

The theoretical concepts in this paper were motivated by the author's observation that vocal harmony could result in rich and, to him, beautiful acoustic patterns that went beyond the harmony present when musical instruments blended their tones. Yet I had not read of any attempt to explain the richness of vocal harmony by mathematical modeling. I wrote the following paper in May, 2004, and passed it around to some colleagues who were more knowledgeable in the acoustics of the singing voice than I was, asking for some reference to similar (or alternate) explanations that I may have missed. To date, none have appeared in my email box, so I am posting this on my website in the hope that it might stimulate some research along the lines indicated by the theory

Martin Rothenberg.

The Role of Vowel Value in Vocal Harmony

Martin Rothenberg

Rev. 5/27/04. Edited for posting 7/9/06

Background

It is generally understood that in many types of singing relatively small variations in vowel articulation can have a strong role in determining the quality of a sung note. For example, some operatic-style singers report using a small amount of nasalization to modify their voice quality on some notes. (This is probably more common in American country-western singing.) Some bass or baritone singers can also enrich voice quality from that of a normal spoken vowel by modifying vowel articulation so as to add what has been called a 'singer's formant' near the third formant. Vowels can also be modified so that the first or second formant acts to augment the fundamental frequency component or lower overtones of a particular pitch.

In the above examples, the voice source and vocal tract are considered to act relatively independently, with the vocal tract modifying the acoustic pulses produced by the laryngeal sound source. However, the author has also shown that some sopranos in their upper range can use a tuning of the first formant to the frequency (F_0) of the note sung to change advantageously the nature of the wave generated by the laryngeal sound source. This tuning can be used by the singer to reduce the average and peak air flow at the larynx, as well as to strengthen the higher harmonics at the level of the larynx, to produce a richer sounding note.¹

The above comments refer to previous attempts to use mathematical models to explain the quality of an individual voice. In this paper, we present a possible mathematical model for the strong perception of vocal harmony that can be attained by two or more singers simultaneously singing different notes separated by less than an octave. This model helps explain how the perception of harmony may relate to the vowel articulation used by the singers, and suggests how modifications in vowel articulation from normal spoken vowels might be used to enhance the acoustic effects of multiple voice harmony.

Tutorial on the acoustic analysis of quasi-periodic waves

Much of music and singing is based on the production of quasi-periodic acoustic waves, that is, waves in which the pattern of acoustic pressure variation tends to repeat at some time interval T_0 , resulting in

an acoustic sensation of pitch closely related to the repetition frequency $F_0 = 1/T_0$. Thus a piano on which the note A220 is struck will produce a decaying acoustic pressure pattern in which the variation of pressure closely repeats at an interval of $1/220$ second. The frequency of 220 Hz is referred to here as the fundamental frequency of this decaying tone.

However, because of the manner in which the auditory system perceives sounds, the A220 piano note will also be perceived to have energy at many multiples of the frequency 220 Hz, as 440 Hz, 660 Hz, etc. The tone of the piano note is then said to be a sum of a component tone at the fundamental frequency and component tones at the higher harmonics (multiples) of that frequency (440 Hz would be the second harmonic, etc.). (The component tone at the fundamental frequency can also be thought of as the *first* harmonic, but this terminology is rarely, if ever, used.) A machine or computer program which mimics this property of the ear to break down a complex tone into a sum of harmonic component tones is called a frequency analyzer or spectral analyzer, and the mathematical process which it uses is called frequency analysis or spectral analysis, or sometimes Fourier analysis, after the mathematician who first described the process.

Fourier described a mathematical process in which an arbitrary periodic wave can be approximated as closely as desired by a sum of 'component' *sinusoidal* waves having harmonically related frequencies, where a sinusoidal wave has a shape generated by varying the time t in the trigonometric sine function **$A \sin [2\pi F_0 (t + T)]$** , where A determines the amplitude of the wave and T determines a possible time displacement.

