Pitch

There is perhaps no aspect of music more important than pitch. It is notoriously prescribed by composers and meaningfully recomposed by performers. It would be near impossible to move through a day without experiencing pitch, even if just in the call of a bird, the buzz of a fly’s wings, or the beep of an alarm clock. Yet what is pitch? Even non-musicians are frequently aware of the idea captured in the official definition: Pitch is, “…the auditory attribute of sound according to which sounds can be ordered on a scale from low to high.”¹ Many singers are aware that pitch has a relationship to frequency, usually measured in Hertz (Hz), or repetitions per second. In the voice pedagogy community, pitch is frequently differentiated by its perceptual characteristics. Frequency is objectively measurable; pitch is how the brain perceives that frequency.² This is relevant to phenomena as ubiquitous as vibrato, which humans may perceive as having a central pitch despite regular oscillations in frequency. For many, this is where the definition ends.

The official definition of timbre specifically controls for pitch, and the timbre of a complex sound can change independent of pitch. This may give the impression that pitch and timbre are completely separable; however, this relationship is complex and must be

² This is a common sentiment in the voice pedagogy literature, exemplified in Scott McCoy, Your Voice: An Inside View (Delaware: Inside View Press, 2012), 18, “Frequency is an objective measurement of vibrations per second; pitch, however, is a subjective perception that can be influenced by factors ranging from vibrato to timbre and intensity.”
qualified. As Plomp (2002) writes, “For the extreme case of a sinusoidal tone without harmonics, the tone’s frequency as its single variable determines pitch as well as timbre.” This echoes the long held assumption in the psychoacoustics literature that, “…simple tones have timbre.” The fundamental frequency of a sound determines the frequencies of the perceived harmonics, which in turn constrains the timbre that can be created by their interplay as parts of a complex sound. For example, play the highest and lowest keys on the piano. Note that the former creates a bright, /i/-like sound while the latter a rich, dark, and complex sound. The former can never elicit the timbre of the latter because of the frequency of the fundamental. The timbre of the harmonics generated by the lower fundamental does not exist within the sound produced by the high fundamental.

Locating this question within voice perception similarly challenges the independence of pitch and timbre. Changing pitch while keeping the vocal tract stable may not change the phoneme, but changes in timbre definitely take place. Bozeman both initially observed this phenomenon, and has also explored its pedagogical value in depth. He calls it *passive vowel modification*, or more recently, *passive vowel migration*. Pitch and timbre may be best thought of as semi-independent, semi-related aspects of sound. Certain timbres are only possible within a given pitch range. The converse is also true, that certain pitch ranges limit the potential timbre.

---

Generally speaking, pitch as a percept arises from the portion of the spectrum perceived as the lowest eight harmonics. From about the ninth harmonic and higher, the energy of a voice progressively turns into pitch-less, bright noise akin to the sound of a cricket. It is not the same as other forms of noise (e.g. air turbulence or white noise), but it is conspicuously unresolved from the pitch. This change is not sudden; crossing from harmonic eight to nine does not immediately introduce this noise. However, as Plomp (2002) notes, “…the unresolved higher harmonics may also contribute [to the pitch], but to a lesser extent.” These unresolved harmonics may be identified as a separate percept, occupying a perceptually bright space above and distinct from the pitch. Note that the pitch percept does not exist in the compression wave in the air; the closest correlate is the frequency of repetition of the waveform. Pitch arises after the processing of this wave by the inner ear.

Nothing on a spectrogram or power spectrum would necessarily indicate this. In fact, we are likely to assume the opposite, that the fundamental is the pitch and the higher harmonics the timbre. The fundamental most likely shares a frequency with the perceived pitch, and depending on the pitch and vowel, may also contribute a great deal of energy to the pitch. But the portion of the sound perceived as the lowest eight harmonics

---


7 In David Howard and Jamie Angus, Acoustics and Psychoacoustics (New York: Routledge, 2017), 137, the authors outline two competing theories for pitch perception: the place theory and the temporal theory. Neither theory completely explains pitch perception, but both locate the emergence of pitch between the inner ear and the brain.

8 A curious counterexample that the voice science community is currently investigating is the phenomenon of subharmonics. Jan G. Svec, Harm K. Schutte, and Donald G. Miller, “A Subharmonic
literally forms the bulk of the perceptual experience of the pitch, and this percep is more than the contribution of the fundamental. In almost all cases, the pitch percept of a voice may be thought of as the tone color of the fundamental plus the tone color contribution(s) of these higher harmonics. Remove the fundamental from a voice, and the pitch will ring through unaffected. The color of the sound will change, but not the pitch. These two parts of the pitch phenomenon are frequently separable perceptually, which is to say that in many cases the fundamental may be identified as a separate percept.

