**Gary S. Kendall,* Christopher Haworth,†
and Rodrigo F. Cádiz****

*Artillerigatan 40
114 45 Stockholm, Sweden
garyskendall@me.com
†Faculty of Music
Oxford University
St. Aldate's,
Oxford, OX1 1 DB, United Kingdom
christopher.p.haworth@gmail.com
**Center for Research in Audio Technologies,
Music Institute
Pontificia Universidad Católica de Chile
Av. Jaime Guzmán E. 3300
Providencia, Santiago, Chile 7511261
rcadiz@uc.cl

Sound Synthesis with Auditory Distortion Products

Abstract: This article describes methods of sound synthesis based on auditory distortion products, often called combination tones. In 1856, Helmholtz was the first to identify sum and difference tones as products of auditory distortion. Today this phenomenon is well studied in the context of otoacoustic emissions, and the "distortion" is understood as a product of what is termed the cochlear amplifier. These tones have had a rich history in the music of improvisers and drone artists. Until now, the use of distortion tones in technological music has largely been rudimentary and dependent on very high amplitudes in order for the distortion products to be heard by audiences. Discussed here are synthesis methods to render these tones more easily audible and lend them the dynamic properties of traditional acoustic sound, thus making auditory distortion a practical domain for sound synthesis. An adaptation of single-sideband synthesis is particularly effective for capturing the dynamic properties of audio inputs in real time. Also presented is an analytic solution for matching up to four harmonics of a target spectrum. Most interestingly, the spatial imagery produced by these techniques is very distinctive, and over loudspeakers the normal assumptions of spatial hearing do not apply. Audio examples are provided that illustrate the discussion.

This article describes methods of sound synthesis based on auditory distortion products, often called combination tones—methods that create controlled auditory illusions of sound sources that are not present in the physical signals reaching the listener's ears. These illusions are, in fact, products of the neuromechanics of the listener's auditory system when stimulated by particular properties of the physical sound. Numerous composers have used auditory distortion products in their work, and the effects of these distortion products—often described as buzzing, ghostly tones located near to the head—have been experienced by many concert audiences. Historically, the technology for generating auditory distortion tones in musical contexts has been rather rudimentary, initially constrained by the limitations of analog equipment and always requiring high sound levels that are uncomfortable for most listeners. In this article, we describe methods of sound synthesis that both exploit the precision of digital

signal processing and require only moderate sound levels to produce controlled auditory illusions. Our goal is to open up the domain of sound synthesis with auditory distortion products for significant compositional exploration.

Auditory Distortion Products

There is a long history of research into what has commonly been called *combination tones* (CTs). Most studies of combination tones have used two pure tones (i.e., sinusoids) as stimuli and studied the listener's perception of a third tone, not present in the original stimulus, but clearly audible to the listener. In 1856 Hermann von Helmholtz was the first to identify sum and difference tones (von Helmholtz 1954). For two sinusoidal signals with frequencies f_1 and f_2 such that $f_2 > f_1$, the sum and difference tones have the frequencies $f_1 + f_2$ and $f_2 - f_1$ respectively. Later, Plomp (1965) identified many additional combination tones with the frequencies $f_1 + N(f_2 - f_1)$ Originally, it was thought that CTs occurred only at

high intensity levels that then drove the essentially linear mechanics of the physical auditory system into a nonlinear region. The original theory was that a mechanical nonlinearity was located in the middle ear or in the basilar membrane.

Goldstein (1967) provided a particularly thorough investigation of CTs produced by two pure tones. The frequency, amplitude, and phase of the distortion tones were determined using a method of acoustic cancellation, first introduced by Zwicker (1955). Importantly, Goldstein demonstrated that CTs were present at even low stimulus levels and thus could not be products of mechanical nonlinearity in the way they were originally conceived.

