Voice Based Affinity

A Recipe for Auditory Cheesecake

Joel Edelman

This document is Chapter 9 from the essay-in-progress:

My Music, Your Music;
A Technical Framework for the Musical Experience
Revision 28

October 25, 2022
9. Voice Based Affinity

So why, we may wonder, are harmonics so central to our enjoyment of music? Why are we not enamored of electronic music with raw, simple tones, or beating drums with vastly non-harmonic overtones? What is it about some combinations of tones that sets our ears on edge (synchronizes our cortical neuron firing rates), that sounds awful or leaves us unsatisfied? Let's start at Day One or, actually, a bit earlier.

Hearing is the first of our senses to become functional in utero. The mind of a fetus at this time is an unimaginably blank slate. Combined with a genetic hunger to learn, this is a perfect setup for a strong and lasting imprint of incoming aural content. There is ample and increasing evidence that the fetus hears, responds to, and remembers the sounds that reach it during the latter portion of pregnancy. "...ear development is primary in a fetus and sound is its first real sensory input. Already at thirty-eight weeks, the fetus reacts with limb motion to the sounds of music." (Donahue p.213).

The human voice, especially that of the mother, is a major contributor to the early hearing experience, and the association of this sound with the comfort and security of the womb creates a predisposition to favor such sounds after birth; this recognition, or affinity, supports our ability to identify generic creatures and specific individuals who are relevant to our survival. "Vocal properties that are available prenatally are sufficient to support newborns' preference for the maternal voice, and they suggest that fundamental frequency is important for neonatal voice recognition." (Spence and Freeman).

Every voice has a somewhat different overtone profile, and of especial importance is that there is an enormous complement of overtones available; this provides individuality in a very large population. But the overtones in every voice are always harmonic overtones, and thus the primitive ear is uniformly predisposed to recognize and prefer sounds with a harmonic structure or texture. Such favor would extend to those musical instruments which have linear resonances and, thus, harmonic overtones. It's also noteworthy that the pre- and post-natal exposure of the ear results in an ongoing evolution of predisposition for "local" sounds - tonal and rhythmic - supporting the cultural contribution to our musical sensitivities and preferences.

Prenatal exposure to the mother's voice does carry over to a preference for that specific voice; hence the exposure to timbre is not unnoticed. The predilection for harmonic tones in general evolves in the same way that comfort with less familiar voices develops: by the acceptance of partial familiarity in the composition of a tonal sound. This ability to expand one's comfort with deviations from the exact profile of the mother's voice is an absolutely essential capability to permit acceptance of other human voices, and this propensity, unstoppably, also allows comfort with the timbre of musical instruments and, of course, tones.

External sounds are, of course, subject to high frequency damping by the mother's flesh and clothing. "...memory storage of acoustical experiences and associations may start ...in the late intrauterine stage, with the fetus perceiving the mother's heartbeat, voice, and gastric sounds, as well as external sounds, filtered through the amneotic fluid." (Roederer, P. 165). Because prenatal experience is thus dominantly lower in frequency it is consistent that lower voices would continue to be experienced as more comforting than high pitched voices.
"Studies in pregnant humans leave little doubt of the existence of a varied foetal sound environment, heavily dominated by mother's voice and other internal noises... By the 24th week of intrauterine life, the cochlea and peripheral sensory end organs have reached their normal development. ...From 60 to 250Hz ...foetal sound isolation is less than 20dB. For frequencies of 500Hz and above, foetal sound isolation is quite high, exceeding 40dB." (Abrams pp 83-84, 99).

It has been observed that whispering, which is absent the fundamental and lower harmonics of a mother's voice, does not stimulate any sign of preference in newborn infants. I also remind the reader of an earlier observation that the dominance of the lower harmonics explains our tolerance for scales constructed of sequential second overtones and not the higher pure harmonics of a single tone.

The range of stimuli that newborns can recognize, based evidently on what they have heard in utero, is remarkable and will probably continue to be seen to expand with further research. "Interpretation of acoustic and linguistic information on intrauterine recordings suggests that the prosodic features of speech (pitch contours, rhythm, and stress) are available to the fetus" (Moon and Fifer) "Newborns can recognize the sound of a familiar voice, a story, or a melody. It appears that not only do newborns respond preferentially to specific voices, but also to more speaker-general properties of speech (e.g., intonation patterns characteristic of their native language)" (Moon, Cooper, and Fifer). Much of this recognition is attributable to their familiarity with the tonal structure of the voices they have heard, specifically the harmonic profiles of those voices, and this is the key to the importance of the harmonic series and the presence of those overtones in the human voice.

