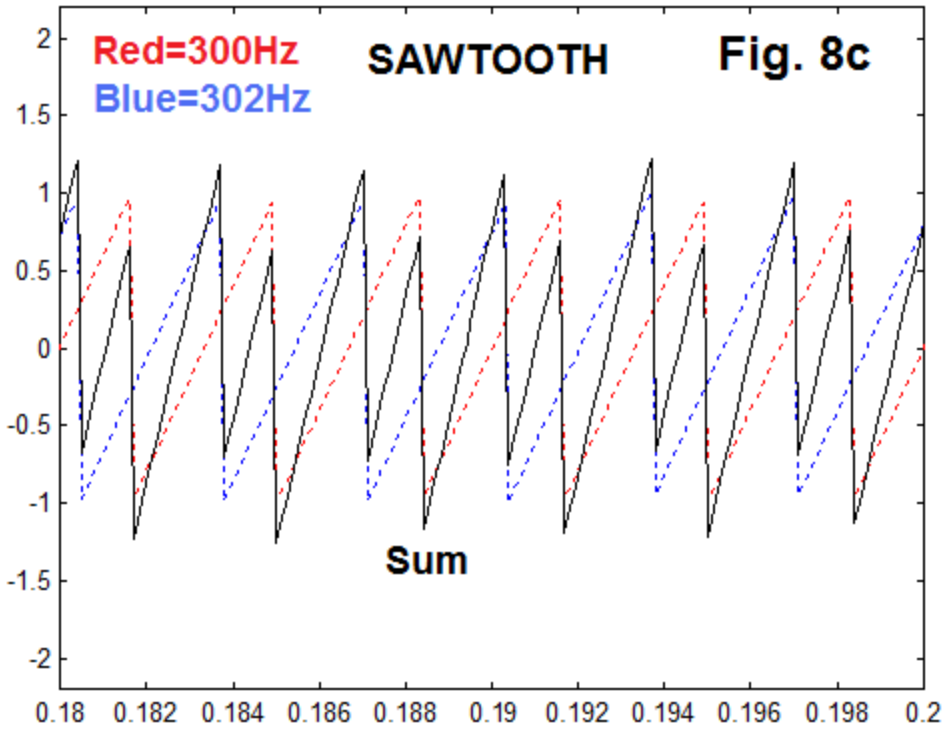
This means that when the fundamental cancels, so do all the harmonics, which have the same symmetry.

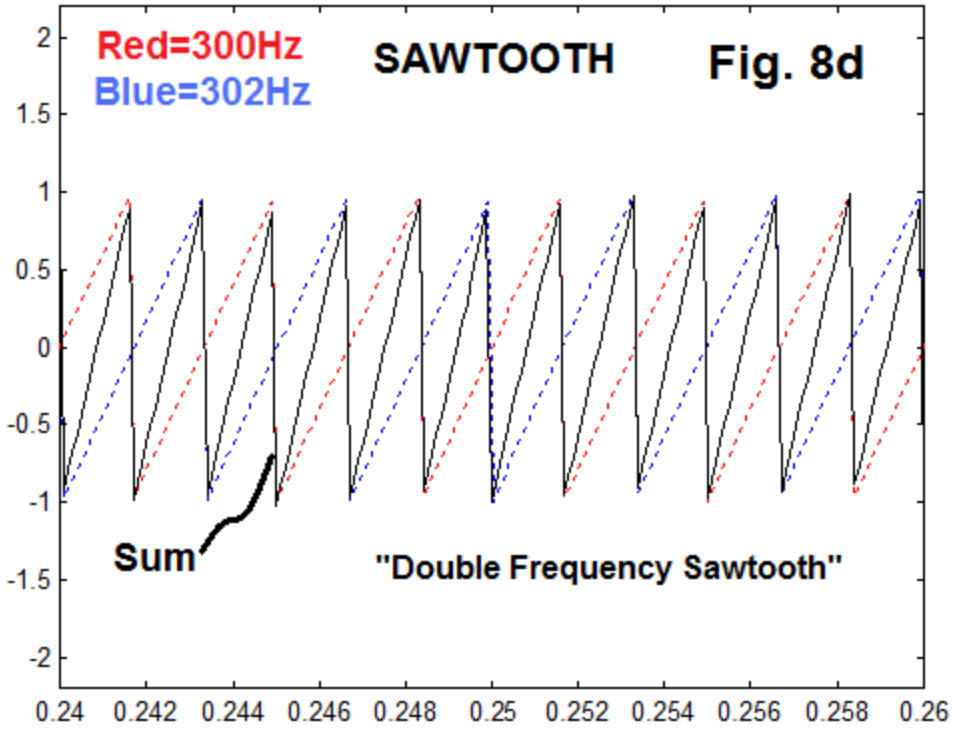
Fig. 8a gives us the case of the sawtooth waveform. Here it is quite apparent that the amplitude varies by a 2:1 ratio, but never drops below 1/2 its peak value. Again, Fig. 8b, Fig. 8c, and Fig. 8d provide the useful details, and they make perfectly good sense in the time domain. In fact, this lovely result shows what is clearly an amplitude minimum accompanied exactly with a doubling of the frequency (Fig. 8d). This sharpening of the spectrum would in itself (even without the fact that half the amplitude remains) reduce the aural impact of the beating.