The power of Fourier's method of analysis lies in the fact that the component sinusoidal waves have unique mathematical and acoustical properties. For example, a sinusoidal tone can be clearly heard to be different from all other periodic waves of the same fundamental frequency, since a sinusoidal tone is the only tone that will not change in quality in an arbitrary acoustical environment. This implies that if you sound a continuous sinusoidal tone in a chamber with many reflections or echoes, it will still sound the same, except for a possible change in amplitude and you will not be able to perceive that the chamber has echoes. As another example, if the laryngeal voice source were to generate a sinusoidal tone, the acoustical chambers of the mouth and nose would have little effect on the quality of the sound emitted by the mouth, other than to change the loudness of the emitted sound, and the various vowels would not be well differentiated. (This is a problem encountered by some singers in falsetto voice.) Another property of a sinusoidal tone is that in a Fourier analysis, whether by machine or by the auditory system, there are no harmonic components to be found at frequencies above the fundamental frequency. Perhaps for this reason, the sinusoidal tone has been long referred to as a 'pure tone', even without the mathematics of Fourier to explain its uniqueness mathematically.

Thus, an arbitrary periodic tone or note can be considered to be the sum of a series of 'pure tones' that are harmonically related (have frequencies that are multiples of the fundamental frequency). Central to the understanding of harmony is the assumption that the auditory system can attend to these various component tones as if they were present as distinct acoustic entities. (More exactly, the auditory system is able to hear component tones as separate entities if they are more than a "critical bandwidth" apart in frequency. This a nicety that can be conveniently ignored in the present discussion.) Discussions of musical acoustics tend to take this assumption for granted, but it is really a complex concept that may take some understanding to appreciate.

The separation of notes by an octave and by a “fifth” as a basis for harmony

Any musician or singer understands the harmony that exists between two notes separated by one or more octaves, with an octave defined as a factor of two in frequency. Fourier’s decomposition of periodic functions helps explain why this is so in mathematical terms. Since the lower note in the octave can potentially contain components at all the component frequencies of the higher note, the two notes will fuse together to form one periodic tone at the fundamental frequency of the lower tone.

More subtle is the strong harmony heard between two tones a ‘fifth’ apart, as C and G, or A and E. The significance of what is termed an interval (distance between tones or notes) of a fifth results from the selection of a 12-tone scale for the construction of western music, with each successively higher tone in the 12-tone or ‘chromatic’ scale being a factor of almost 6% higher in frequency than the previous note. If we go up 6% twelve times, we are a factor of 100% higher, or up an octave, as expected (since the figure of 6% was obtained by breaking up an octave into 12 intervals). But, important for the study of harmony is the fact that if we go up 6% only 7 times, we end up a factor of almost exactly **1.5** times higher in frequency. ($1.06 \times 1.06 \times 1.06 \times 1.06 \times 1.06 \times 1.06 \times 1.06$ equals 1.5036)

Thus, if we compare the overtone structure of a given tone, say at a frequency f , and a tone 7 steps higher on the chromatic scale, at $1.5f$, we see that the third harmonic of the original tone, at $3f$, matches the second harmonic of the higher tone, at $2(1.5f)$. Likewise, the 6th harmonic of the original tone matches the 3rd harmonic of the higher tone, and so on, and the ear hears a strong merging between the two sounds (though of course not as strong a merging as between a tone and its octave). A tone 7 steps higher than the original tone, on the 12-tone scale, is commonly termed to be a ‘fifth’ higher. (This musical terminology, with the term ‘fifth’ referring to a difference of seven tones, is related to the construction of the piano keyboard and should not be allowed to cause confusion for the reader not familiar with musical terminology.)

Harmony in sung vowels

We here consider harmony between two notes to depend on the agreement, at least to within about one or two Hz, of one or more of the harmonic components of the two notes. This is a strong assumption, since there are undoubtedly many other facets of harmony to be considered, but it will suffice for the points to be made in this paper.

Consider the harmony between two notes a fifth apart (seven tones apart on the chromatic scale), using the example of A220 and its fifth E330 (approximately), as notes included in the singing range of both male and female adult singers and in the range of most instruments. Both notes will contain harmonics at 660Hz, 1320Hz, 1980 Hz, etc.