As a newborn, an infant normally continues to gather supportive exposure to the human voice. Beyond the mother, the infant will begin to recognize the father, other family members, the family doctor, and so on. Thus we recognize the human voice at large as a sound associated with a positive or secure environment, and the event of such recognition is a prime candidate for the release of dopamine associated with the reward response (Zatorre). Musical instruments might initially be discomforting to the newborn but with repeated exposure would become, perhaps, first a curiosity and then a pleasure as the differences cease to threaten and the similarities to voice dominate and are embraced.

This "mimicking" is well-recognized but persistently vague; some elaboration and clarification is in order. To begin, it would be only fair to ask how the timbre of a human voice, constantly sliding in pitch and varying in loudness, can be compared to the relatively steady sound of a musical tone. "...music is differentiated entirely and essentially from sing-song speech by its need of fixed steps, whereas speech ...knows no fixed intervals and in many cases occurs in the form of continually gliding pitch movement." (Stumpf P. 38). In his view " all documented music displayed organization and variety and, first and foremost, all of the samples of music known by then used distinct pitches." (Kursell, P.335).

To the human ear the harmonic mix of overtones and other components in the human voice, e.g. the "texture", is invariant in the pitch and volume domains, suggesting that the ear is capable of responding to what might be called "momentary quality" or a "running" structure (or "texture"). The temporal persistence of this structure allows us to internalize the characteristics of a voice that we need to be able to recognize, and to recognize those characteristics in motion, e.g. dynamically, as well as statically in tone form.
Melody would be derived from human speech, as prosody confined to discrete steps; the best and more favored steps have been those which fell into a harmonic pattern. The use of the voice for a sustained tone certainly was the result of imitation, the attempt to replicate what was heard in nature; the mockingbird does this to this day. It's very possible that the mockingbird's drive to replicate is related to our singing of tunes we have heard and liked, as well as those notorious earworms. Thus the natural occurrence of sustained tones combined with the overtone-rich wealth of the human voice have jointly contributed to the development of the specific musical vocabulary that we now have.

Harmony is known to have developed later than melody and has been documented to follow the early scales, most of which were based on lower harmonics or the circle of fifths (Benson P. 203). Primitive harmonies appeared around 800 C.E. followed by polyphonic music in the 11th century and the acceptance of the major third circa 1250.

It has been argued that language (and thinking) strongly influence our sense of rhythm, and similarly our preference for certain harmonic idioms would follow this influence. To the extent language affects our music locally, prenatal language exposure in and of itself would predispose a newborn to develop affinity for the music of the local culture, even absent musical exposure specifically.

"...training in a foreign music will remain necessarily limited if it is not matched with a good knowledge of its context, and possibly its language or one of its operative languages. The hazards ...are real and numerous, particularly due to differences in ways of thinking" (Aubert pp 69-70).

**The Invariants**

Let me identify and elaborate on those specific invariant qualities, a.k.a. the "temporal invariants". I will be giving special attention to the representation of sounds not as a waveform over time but as a diagram of the overtones, or spectral content, of the sounds. I do this because looking at the spectral content diagrams is similar to looking at an x-ray picture of the human body; it allows us to see the internal structure and gives us additional clarity. Looking at tones in the language of spectra vs waveforms gives us perspective in the same way that using arabic numerals instead of using roman numerals does.

Now let's look at some charts of a model of the human voice, shown below (parts of this graphic have been used earlier, in "Tone Quality"). The two top figures show a cross section of the vocal tract, for the expression of two different vowels. This cross section is through the center of the head, vertically, and midway between the two ears. The next two figures show the natural resonances of the vocal tract as shaped for that vowel; the tall peaks are the resonances or formant of the tract. The third pair of figures show the resonances of the voice, resulting from the harmonic vibration of the vocal cords passing through the resonance chamber of the vocal tract. The predominant resonances are labeled and these are regarded as factors in the expression of the vowel of interest; the relative heights of the harmonics determine the timbre of the voice at that instant. The last pair of figures are the waveforms of the actual sounds.
The third pair of figures is of specific interest; they represent the spectral structure of the sound of the voice. Of note is that the harmonic structure is different for the two vowels and the timbre associated with them. Thus, as time passes, the spectral structure will change from one shape to another as the vowels change and consonants are introduced. Changes in vocal pitch and volume will also be reflected in these figures. Thus the spectrum will be continuously changing shape during speech, almost flickering like the mottled sunlight under a tree on a breezy day.

Let's now look at the figures below.
This first figure, above, is a constructed sample voice spectrum for purposes of illustration; it is an instantaneous snapshot. The harmonic profile does not show a steady decline in volume with increasing pitch; this is characteristic of the timbre of a voice and its expression of vowels.

Next, the above figure represents the spectrum for a louder voice, at a higher pitch; the timbre also shows some variation; this would be a snapshot, for example, at a different instant from the first figure.

Looking at the above two graphs we can identify four properties of the human voice spectrum which are persistent, evidently, at most instants in time. They may seem obvious but they are quite important.