The theory of mechanical nonlinearity has been displaced after the recognition that parts of the inner ear, specifically, the outer hair cells of the basilar membrane, act as an active amplification system. So, rather than being a passive system with nonlinearities, the ear is an active one, and these nonlinearities are best explained in terms of the workings of the cochlear amplifier (Gold 1948; Kemp 1978). Seen from this perspective, CTs can best be understood as subjective sounds that are evoked by physical acoustic signals and generated by the active components of the cochlea. Combination tones are exactly the same as otoacoustic emissions, or, more specifically, distortion product otoacoustic emissions. Incidentally, distortion product otoacoustic emissions propagate back through the middle ear and can be measured in the ear canal. They are typical of healthy hearing systems and their testing has become a common diagnostic tool for identifying hearing disorders (Kemp 1978; Johnsen and Elberling 1983). To be perfectly clear, however, when experiencing distortion products as a listener, it is the direct stimulation of the basilar membrane that gives rise to the perception of sound, not the acoustic emission in the ear canal. This is why we use the term "distortion products" to refer to the general phenomena throughout this article.

Of the many distortion products, two types are particularly useful for music and sound synthesis due to the ease with which listeners can hear and recognize them: the quadratic difference tone ($f_2 - f_1$), QDT, which obeys a square-law distortion and the cubic difference tone ($2f_1 - f_2$), CDT, which obeys

cubic-law distortion. Despite the commonalities of their origins, there are considerable differences between the two. The CDT is the most intense distortion product and is directly observable to the listener even when acoustic stimuli are at relatively low intensity levels. However, because the tone's frequency ($2f_1 - f_2$) generally lies relatively close to f_1, it has seldom been commented on in musical contexts (a significant exception being Jean Sibelius's First Symphony, cf. Campbell and Greated 1994). The level of the CDT is highly dependent on the ratio of the frequencies of the pure tones, f_2 / f_1, with the highest level resulting from the lowest ratio and thereafter quickly falling off (Goldstein 1967). There is a loss of over 20 dB between the ratios of 1.1 and 1.3.

The QDT ($f_2 - f_1$) requires a higher stimulus intensity to be audible, but because the resultant tone's frequency generally lies far below the stimulus frequencies and thus can be more easily recognized, it has been a topic of musical discourse since its discovery by Tartini in 1754. The QDT shows little dependence on the ratio of the frequencies of the pure tones; levels are again highest with the lowest ratios and there is a roughly 10 dB loss between ratios of 1.1 and 1.8 (Goldstein 1967). Even simple characterizations of the differences between the CDT and QDT are subject to debate, and our understanding is frequently being updated by research.

A Study with Musical Tones

In a study easily related to musical tones, Pressnitzer and Patterson (2001) focused on the contribution of CTs to pitch, especially to the missing fundamental. They utilized a harmonic tone complex instead of the usual pair of pure tones. In their first experiment, they used a series of in-phase pure tones between 1.5 kHz and 2.5 kHz with a spacing of 100 Hz, as shown in Figure 1. Using the same cancellation technique as Goldstein, they measured the resulting amplitude and phase of the first four simultaneous distortion products at 100 Hz, 200 Hz, 300 Hz, and 400 Hz.

One consequence of employing the complex of pure tones was that each adjacent pair of

Figure 1. Representation of the harmonic tone complex used by Pressnitzer and Patterson (2001) in their Experiment 1 to measure