Again, this may seem like an overly complex distinction for the voice pedagogy community, especially when pitch and the frequency of the fundamental are tightly intertwined. However, we interact with this phenomenon, called the missing fundamental, every day. Analogue telephones regularly remove all information below about 300Hz (roughly equivalent to the pitch D4). Every phone conversation that ever happened over such a phone line removed the fundamental for a majority of people. However, the pitch of these voices do not suddenly jump an octave. This is why one may listen to music on the tiny speakers built into our smart phones. While doing so removes qualitative aspects of, for example, a bass guitar, it does not change the pitch of any of the instruments. Or, consider an unamplified male voice singing a very low pitch in a large space. Very little

Vibratory Pattern in Normal Vocal Folds,” *Journal of speech and hearing research* 39(1), 135-43. DOI: 10.1044/jshr.3901.135, explore semi-chaotic patterns of vocal fold vibrations that produce a weaker glottal source puff every other puff. This results in a perceived pitch an octave lower than the actual frequency of puffs in some cases. In others, the percept is of two pitches an octave apart sounding simultaneously.

of his fundamental will radiate to an audience member. Not only is his lowest vocal tract resonance unlikely to be lower in frequency than around 270Hz,\(^{10}\) his lowest frequency energy will be the least directional part of his sound.\(^{11}\) In daily life we frequently experience pitch in the absence of a strong fundamental. This means that the energy perceived as harmonics two through eight contributes a different aspect of the same pitch percept, one separable by qualitative characteristics. Another way to consider this is that a pitch will have multiple aspects that differ qualitatively.

I would like to explore this idea with an example of a real singer (see Figure 1). In this power spectrum image, note that the complex wave (see Figure 2) has been translated into an easier to read spectrum of harmonics by a fast Fourier transform. Neither of these images display the pitch as perceived. Figure 2 displays the frequency of repetition of the waveform pattern as a function of time (i.e. how many milliseconds long is each repetition of the pattern in the waveform). Figure 1 displays this same information in the frequency domain. In fact, this image suggests that only the fundamental (the harmonic at the frequency equivalent to Ab4) aligns with the pitch; all the other harmonics visually align with higher pitches on the piano keyboard. The image would better display the pitch if the lowest eight harmonics were grouped together, as shown in Figure 3. These harmonics are resolved into the pitch.

\(^{10}\) An average frequency for the lowest male vocal tract resonance as found in McCoy, Your Voice, 42.

Figure 1: Power spectrum view of a female singing Ab4 [a]. Each colored line represents a harmonic that corresponds to the frequency equivalents of the pitches on the keyboard. Source: Author.

Figure 2: Waveform of a female singing Ab4 [a] from Figure 1. Note the repeating pattern (about two and a half repetitions in this image). Source: Author.
Figure 3: Power spectrum view of female singing Ab4 [a] from Figure 1. The lowest eight harmonics have been bracketed together to show that they form the pitch percept. Note this percept arises after the processing of the inner ear. It does not exist in the compression wave in the air. Source: Author.

The Singer’s Formant Cluster and Pitch

The perceptual difference between resolved (pitch) and unresolved (not pitch) harmonics points to an interesting paradox in what is thought to be a commonly understood phenomenon. Since Sundberg (1974), the singing community has largely accepted the idea that the singer’s formant (later changed to singer’s formant cluster by Sundberg (2003)) is the reason that a classical male voice can be heard over an orchestra. In fact, this bright, ringing quality frequently stands in for the aesthetic of classical singing, listed along such genre defining technical choices as a lowered larynx.

---

In contrast, classical female singers (and singers who use amplification) are said to not need or use a singer’s formant, especially above the pitch C5.\(^\text{13}\)

This is demonstrated by means of filtered masking noise that imitates the theoretical long-term average spectrum of an orchestra. When applied to a voice free of the singer’s formant, the voice is masked. When applied to a voice with a singer’s formant, that portion of the singer’s spectrum noticeably occupies a relatively ‘orchestra-free’ frequency band.\(^\text{14}\) Setting aside the question of whether an orchestra ever produces a continuous sound equivalent to white noise, a few other issues with this model are worth exploring. These issues neither discard the importance of brightness in the male voice, nor the specific phenomenon (both perceptually and acoustically) that is the male singer’s formant cluster as described by Sundberg.\(^\text{15}\) However, I would like to suggest that our understanding and appreciation of brightness (specifically in the male classical voice, but also in all other singing voices) can be given much more detail and specificity.

The reason for this is eminently practical. At the recent 2018 NATS National Conference, attendees sitting around me at a presentation applied the term singer’s formant to any instance of brightness in the samples of a male voice. Sometimes the phenomenon they heard was the SFC. But at other times it was a pronounced harmonic lower than the SFC.\(^\text{16}\) And in some cases it was actually the contribution of harmonics higher than the SFC. If the SFC has a technical definition (a specific frequency band

\(^{13}\) McCoy, Your Voice, 49-50. \\
\(^{14}\) Sundberg, The Science of the Singing Voice, 123. \\
\(^{15}\) Sundberg, “Research on the Singing Voice in Retrospect,” 011–022. \\
\(^{16}\) As may be expected from male singers employing a third harmonic/second vocal tract resonance strategy as outlined in Donald Miller, Resonance in Singing (Princeton: Inside View Press, 2008), 4.
amplified by a specific mechanical action), we ought be able to attach a specific perceptual aspect of the sound to the specific phenomenon the term references. Otherwise it has no practical pedagogical value and we may as well revert to a general term like, *brightness*.