1. The total number of overtones remains substantial. There may be a possible loss of one or two low-volume harmonics at the top of the hearing range.

2. The evenness of distribution of the overtones is unchanged. The distribution of overtones does not have to be harmonic, only "even". That is, they may be in the proportions 1.1, 2.2, 3.3, etc.

3. The presence of an unrelenting and irregular attenuation of volume (fade) with increase in frequency remains in place.

4. The pattern of fade of overtones with increasing frequency will normally include some variation. This is generally associated with timbre, but it is important to note that a voice pronouncing one vowel is easily recognized as the same voice if it were to pronounce another vowel.

These four properties are visible in the simple model I have constructed, a model similar to that shown in the diagram of the human voice, above. But, like most models, it is incomplete. It does not display either of the following, both of which are known and significant aspects of all sound:
5. The non-zero width of each of the harmonic spikes or whiskers.

6. The non-zero noise visible inbetween each pair of harmonic spikes.

All six of these properties are ubiquitous in tones produced in the physical world, and even the last four are invariant, not so much in their instantaneous qualities but in their persistence and continuity in both dynamic and sustained tones. Every voice and every instrument has a recognizable sound despite the range in which the last four properties may vary, and within this limitation they can be viewed as invariant.

These six invariants comprise the "harmonic texture" of a tone. They provide us with the critical link between prosody in speech and tones in music. Because of their invariance and their persistence, they are the link between our affinity for a familiar voice and our affinity for harmonious tones, by voice or by instrument, and, further, for the scale of tones based on the harmonic series. This extended affinity has been observed from birth and is carried with us throughout our lives.

_Elaboration on the Invariants_

The above list of the invariant properties suggests to me a number of factors that need to be considered in any discussion of consonance or dissonance, and I will elaborate on these now.

_...Quantitative Density_

Consonance for the human ear seems to lie partially in the total number of tones experienced, including all overtones of all fundamental tones. A low value of consonance may be reported if there are too few tones; for example, there may be no overtones from the source (e.g. a tuning fork or an electronic signal generator) and thus there is a total of one tone experienced. Although not unpleasant the sound of a tuning fork is not engaging in any way; consider how rare the tuning fork is in ensemble instrumentation.

It follows that the concept of two pure tones generating either consonance or dissonance is meaningless. To the human ear such spare sound carries little musical content, positive or negative. The observation of beats or roughness from two nearby pure tones is no more than a cold observation of a specific quality and, in the absence of context, has no aesthetic value. This explains the ambiguities in consonance and dissonance research using pure tones and reinforces the earlier discussion of the weakness of such experiments; such data is being gathered in an aural region with an extremely low consonance signal-to-noise ratio, and reliable data in those circumstances is elusive.

An instrument tone typically comes with a full complement of harmonics and is widely experienced as consonant. If two or more instrument tones occur at frequency values belonging to the same root scale (e.g. a fifth or a third), then they have numerous harmonics in common and the total number of distinct tones experienced is thus increased by only a modest amount. The overtone texture in this instance has been strengthened or enriched without becoming cluttered or muddied and may be more highly valued for its consonance, often even more so than a single instrument tone. I will elaborate on this in the next section.
Now consider the highest of the notes that we use in our music. We know that higher tones contain fewer overtones than lower tones, partly because they originate closer to the upper limit of the human hearing range but also because the spacing between their overtones is wider. Further, they reside in a frequency region where their sound energies are more vulnerable to damping and obliteration by the environment, the instrument itself, and our ears. In this way these tones violate the consonance requirement of "quantitative density" and, consistent with experience, are generally less consonant than lower tones and often fit well into labels such as "tinny" or "shrill".

The inclusion of too many tones or overtones reduces the aesthetic value of a sound because it increases the harmonic density, reduces our ability to identify the primary tone(s), and thus has less association with a human voice. It shifts the sound in the direction of noise. The noise from an excess of tones and harmonics results in auditory chaos, or acoustic anarchy. This experience can be suggestive of danger or, even more likely, simply uncomfortable because the ear is not able to recognize a familiar texture and identify a familiar source. Musically there will be a point beyond which the ear ceases expecting a tone experience and accepts and assesses the chaos on its own terms. Thus noise can be a less dissonant experience than mistuning!

...Regularity

The consonance embedded in a tone with multiple overtones is largely the result of the regularity, or equal spacing, between the frequencies of the overtones. This may be the dominant factor in consonance; I don't know, but it clearly correlates with the regularity of the overtones in the human voice.

Note that there is no explicit requirement that the texture be comprised of harmonic overtones, only that the spacing between overtones be constant. Scales have been constructed digitally with arbitrary overtones -- typically stretched or squeezed harmonics -- and the result is experienced as pleasing, or consonant (Benson 2007 153-4). Auditory demonstrations of these scales are available on CD. (Houtsma et al, tracks 58–61). It is unclear at present how far from harmonic or how irregular a fabricated texture can be and still sound consonant. It is also not clear how significant the regularity of the overtone distribution must be for consonance.