Obvious summation of sawtooth waves





Here we have the half-amplitude/ double-frequency limiting case

What has happened in the case of the sawtooth (Fig. 8d) is that the fundamental has exactly been cancelled, as have all the odd harmonics, just as the odd harmonics were all cancelled in the triangle and square. But in the case of the triangle and square, there were no even harmonics remaining, so we had a dip to zero amplitude. Here the even harmonics are not cancelled, and this, while understandable in the time domain, is quite neat in the frequency domain.

The Fourier series for the sawtooth wave is:

$$x_{sawtooth}(t) = \left(\frac{-2}{\pi}\right)[\sin(2\pi t) + \left(\frac{1}{2}\right)\sin(4\pi t) + \left(\frac{1}{3}\right)\sin(6\pi t) +$$

$$\left(\frac{1}{4}\right)\sin(8\pi t) + \left(\frac{1}{5}\right)\sin(10\pi t) + \left(\frac{1}{6}\right)\sin(12\pi t) + \left(\frac{1}{7}\right)\sin(14\pi t) +$$

$$\left(\frac{1}{8}\right)\sin(16\pi t) + \left(\frac{1}{9}\right)\sin(18\pi t) + \left(\frac{1}{10}\right)\sin(20\pi t) \ \ldots \ ] \qquad (4a)$$