distortion products. The signal is comprised of 11 pure tones separated by frequency internals of 100 Hz between 1.5 kHz and 2.5 kHz.
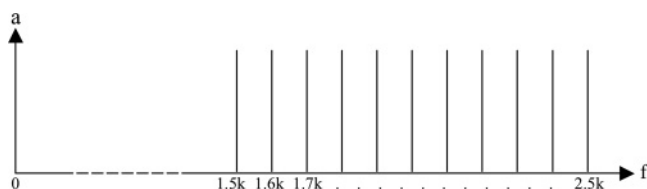
sinusoids contributed to the gain of the resulting fundamental (which was verified in their subsequent experiments). They report that "an harmonic complex tone … can produce a sizeable DS [distortion spectrum], even at moderate to low sound levels." They go on to establish that the level of the fundamental is essentially "the vector sum of the quadratic distortion tones … produced by all possible pairs of primaries." (This is a good first approximation in which the influence of CDTs is ignored.) Another consequence was that the resulting distortion products contained multiple harmonics of the fundamental (1,700 – 1,500 = 200 Hz; 1,800 – 1,500 = 300 Hz; etc.). These too were approximately vector sums of the corresponding pairs of pure tones. And, significantly, phase has a critical influence on these vector products because out-of-phase pure-tone pairs create out-of-phase distortion products that can cancel out the in-phase products when summed together. Therefore, to create distortion products with the highest gain, all acoustic components should be in phase with each other.

In addition, Pressnitzer and Patterson verified that there was relatively little intersubject variability. The predictability of QDTs and distortion spectra provides a practical foundation for the synthesis of more complex tones and dynamic sound sources that are heard by the listener yet are completely absent from the acoustic sound. In fact, from the listener's perspective, QDTs might just as well be externally generated sound, albeit a sound with some illusive perceptual properties.

The Missing Fundamental versus Combination Tones

(Refer also to Audio Examples 1 and 2a–b in Appendix 1.)

The "missing fundamental" is a perceptual phenomenon that is superficially related to CTs. In the psychoacoustic literature, the missing fundamental is most commonly referred to as "residue pitch," where "residue" refers to how the perceived pitch of a harmonic complex corresponds to the fundamental frequency even when the fundamental component is missing from the acoustic signal. The simplest way to illustrate the phenomenon is to imagine a 100-Hz periodic impulse train passing through a high-pass filter. Unfiltered, the sound will clearly have a perceived pitch corresponding to the 100-Hz fundamental as well as harmonics at integer multiple frequencies. But setting the high-pass filter's cutoff so that the 100-Hz component is removed does not cause the pitch to disappear; what changes, rather, is the perceived timbre of the tone. When raising the cutoff frequency even further, the pitch persists until all but a small group of mid-frequency harmonics remains (Ritsma 1962). Now, relating this back to Pressnitzer and Patterson's experiment with the harmonics of a 100-Hz fundamental, we might ask whether residue pitch and combination tones are essentially the same phenomenon.

It is true that the missing fundamental and combination tones have an intertwined history. Early researchers (Schaefer and Abraham 1904; Fletcher 1924) assumed that residue pitch was itself a form of nonlinear distortion, reintroduced by the ear when the fundamental was removed (Smoorenburg 1970). Schouten (1940) disproved this, however, by showing that the residue is not masked by an additional acoustic signal. This is illustrated in Audio Examples 1a–d (Appendix 1) where the residue is not masked by noise, whereas the combination tone is. This is a very important point for composers, because for CTs to be easily perceived by the listener, other sounds must not mask the distortion spectrum.

Importantly, Houtsma and Goldstein (1972) established that residue pitch is not dependent upon interaction of components on the basilar membrane. But recall that Goldstein (1967) as well as Pressnitzer and Patterson (2001) measured the properties of combination tones by cancelling them with acoustic tones. Combination tones require the interaction of components on the basilar

Kendall et al. **7**

membrane while residue pitch is then the product of a higher auditory "pattern recognition" mechanism (Houtsma and Goldstein 1972).

An equally compelling difference is illustrated in Audio Examples 2a and 2b, where each set of acoustic tones produces combination tones with the same subjective pitch. In Audio Example 2a the acoustic tones are higher harmonics of the perceived fundamental, $f_1 = 10 \ F$. In Audio Example 2b the acoustic tones are inharmonic to the fundamental, $f_1 = 10.7 \ F$, while still maintaining a frequency separation of the fundamental frequency, F. The subjective impression of the combination tones is essentially the same. This illustrates that CTs do not depend on harmonic ratios. In sum, although CTs and the missing fundamental may appear to be related, their underlying neurological mechanisms must be quite different.