However, the sound of a drum, with very irregularly distributed overtones, is generally never viewed as consonant, and many tone percussion instruments are strident and used for contrast or emphasis as opposed to harmony; thus there must be an upper limit, somewhere, to the overtone chaos that is consistent with consonance.

A tone with artificial harmonics (stretched, e.g.) may be experienced as a rich or fused tone, but if the harmonics are stretched too far the tone and harmonics will be experienced as a set of individual pure tones. The ear has only so much forgiveness. The fact that such an artificial tone may be heard as a fused tone does not mean that it is a harmonically rich tone. Each component is a pure tone and there are no Fourier components to any of the tones.

Often there may be scattered overtones which are reduced in amplitude or omitted as a function of timbre, but this does not seem to affect the consonance of the sound. Many open-ended wind instruments produce only the odd harmonics, and this is a dramatic decrease in density. Nevertheless the sounds of these instruments are normally highly consonant; this would be the result of the tight regularity of the remaining harmonics.
Every chord has its own level of consonance which is dependent on how thoroughly its tones complement each other harmonically. Thus chords with minor seconds, for example, would receive a lower rating than one containing a perfect fifth. In the figure below, the introduction of the minor second disrupts the regularity of the overtone pattern. In contrast the introduction of the perfect fifth maintains good regularity. Note that the absence of some harmonic elements with the perfect fifth has the effect of creating timbre, not dissonance.

The disruption of the regularity of overtones decreases the resemblance of a sound to a harmonic tone and, significantly, to the harmonic human voice as well. It also follows that a reasonable overtone density is required of any sound to be consonant, since a sufficient number of overtones is needed to establish a regularity of pattern that the ear can recognize.

...Fade and Timbre

Most naturally produced tones, including the human voice, a plucked string, or a sustained tone, produce an overtone series which for the most part decreases in amplitude or volume as the frequency rises. Thus the first few harmonics in a musical tone might be somewhat similar to the fundamental tone in amplitude but the eighth, ninth, tenth and so on would be significantly diminished in strength, and at some point the overtones would become almost inaudible. This quality is an important part of the experience of consonance.

If we examine the harmonic series for an electronically created tone, such as a simple square wave or triangular wave, we will observe a series of overtones which seem to meet all of our requirements for consonance: an infinite collection of overtones, an absolutely perfect even distribution, and a smooth, steady attenuation in the amplitudes of the harmonics with increasing frequency. Yet, these tones are anything but consonant, or pleasing to the ear in any way, especially as simultaneous sounds. They are harsh, metallic, and abrasive, and no one would imagine that they were produced by a traditional physical instrument. This has, in fact,
been one of the greatest challenges to the electronic music genre over the decades: to replicate in some way the qualities of consonance that are associated with physical instruments.

How bad are the simple tones? Listen for yourself:

https://en.wikipedia.org/wiki/Square_wave  
https://en.wikipedia.org/wiki/Triangle_wave  
https://en.wikipedia.org/wiki/Sawtooth_wave

Below is a figure showing the waveforms and spectra for three instruments, as well as a tuning fork. The latter has, of course, only one significant harmonic, and the remaining signal is from very small higher resonances plus unavoidable noise resulting from the digitization process by which this graph was produced. The other three instruments, however, have much more interesting spectra, accompanied again by some systemic noise but also some mechanical noise from the instrument.

Note that data which is smaller in magnitude is actually much, much smaller because the data is presented on a logarithmic scale and thus, for example, data which appears to be half as large may actually be only one one-hundredth as large. But also be reminded that hearing, like vision, responds to logarithmically-scaled stimuli.

The most relevant property of these three instrument spectra is that the decline in overtone amplitudes with increasing frequency, despite its persistence, is irregular; that is, the decline is not a steady and smooth decline but rather moves up and down unpredictably. Remarkably some of the overtones are actually louder than the fundamental.

There are two reasons why we hear the fundamental as dominant even though some of the harmonics are louder. One is the propensity of the ear to construct a tone experience based on the spread of harmonics (Section 2, "Missing Fundamental"), and the other (possibly related reason) is the cumulative effect of difference tones (see "Notes on the Human Ear," below). "The fact that we hear the fundamental tone (the first harmonic) as loudest is a
function of our brain which creates and adds the amplitudes of the difference tones to the actually sounded harmonics. Because the difference between frequencies of the adjacent harmonics always is equal to the frequency of the fundamental tone, the insufficient amplitude of the first harmonic is compensated by the difference tones of all adjacent pairs of the Harmonic Series.” (Sergei)

Timbre - the irregularity in the fade of the harmonics - is a card-carrying player in the game of consonance, despite that it is not mathematically invariant in the experience of the human voice. Its irregularity in a voice is generically consistent and over short segments of time recognizably stable. Note that we are all able to identify a familiar voice despite its roaming timbre. And most instruments have different timbres in different ranges, and often variations in timbre depending on how the performer manipulates them; yet we can easily recognize the instrument.