However, if we consider the perception of harmony between these two simultaneous notes when played on a musical instrument and when sung, there is an important difference. With some few exceptions, musical instruments produce notes having most of their energy at the fundamental frequency and lower harmonics, while the reverse is true for sung vowels. Basically, this is true because, unlike the case of the voice, in most musical instruments the note is generated by the resonance of some part of the instrument – as the string in a piano or guitar or the tubing in a wind instrument. In other words, the note sounds at the frequency at which the response of the instrument is

greatest. In the voice, however, the note is determined almost entirely by the tensions and masses of the vocal folds. The resonances (or formants) of the vocal tract do not significantly affect the note produced and instead act to selectively amplify various harmonics of the laryngeal tone.

The result is that in the voice, in all except the highest registers, the components at or nearest to the fundamental frequency are comparatively weak compared to components near the frequencies of the vocal tract resonances or formants. As an example, consider the example of A220 and its fifth E330. The coincidence frequencies of 660, 1320 and 1980 Hz are right in the middle of the range of the lowest (and strongest) three formants for both male and female singers. Similarly, with A110 and its fifth E165, notes more related to *male* choral singing than the notes an octave higher, the four lowest coincidence frequencies are approximately 330, 660, 990 and 1320 Hz. These frequencies are all in the range of the lower two formants of adult male vowels.

Thus by tuning the vowel being sung to accentuate one or more of these coincidence frequencies, two singers can conceivably attain a harmony much stronger than can two musicians playing the same notes on, say, a typical wind instrument. It is at least possible that some of the beautiful effects in vocal harmony, effects well beyond the harmony heard in instrumental chords, is due to this interaction of voice formants with the various vocal harmonics that coincide in sung notes spaced by certain intervals.

Adding a major or minor third to the harmony

If a third voice is to be added to a note and its fifth, it is often at an interval that is called a third from the note. A 'major third' is 4 steps above the original note on the chromatic, 12-tone scale, which is a factor of 1.26 higher than the note. A minor third is only 3 steps above the original note, which is a factor of 1.19. Both of these notes have harmonic components that closely match components of the original note and its fifth, to fit our requirements for harmony. Considering only the lowest 6 or 7 harmonics, we find that for the major third, its fourth harmonic matches fifth harmonic of the original tone, and its sixth harmonic matches the fifth harmonic of the note a fifth above the original tone. For the minor third (a factor of 1.19 above the original tone), its fifth harmonic closely matches both the sixth harmonic of the original note and the fourth harmonic of the note an interval of a fifth above the original note.

The matches in frequency listed above for the note a third above the original note are not as close as those between the note and its fifth, but in all cases mentioned match within about 1%. This is sufficiently close for a harmony to be heard, especially considering that a sung note can be sung a trifle sharp or flat if the harmony is perceived to be stronger. The matches mentioned also fall in the normal formant range of first or second formants.

Summary and Conclusions

A theoretical model for vocal harmony has been described, based on the premise that harmony is built on the coincidence of harmonics falling within the range of the lower vowel formants. An implication of this model is that because of the emphasizing of the higher harmonics produced by the resonances

of the vocal tract, the harmony effects in vocal harmony can be stronger than is the case with the same notes played by most musical instruments. A further implication of this model is that the perception of vocal harmony can be augmented by adjusting vowel articulation so as to match vowel formants to the frequencies of the coinciding harmonics. These conclusions can be tested both by generating synthesized sung notes and rating the quality of the harmony and by measuring the formant frequencies of highly qualified choral singers.

1. M. Rothenberg, Così Fan Tutte and What it Means - or - Nonlinear Source-Tract Acoustic interaction in the Soprano Voice and Some Implications for the Definition of Vocal Efficiency, in *Vocal Fold Physiology: Laryngeal Function if Phonation and Respiration*. T.Baer, C. Sasaki, and K.S. Harris, eds., College Hill Press, San Diego, pp. 254-263, 1986.