Musical Applications

Although auditory scientists have expanded our knowledge of CTs, it was a musician who first discovered them, and the many composers and performers who have utilized the phenomena in their work inherit Giuseppe Tartini's early fascination with what he called the *terzo suono* [third tone]. As we will see, the computer musician is technologically better equipped to exploit the phenomena, given the exacting control one can exert upon all aspects of the acoustic sound. For historical reasons, however, it has tended to be improvisation that has afforded creative experimentation with auditory distortion. This is reflected by the many instrumental improvisers—for example, Yoshi Wada, Matt Ingalls, John Butcher, Pauline Oliveros, and Tony Conrad—who describe the role of the phenomena in their practice. Conrad has described his Theatre of Eternal Music improvisations with La Monte Young and others as a practice of working "on" the sound from "inside" the sound (Conrad 2002, p. 20), and his characterization indirectly illustrates why auditory distortion flourishes in this context. Where accidents and artifacts can be accepted or rejected, or enhanced or attenuated immediately, the opportunity for a subjectively heard "musical

layer" to be developed is greatest; greatest, that is, when the performer is free of a score. Evan Parker's *Monoceros* (1978) is a great example of a work in this tradition. Recorded from the microphone directly to the vinyl master using the "direct-cut" technique, the album comprises four solo soprano saxophone improvisations that explore a range of performance techniques including circular breathing and overblowing. These enable him to achieve a kind of polyphony from the instrument, with three or more registers explored simultaneously. When listened to at a high enough volume, the rapid cascades of notes in the altissimo range of the saxophone create fluttery distortion tones in the listeners' ears. The sheer melodic density of the piece, however, lends the distortion products a fleeting quality here: Listeners who do not know to listen for them could easily miss them. This is perhaps emblematic of the overall status of auditory distortion products in musical history—more "happy accidents" than directly controlled musical material.

Jonathan Kirk (2010) and Christopher Haworth (2011) have both described several instances in 20th-century music where this is not the case, and the auditory distortion product has been treated as a musical material in itself. Artists like Maryanne Amacher and Jacob Kierkegaard achieved this with the aid of computers, and for accurate control of the distortion product, the use of a pure-tone generator, at the very least, is essential. Phill Niblock is particularly worthy of note in this context, an artist whose approach falls squarely between the ear-guided instrumental work of Parker and the more exacting approach of somebody like Amacher. His work is composed of dense layers of electronically treated instrumental drones. He applies microtonal pitch shifts and spectral alterations in order to enhance the audibility and predominance of the naturally occurring combination tones, as well as to introduce new ones. Volker Straebel (2008), in his analysis of works by Niblock, counted as many as 21 CTs of different frequencies in *3 to 7 - 196* for cello and tape (Niblock 1974).

Niblock's drone music illustrates an important point concerning combination tones and perceptual saliency. A formally static, apparently stationary composition can reveal a multiplicity of acoustic

detail when listened to intently, and auditory distortion may often be noticed in this situation. Freely moving the head, one can easily recognize how this movement changes the intensity and localization of the resultant distortion products. Were the musical form rapidly changing and developing, this kind of comparison would not be possible, and so in many cases auditory distortion may simply go unrecognized. Niblock's approach therefore magnifies the conditions for the discrimination of auditory distortion from acoustic sound. Engineered during the editing process, the serendipitous quality of auditory distortion in music performance is, therefore, subtly effaced.

Like most techniques for creating auditory distortion, Niblock's approach can be considered to be "inside out," that is, he starts with the acoustic sound and manipulates it until the distortion product is rendered audible. Whether one is (like Tartini) playing the violin, or (like Niblock) digitally pushing partials to within close ratios, the fact remains that the distortion product as a musical material is fundamentally elusive here, controllable only in terms of its pitch and loudness. In order to achieve fine-grained control, one needs to reduce the acoustic variables to just those that are necessary.