We had observed earlier (in Section 7) that the irregularity of the overtone amplitudes is one of the factors in the quality of a sound that allows us to identify the instrument (Section 7, Tone Recognition). It is, thus evidently, also one of the invariant factors in the quality of a sound that makes it consonant to our ears. The absolutely smooth and regular attenuation of overtones from the basic digital tones diminishes the experience of consonance as much as, or more than, the absolute absence of overtones altogether, e.g. the pure sine wave. Note that, on a logarithmic scale, the smooth attenuation of the harmonics from a basic synthesizer tone form a straight line at their peaks (see figure below); the dissonance of such tones is thus easily defined mathematically. I also speculate that no voice or physical instrument is capable of producing the smooth attenuation of the simplest electronic tone.

![Logarithmic graph of harmonics for a triangular wave displays linear attenuation (Hoffman).](https://www.blackghostaudio.com/blog/the-beginners-guide-to-audio-synthesis)

Why is this the case? Why does that special smoothly attenuated "timbre" not produce a pleasant sound? It is much like a cartoon image in which, for example, a face is a uniform color and texture and does not have any irregularities of features or any shadows, least of all any subtle variations in the color itself. It is like a cartoon landscape as opposed to a work of art. Similarly many viewers are actually bothered by the artificial qualities of "colorized" black-and-white films and often would rather see the original unaltered version of the film. So we demand an "organic" experience visually and aurally. We expect and prefer some level of irregularity or imperfection in our experiences, certainly because that is the nature of the
physical world we live in. The instances of undisturbed perfection and linearity in the natural environment are vanishingly small. We are not easily fooled and can recognize quickly the results of a mechanized process; we require faithfulness to natural phenomena, a certain level of chaos in the patterns around us.

Modern synthesizers are becoming quite good at producing consonant sounds, just as cartoonists have at times been very good at producing artistic work (consider Walt Kelly or Bill Watterson). For us this effort to make synthesized sounds consonant is a window into the content of consonance.

Fade or attenuation likely has a pronounced but indirect role in that it is a factor in the tendency of the ear toward fusion of tone sets, discussed in Section 2. Our ability to fuse multiple tones into a single-source experience supports our experience of consonance by replicating the prenatal aural environment which is largely influenced by voice.

The presence of a pronounced harmonic component at an unusually high pitch is likely to be attributed by the ear to a separate aural source, and the experience of consonance would now be determined by the parameters of the ensemble experience, discussed below.

...Fat Spikes

Note in the above figure, and several others earlier in this essay, that the harmonic frequencies of physical instruments do not appear as a single line, or a whisker or spike, as they do in mathematical models. They are fat. This is in part because of the process of generating the graph with computers, but it is also in part because of the nature of the tones from which these graphs have been generated. Instrument tones are, simply, never absolutely constant and unvarying (the exception, of course, being "now", which is a time period of zero duration). It is the nature of the physical reality of instruments. And the graphs can only be generated by taking a sample of the sound over a nonzero period of time, several seconds at least, and more if greater accuracy is desired. During this time the tone will vary, if only minutely. For example, a violin tone will vary because of slight changes in bow pressure, bowing speed, bow angle, and even the varying amounts of rosin on different areas of the bow. A clarinet tone will vary because of changes in air pressure and speed due to the emptying of the performer's lungs, or the tiring of the lips and cheeks. A plucked or hammered string has a natural variance in upper harmonics due to the change of tension in the string as it swings between the extreme positions of the fundamental tone. "When a string vibrates, it stretches slightly at each back and forth cycle of the fundamental frequency. So when a string is at a far excursion for its fundamental frequency, the pitch of each harmonic rises due to the increased string tension. This causes the waveform to "roll around" over time" (Winer P. 359).

Thus the spikes in the graphs for the overtones are not gratuitously spread out, but rather each overtone does have some systemic variation independent of the computational mathematics in use. In this way we see that the tones produced by a synthesizer must also be gently "ruined" although not in an arbitrary manner.

"One important way to make a synthesizer sound more interesting is to use two or more oscillators playing in unison but set slightly out of tune with each other. This is not unlike the difference between a solo violin and a violin section in an orchestra. No two violin players are ever perfectly in tune, and the slight differences in pitch and timing enhance the sound" (Winer P. 350).
...Noise

The last invariant feature of interest is low-level noise, evidenced in the graphs by the fact that the spectral energy curve is never zero, at any frequency. I mentioned this much earlier in discussing tone identification and observed that the flute displays a substantial amount of noise due to the steady wind across its mouth. Similarly string instrument tones are accompanied by the rasping of the bow across the string.