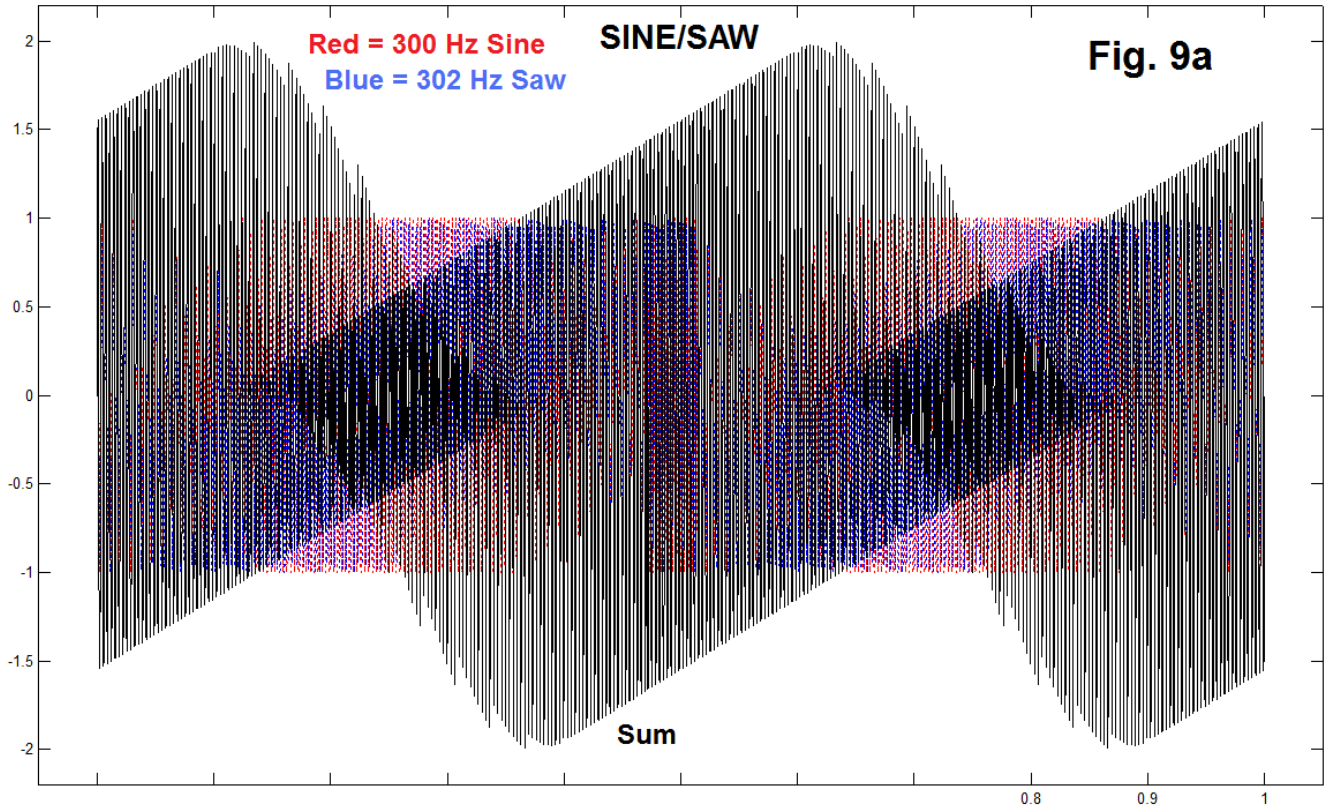
When the fundamental disappears, so do all the odd harmonics. [Note that the harmonics are the <u>integer multiples of 2π</u>. Thus 2π, 6π, 10π, … are the odd harmonics, and 4π, 8π, 12π… are the even harmonics.]

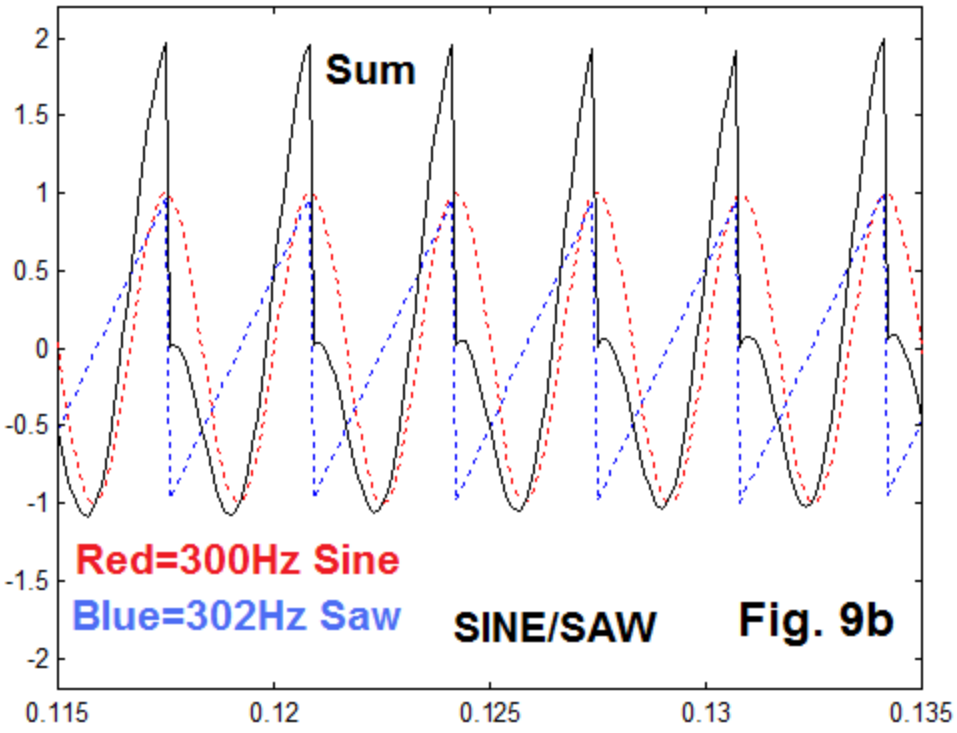
What is left is:

$$x_{sawtooth}(t)' =$$

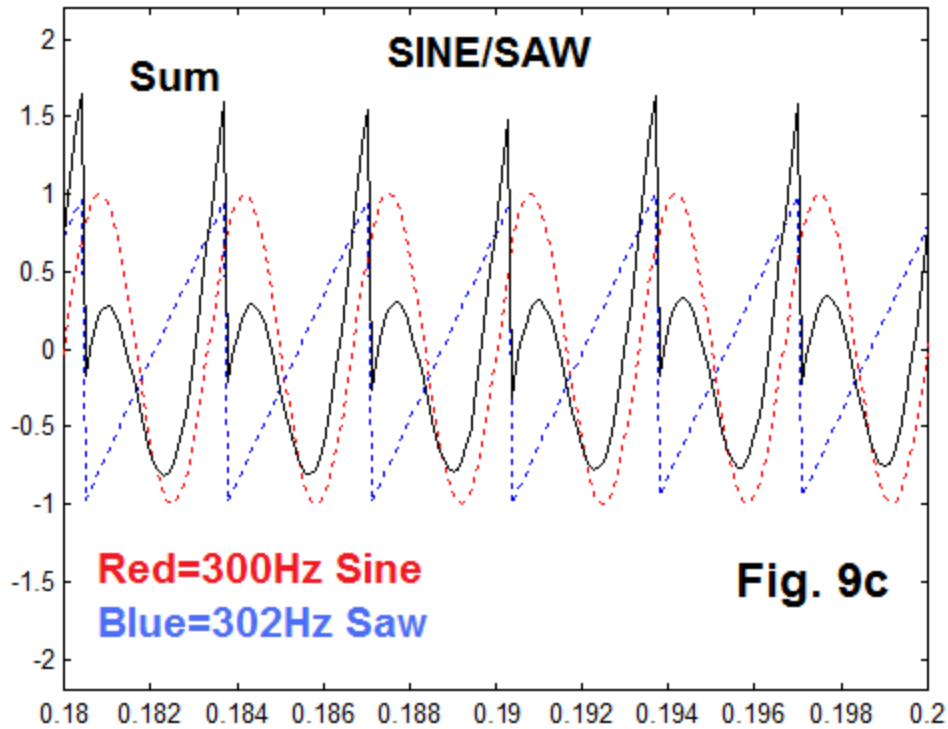
$$\left(\frac{-2}{\pi}\right)\left[\left(\frac{1}{2}\right)\sin(4\pi t) + \left(\frac{1}{4}\right)\sin(8\pi t) + \left(\frac{1}{6}\right)\sin(12\pi t) + \left(\frac{1}{8}\right)\sin(16\pi t) + \ \ldots\right] \qquad (4b)$$