Electronic musicians were quick to see the musical possibilities of the evolving notions of auditory nonlinearity. For instance, the British Radiophonic Workshop composer Daphne Oram devotes two chapters to the consideration of sum and difference tones in her book, *An Individual Note* (Oram 1972). Some years later these ideas were born into fruition by the late Maryanne Amacher, who made the solicitation of auditory distortion into an art form in its own right. Her sound installations and live performances became notorious for their utilization of interlocking patterns of short sine tone melodies reproduced at very high volumes, which induced prominent distortion tones in the ears of listeners. In the liner notes to *Sound Characters (Making of the Third Ear)*, Amacher gives a vivid description of the subjective experience of these tones:

> When played at the right sound level, which is quite high and exciting, the tones in this music

will cause your ears to act as neurophonic instruments that emit sounds that will seem to be issuing directly from your head . . . [my audiences] discover they are producing a tonal dimension of the music which interacts melodically, rhythmically, and spatially with the tones in the room. Tones "dance" in the immediate space of their body, around them like a sonic wrap, cascade inside ears, and out to space in front of their eyes . . . Do not be alarmed! Your ears are not behaving strange or being damaged! . . . These virtual tones are a natural and very real physical aspect of auditory perception, similar to the fusing of two images resulting in a third three dimensional image in binocular perception . . . I want to release this music which is produced by the listener . . . (Amacher 1999, liner notes).

The tones Amacher used to produce these effects were generated using the Triadex Muse, a digital sequencer instrument built by Edward Fredkin and Marvin Minsky at MIT. Amacher's is the first sound work to elicit a truly separate musical stream from the auditory distortion, a subjective "third layer," which she sometimes referred to as the "third ear" (Amacher 2004). This objectification of these previously ignored, subliminal sounds is very successful in Amacher's work, and is the point that we have taken forward in this research.

Practical Observations

In order for auditory distortion products to be musically meaningful, the listener must be able to distinguish them from acoustic sounds; otherwise, why not simply use ordinary acoustic signals? As already stated, fixed combinations of acoustic pure tones will produce sustained distortion tones with fixed frequencies. In this situation, the listener's head and body movements will produce important streaming cues for segregating the two sound sources (see the Spatial Imagery section, subsequently). For musical purposes, however, we may want to create sequences of pure tone complexes, thereby producing distortion-tone patterns that change over

time. Musically speaking, sequences of tones are more noticeable, as is illustrated by Amacher's 1999 piece "Head Rhythm/Plaything." The piece features a repetitive sequence of crude, pure-tone chirps that elicit a disorientating, subtly shifting rhythmic pattern of distortion tones at different frequencies; easily distinguishable from the tones used to generate them. But among the musical properties that have not been synthesized in any systematic way are dynamic properties of tones such as tremolo, vibrato, dynamic spectra, spatial location, etc. Computer synthesis enables the exploration of these possibilities in a way that was not available to the early practitioners, and has not been previously exploited in music synthesis.

Modeling Auditory Distortion as a Nonlinear System

The exact relationship between physical acoustic stimuli and the resulting auditory distortion products is quite complex, but in developing a systematic approach for synthesis, a good first approximation is to model the production of the distortion products as a general nonlinear system. We start with a classical power series representation (von Helmholtz 1954):

$$y - a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^2 \qquad (1)$$

where x is the input and y the output of the system. The a_n are constants. The nonlinearity of the output increases as the gain of the input level, x, increases.