Our ear apparently wants to hear some of the nonharmonic noise that comes from a physical instrument in the process of creating a tone. We want to hear the whooshing air of a flute, the rasping slide of a violin bow, and as much as possible the unfathomably complex sounds at the onset of an instrument tone. This background noise in many cases may be an essential contributor to our preference for certain sounds. White noise is often a comforting sound all by itself.

The complex onset sounds, for all they contribute to tone identification, have no role that I can identify in the experience of consonance. Neither of these two parts of a tone contribute to our affinity for harmonics or the experience of consonance; these are amply stimulated by the earlier invariants. But they appear to have a notable role in our affection for certain tonal sounds, and disaffection for others.

Recent work at the Indian Institute of Technology was focused on a parametric model for synthesized violin tones, "in particular the generative modeling of the residual bow noise to make for more natural tone quality. ...the residual is in essence that part of the audio signal that cannot be represented by a sum of harmonic partials. ...Examples in musical instruments involve the breathy sound when playing the flute and the scratchy sound the bow makes when it moves against the violin string during note sustain regions" (Subramani).

As much as there has been an effort to match the consonance of physical instruments with synthesizers, there has also been a movement to view the synthesizer as an instrument in its own right. It has been used often in recent years (the square wave was used in Keith Emerson's "Lucky Man" circa 1970) for its specific qualities and not as a replacement for another instrument. Synthesizers have also been incorporating features which allow the introduction of expressiveness in a performance, not unlike the vibrato or rubato widely used in traditional performances. My only concern in this essay is the quality or consonance of tones, not embedded in any context.

An often-asked question is whether a synthesizer will ever produce a sound better than an instrument. The answer is, yes, of course they will, but only after extended vetting similar to the evolution physical instruments have gone through. It has not been a short process by which our favorite instruments have reached their modern configuration and sound.

"Physical modeling uses complex equations that mimic the sound sources, body resonances, and other attributes of acoustic instruments. ...physical modeling more realistically creates the sounds that real instruments make in between successive notes. When a trumpet or clarinet plays a legato passage, the player's breath continues uninterrupted for the duration of the passage, even as the notes change. ...Real violinists often play several notes in a row without reversing the bow direction, and they sometimes slide from one note to another note" (Winer P. 373).
The second half of the answer to that leading often-asked question is whether synthesized music will, in the end, replace live performances. I expect that an imperfect live performance will routinely be preferred to an absolutely perfect automated performance. Would we forever give up the concert hall and just stay home and listen to a recording? How much of our lives will we want to spend between headphones?

**Consonance and Dissonance**

The search for clear definitions of consonance and dissonance is age-old and, from the beginning, has been based on the stand-alone properties of sound waves: frequency ratios, beats, and harmonicity (the degree to which a sound contains harmonic overtones). The measure of consonance and dissonance has always been a matter of individual response and opinion but has never been connected to the individual's experience.

What I have done in this essay is describe the relationship of present sounds to the earliest human exposure to sounds, specifically the maternal voice in utero, and the deep imprint those sounds left in our minds. My considered assertion now is that the **consonance of a sound is best defined by the degree to which it faithfully mimics the qualities of that maternal voice**, in particular as expressed by the parameters of spectral or harmonic texture discussed above. The failure of a sound in any of those parameters contributes to its dissonance. Thus consonance on its surface is a measure of similarity or familiarity. Briefly, you prefer certain musical sounds because they are reminiscent of your Mom's voice.

The extension of harmonic texture from sliding sounds to tones (and subsequently to scales) is the result of our genetic predisposition to accept and embrace sounds which vary - though moderately at every step - from the original voice imprint and to expand our internal library of *nice* sounds. Because of our retention, or working memory, we experience the harmonies suggested by melody, and our initial choices will thus be the melodies that are built on the harmonic structure of that familiar maternal voice. This answers the perennial question of "nature". Conveniently this definition also accommodates the impact of the unique music and language (and thought) of a culture on an individual's internal library and working definition of consonance, assembled both before and after birth. This answers the perennial question of "nurture".

* see Section 1, *Tone*, for a concise definition of "nice".

This approach to consonance does not contradict or invalidate earlier ideas and work. Rather it envelops them in a way that accommodates their shortcomings. Pairs of tones with low ratios of pitch or frequency are still considered consonant because their combined distribution of harmonics preserves regularity. Some lesser pairs of tones with mathematically compromised ratios have been accepted into the club, but pairs of tones with hideously large ratios disrupt the regularity and do not dance well together. This was illustrated in the chart in the paragraph "Regularity", several pages back.

The presence of beats and roughness in pairs of pure tones is also consistent with the failure to address the invariants. Specifically, they do not have enough overtone and background content to be consonant in any configuration; the supposed consonant points,
generally at unison and one octave, are points at which the sound is less awful and not sounds one would relish in another circumstance. As noted in Section 8, dissonance from beats is being measured in a very low signal realm of sound quality, and any effect of perception will generate an outsized bit of information.