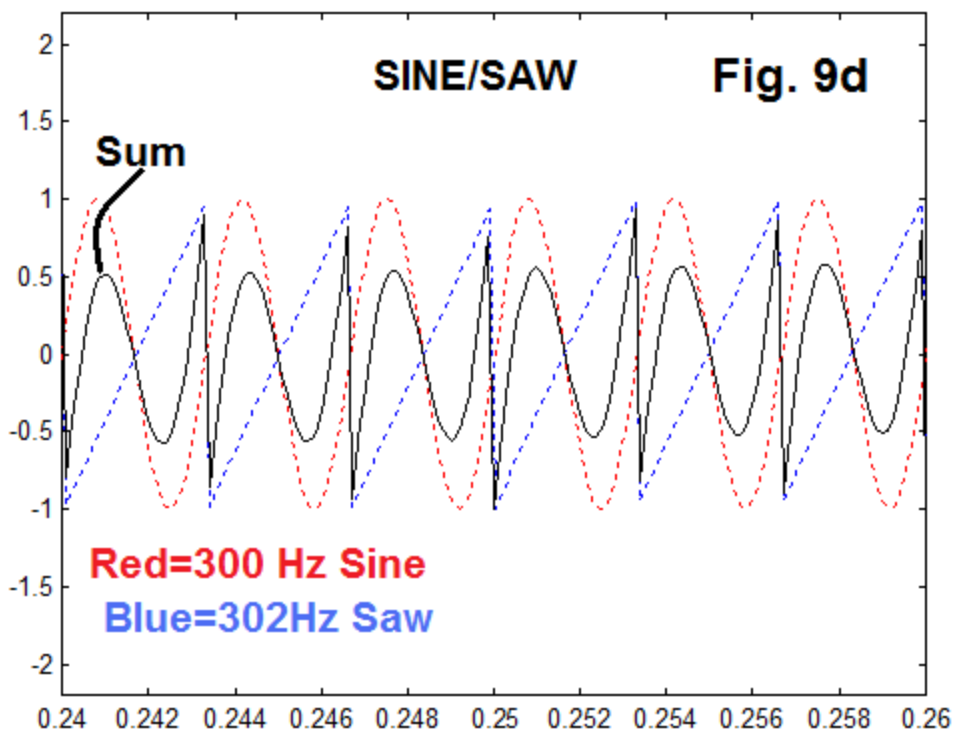
which is exactly the Fourier series of a sawtooth of twice the frequency and 1/2 the amplitude of the original case.





Following the discontinuity in the sawtooth, the fall of the sine is contrary to the rise of the sawtooth. Prior to the discontinuity, the rises added, <u>at this phase</u>.





Here is the minimum amplitude

Note that considerable fundamental remains, and we have that discontinuity.

We have looked at a lot of cases, and could study many many more. However we have likely looked at enough representative case to make the point we need. In particular, the answer to question (4) appears to be that in cases of rich harmonic structure (even and odd harmonics – such as the sawtooth) we can expect that large amplitude beats may not be present, or not have much aural impact, even at the mistuned unison. In many, and probably in most cases, we are left with more of a subjective beating.

Note that while the sawtooth was our main (only!) example of a waveform containing all harmonics (even as well as odd), the "pulse" is well known in music synthesis as having a potentially very wide spectrum of all harmonics and corresponding "aural bite". Such a pulse is merely a square wave with a duty cycle that is not 50%. Typically it might be something like 10% and thus contains all harmonics except every 10th. So its spectrum can be very wide. It too does not have large beats in amplitude. It was not used in this study because it was too simple – just two generally non-overlapping narrow pulses added together. It is left as the proverbial exercise for the reader.

So while we have answered all four questions, one fifth question remained unasked, but is addressed in Fig. 9a: what happens if we sum two different waveforms? In this case, we summed a sinewave with a sawtooth. As one might expect, we do not see perfect symmetry, and we certainly do not see complete amplitude cancellations. Complete amplitude cancelations would not occur first because the sinusoidal has amplitude 1 while the amplitude of the fundamental of the saw is ($2/\pi$), and secondly, because at the minimum of the fundamental, all the rest of the harmonics in the sawtooth remain. A mixture such as this is likely not at all uncommon in everyday use of analog synthesizers.

SUMMARY:

When we add together two signals, the spectrum of the sum is never anything other than the sum of the spectra (linearity). If the frequencies of signals added together are close, or close to a small integer ratio, a phenomenon of "beating" may occur. The beating may be an annoying large variation in amplitude, called first-order beating, in the case of close frequencies. This occurs with sine waves and similar (such as triangles) that do not involve sudden jumps. With other waveshapes (square, pulse, saw) these do not occur for a variety of reasons (harmonics that do not cancel and/or amplitude nulls that are very short in time). In the case of ratios of small integers, (like 2:1, 3:2. etc., not 1:1) the beating is second-order or subjective: a feeling that something periodic is occurring. The "beat rate" is the difference frequency, but no spectral energy at that difference frequency is present.