Here is a specific example to tease out this issue. Given a singer’s formant cluster (SFC) centered around 2800Hz, a tenor’s SFC on a high B♭4 will be composed of harmonics resolved into the pitch (see *Figure 4*). The same SFC of any pitch below about B3 (see *Figure 5*) will be perceived as unresolved. It is a bright and ringing sound, but an unresolved sound nonetheless. This means that the middle and lower range of most male voices elicits increasingly pitch-less sound in the singer’s formant band as fundamental lowers. Were the SFC truly the only part of the voice that carried over an orchestra, we would hear the voice switch between pitched and pitch-less sound based on \( f_o \) and the frequency range of the SFC. The fact that this is obviously not how the voice behaves punches a hole in the idea that the non-SFC spectrum of the classical male voice is significantly masked by orchestral instruments in practice.
As of writing this document, I am unaware of studies specifically exploring the pitch resolution of harmonics in the singing voice in this regard, especially at high fundamentals, and further research is certainly warranted. However, this phenomenon can be readily observed in the wild, as found in this sample of an A3 (Figure 6 top) and C#4
(bottom) sung by Dmitri Hvorostovsky. The SFC of the latter far more clearly resolves into the pitch than the former, which takes on a confusing, noisy quality.

Figure 6: Long term average spectrum of sample of A3 and C♯4 sung by Dmitri Hvorostovsky. As this is a live recording with orchestra, Hvorostovsky’s harmonics are indicated in each image. Note that the SFC for the A3 (top) begins at about the ninth harmonics and extends up to the thirteenth harmonic. The SFC for the C♯4 (bottom) ranges from the fifth through tenth harmonics. Source: Met Live HD Broadcast of Ernani
Beginning to Understand Pitch in Other Voices

Setting aside the more thoroughly studied classical male voice for a moment, we must consider the implications of pitch perception for the study of female classical and contemporary singers (regardless of gender). Most spectrographs default to a frequency display that shows no information above 4 to 5kHz. While this range makes sense in the context of the analogue telephone bandwidth, the spectrum of a singing voice frequently exceeds the upper frequency range of speech. Consider (see Figure 7) the same female singer from Figure 1 singing an A♭4 (above) and an A♭5 (below). When the spectrograph caps the visible frequency range at 5kHz, up to twelve harmonics could be visible for the pitch A♭4. However, only six harmonics are visible for the A♭5. This means that any unresolved harmonics are not captured on this graph.

I must reiterate that unresolved harmonics will contribute something perceptually different from resolved harmonics. This is a qualitative difference relevant to the sound of a singing voice. Perhaps there is no energy above 5kHz. Or perhaps any such energy is perceptually irrelevant; it almost certainly will not dramatically impact language cognition. But the artificial visual limit similarly constrains the imagination. This graph, while logically extrapolated from the study of speech, is inherently biased. It suggests that a female voice above the staff will obligatorily elicit a resolved sound; that the broad perceptual quality of unresolved brightness is available to only lower voices. The converse is also problematic. As the SFC model pushes singers to imagine that sub-SFC
energy makes no meaningful contribution to the sound, the harmonics above the fundamental in Figure 7 (bottom) contribute to both the sound and the pitch. One cannot assume that the quality of the classical female voice above the treble staff is captured by the sound of the fundamental, despite its obvious power.

Figure 7: The female singer from Figure 1 singing an A♭4 (above) and A♭5 (below). Note that when the frequency range (horizontal axis) is capped at 5kHz, only six harmonics are visible at the higher pitch.

This is the central challenge the perception focused, fact based voice pedagogy community faces: Linguists would suggest that we have captured enough of the spectrum
in Figure 7 to preserve language cognition. Such a model excludes aspects of the sound potentially critical to the aesthetics of the singing voice. Were this image the spectrum of a violin, we wouldn’t think to limit the view in this manner, or to suggest that such a limited image captures the essence of the instrument. In fact, raising the upper frequency limit of this image to 12kHz reveals an audible portion of the spectrum not previously seen (see Figure 8). It is fair to debate the perceptual relevance of this higher frequency energy, but we do not discuss what we are not aware of. As we move forward to chapters on auditory roughness, tone color, and the ways in which these temporally invariant aspects of timbre interact in predictable patterns to explain familiar phenomena, we will similarly push to reform models to better represent the way in which a human perceives sound.

Figure 8: The female singer from Figure 7 singing an A♭5. Note that an additional spectral peak is revealed when the frequency range (horizontal axis) is extended to 12kHz.