It's too limiting to define consonance as merely the absence of beats. This approach invites the most crude of electronically generated signals to be assessed as consonant, and they are not. A sound profile evidently must include noise and fat spikes to produce sounds that would meet any definition of "nice".

Early thought about consonance is well viewed similar to the parable of the blind men and the elephant. The elephant did indeed have large legs, a tail, and an amazing trunk. But what about the elephant itself? And as I alluded at the beginning of this section, what about the forest?

The ultimate and common failure of prior studies of consonance is that they assume the basis is in the form of a silver bullet - numbers, beats, or harmonics - and it is not. There is no silver bullet. **Consonance is an experience that results from the simultaneous presence of multiple qualities** in a proper arrangement and in proper emphasis. The requirements for consonance are thus as complex as the tones themselves.

The list of these qualities - the invariants - can be compared to the list of ingredients for, say, a cheesecake. If the ingredients are all included in appropriate amounts, and the procedures for combining them are followed, the resulting dessert will be rewardingly delicious. But if the quantities are varied too much from the recipe, the result may be awful. Think about omitting the sugar, or dramatically increasing the salt. Even an excess of vanilla can render the cheesecake distasteful.

There is a recipe for consonance. The invariants are the ingredients. But what of the quantities? This is a difficult question; the preferred quantities are as varied as our fingerprints and, worse, vary for each individual as we grow and age. But this is not an impossible situation. After all, a cheesecake made (carefully) from any of thousands of recipes is still a delicious cheesecake and, further, can easily be distinguished from a pound cake or an angel food cake.

**Ensemble**

One might ask why the sound of an ensemble -- with chords and harmonies -- would appeal to the ear more than a solitary harmonic tone, presumably the ideal consonant sound (and often close to that). This effect is never predicted by consideration of frequency ratios, beats, or harmonicity. It is not even suggested explicitly by my idea of voice based affinity. But it is a significant "why" that needs to be answered.

First, let's consider again what voices might sound like in utero. Because of the damping effects of the womb, the lower harmonics will come through loud and clear but the higher harmonics will suffer from some damping, more as one goes higher. Thus the experience of the fetus' ear is of an emphasis on lower overtones, though not a lower fundamental pitch.
Now, I note that in an ensemble sound the total number of tones and overtones is substantially increased but the regularity of spacing of the tones and overtones is maintained, especially for the chords of a functional cadence. This is illustrated using the figure above, Constructed Sample Solo Tone, in the figure below, Constructed Major Triad, for the root tone A2. In the figure below there is substantial irregularity introduced in the spectrum magnitudes; as discussed above this could significantly increase consonance, even if the original solo tone were not handicapped with the dissonance of too much regularity of fade.
In this last graph we can now see that the simultaneous playing of multiple instruments (on appropriate tones) produces greater volume for the fundamental and the lower harmonics which often match, but less so for the higher harmonics where overlap is less common. The result mimics the prenatal dominance of lower harmonics.

Thus, we have four factors each of which would lead one to expect an ensemble sound to be more consonant than a single harmonic tone: first, that there are simply a greater number of overtones, though not an excessive number, second, that the quality of regularity in the spread of frequencies is maintained, third, that the depth of timbre seems (admittedly by example only) to have been increased, and, fourth, that the lower tones and overtones are relatively stronger replicating the prenatal experience.

In the context of synthesized tones one may wonder whether the sound of the above triad would sound consonant, even though each of the component tones displays all the characteristics of a simple synthesized tone: a steady, smooth attenuation of harmonics with no harmonic timbre. The triad itself shows a great deal of harmonic timbre and thus very well might sound more pleasant to the ear because of this timbre; but the sound will nevertheless seem artificial because it is lacking the minute spectral spread in its harmonic peaks as well as any background noise.

Suspension and Resolution

The musical experience of a sequence of chords - e.g. a cadence - which moves from one level of consonance to a higher level correlates with the experience which we describe as resolution; all other motion of harmonic statements -- either down in level or static -- are experiences of suspension (the sustained tonic generally excepted). It is thus the increase in harmonic consonance level that drives our fondness for cadences which end in the tonic, particularly cadences in which the final steps are of moderate size. This is the lifeforce behind functional harmony and the reason for its persistence as the Lingua Franca of harmony.

The force behind this effect is analogous to the effect observed throughout physics in which materials or particles are always driven to a state which is stable only because of the additional energy required to dislodge them. e.g. they come to rest in a low energy valley or trap.
Musically, motion away from the tonic produces tension and requires energy. The highest value of consonance, then, is associated with the lowest value of tension or suspension.

Some Further Musical Considerations

1. The voice needs less energy to slide in speech than to sustain a steady tone or, especially, to move between two held tones. For example, a sigh naturally produces a slight downward shift, a question or a demand has its own prosody which reflects the existing emotional dynamic, rarely if ever a steady tone. In contrast, mechanical instruments require input, e.g. additional activity or energy, to produce anything other than a single tone at one frequency. This suggests that vocal tones originated in our mimicking of external sources, and the impetus for that mimicking would have come from recognition of the consonant harmonic texture. Furthermore (as long as I am speculating) our adoption of vocal tones and the evolution of singing would thus have far trailed the evolution of speech.
2. Extremely small mistuning between instruments is notoriously dissonant; this would be predicted in the framework since this particular disruption of regularity is essentially never balanced by overtones spreading away from their initial close proximity into a middle zone that even vaguely approaches the appearances of a regular distribution.

3. Harmonic texture or density, the material of voice-based affinity, is associated with activity in the right hemisphere. Voice based affinity emphatically is not based on the concept of prenatal security; it is a non-verbal non-symbolic reaction to recognition of a specific sound texture, a gut feeling so to speak, with pre-natal origins.

This mechanism is in play in adults in other ways. As an example, many folk dances are accompanied by tunes with a straightforward AABB structure, and often the ending cadence and melody are repeated in both the A and B parts. This is a non-verbal, almost subliminal, device that signals the dancers that they are at the end of a figure. It is useful for novice dancers who are more occupied with the details of the figures and for experienced dancers who can keep "tempo" with the flow of the figures more effortlessly.

4. Our response to rhythm has the same origin as our response to harmonic texture, the prenatal sensory experience. In this case the rhythmic experience would come from the skin as well as the ear, and it is possible, even likely, that the skin precedes the ear by some time as the first sensory input organ to become active.

5. The concept of "harmonic sieve" has been proposed to help explain our response to harmonic overtones. The concept is that the "auditory system employs a harmonic template in order to assemble images" (Huron, P.29). Like the protention library it's a somewhat vague concept but it's consistent with the role of harmonic texture, including tolerance for some degree of mistuning. One can easily surmise that this sieve would be formed in the pre- and post-natal exposure to the human voice, if not to music itself.

6. One of the most-repeated comments about music that I've seen in my reading is a statement made by American linguist Steven Pinker. His view of music from an evolutionary standpoint is that music is no more than "auditory cheesecake." "Pinker believes that music flips certain switches in the brain, that have actually been installed for other purposes, especially in connection with our speech ability and control of our body movements" (Lehmann P 29).

This statement, and of course its concept, have ruffled quite a few feathers over time, and any number of academics have responded in protest, arguing one way or another that this idea must be flawed and has weaknesses here and there. It makes us uncomfortable, if only because it shifts music away from a realm in which it has high spiritual value. Thus perhaps there is less about humanity that separates us from lowly animals.

It is true that this idea dispenses with any aspect of spirituality in the origins of music, but this loss cannot be a driver in our thinking. Indeed, as my picture of consonance became clear I have found that my thinking lends support to his brash statement. Our sensitivity to harmonic sound has its evolutionary basis in voice recognition, and its extension to tonal experiences is fallout.

Lehmann asks, "Is it at all possible that a basically dispensable pastime could become such an integral part of all cultures, independent from each other, in different parts of the earth?" One
could ask the same question about recreational drugs, or any other recreational activity. And the answer, of course, is "yes".

**Deaf Parents**

There is some question thrown into this view by a recent study of hearing infants born to deaf parents. The infants were found to have the same response to consonance as infants born to hearing parents, suggesting that this sensitivity is genetic, not instilled by early prenatal experience, e.g. nature and not nurture. However, the fetus is expected to have been exposed to some vocalizations from the deaf mother, albeit less than many other circumstances. "While deaf mothers clearly do not vocalize as often or as clearly as hearing mothers, it is nevertheless likely that the mothers did vocalize while their fetuses were developing. As such, there may have been some prenatal exposure to the musical and/or speech structures that carry relevant harmonic information." (Matasaka, P.50).

Also it is not clear that the fetus was not exposed to voices from other sources such as neighbors, relatives, bus riders, shoppers, or doctors and nurses? Did the fetus hear Christmas music in the drug store, restaurant music, or music in elevators? Further, a paucity of early exposure would increase the influence of what little may have been experienced. We should also be cautious about the exciting and controversial studies suggesting that experience can be genetically captured and conveyed to the next generation. These issues are all clearly areas worth watching and warrant at least a small caution flag in my speculations.

"...there is also accumulating evidence indicating that prenatal exposure to and familiarization with salient auditory stimuli accounts for a young infant's auditory preference for the mother's voice, for a familiarized melody by voice, for the prosody (that is, melody, rhythm and dynamics) of a familiarized spoken text and for the mother's language." (Matasaka, P.46).