Quadratic Difference Tone

The quadratic component, $a_2 x^2$, contributes the difference tone, $f_2 - f_1$, and also components at $2f_1$, $f_1 + f_2$, and $2f_2$, although at lower subjective levels. The level of the quadratic distortion tone (as measured by the acoustic cancellation method) is given by

$$L_{(f_2 - f_1)} - L_1 + L_2 - c \qquad (2)$$

where L_1 and L_2 represent the levels of the acoustic signals in decibels and C depends on the relative amplitude of the quadratic distortion, $C \approx 130$ dB (Fastl and Zwicker 2007). Experimental data in which a cancellation tone is used to determine the amplitude of the QDT exhibit a fairly regular behavior. The auditory QDT is well modeled as a quadratic distortion. With increasing L_1 or L_2 the cancellation level is almost exactly what is predicted and this happens whether the difference between the frequencies of the acoustic signals is large or small. For example, for $L_1 = L_2 = 90$ dB, the level of the cancellation tone is approximately 50 dB. (There is a percentage of listeners for whom this observation breaks down, see Fastl and Zwicker 2007, pp. 280–281.) For our purposes, variances in the effective amplitudes will have a relatively small effect on perceived timbres, especially dynamic ones.

The QDT as Distortion Product

Modeling the QDT as a nonlinear product is quite straightforward. If we consider the situation in which there are two sinusoidal inputs to the simple quadratic equation:

$$y - x^2, \qquad (3)$$

we find:

$$\begin{aligned}
y(t) &= (A_1 \sin(\omega_1 t) + A_2 \sin(\omega_2 t))^2 \\
&= A_1^2 \sin^2(\omega_1 t) + A_2^1 \sin^2(\omega_2 t) \\
&\quad + 2 A_1 A_2 \sin(\omega_1 t) \sin(\omega_2 t) \qquad (4)
\end{aligned}$$

where ω_1 and ω_2 are the sinusoidal frequencies and A_1 and A_2 their respective amplitudes. In expanding Equation 4, one finds that the first two terms yield a direct current (DC) component, and the third term supplies the important combination tones:

$$\begin{aligned}
y(t) &= \frac{A_1^2}{2} - \frac{A_1^2}{2} \cos(2\omega_1 t) + \frac{A_2^2}{2} - \frac{A_2^2}{2} \cos(2\omega_t) \\
&\quad + A_1 A_2 \cos((\omega_1 - \omega_2)t) - A_1 A_2 \cos((\omega_1 + \omega_2)t),
\end{aligned}$$
$$(5)$$

where $(\omega_1 + \omega_2)$ and $(\omega_1 - \omega_2)$ are the sum and difference frequencies. The gain of the difference frequency is $A_1 A_2$ (in decibels: $L_1 + L_2$). In addition

Table 1. Chords Evoked by Two Pure Tones and CDT

f_2/f_1	$f_1: f_2$	*Interval*	$2f_1\text{-}f_2: f_1: f_2$	*Resulting chord*
1.25	4:5	Major third	3:4:5	Major triad
1.2	5:6	Minor third	4:5:6	Major triad
1.166	6:7	~Minor third	5:6:7	~Diminished triad
1.1428	7:8	~Major second	6:7:8	Non-tertian triad
1.125	8:9	Major second	7:8:9	~Whole-tone cluster
1.111	9:10	Major second	8:9:10	Whole-tone cluster

The frequencies f_1 and f_2 generate a third tone, the CDT, at the frequency $2f_1 - f_2$, supplying the lowest note of a three-note chord. The table gives the interval between the first two tones and the kind of chord resulting. Intervals and chords marked with the tilde (~) are slightly out of tune.

to the sum and difference frequencies, the complete output signal of the squarer contains DC and components at twice the input frequencies, components that are inaudible.

Cubic Difference Tone

The cubic component, a_3x^3, contributes the cubic difference tone, $2f_1 - f_2$ and also $2f_2 - f_1$, $3f_1$, etc. Experimental test data do not conform well to what would be predicted for regular cubic distortions. For example, the level of the CDT is strongly dependent on the frequency separation between the pure tones, $f_2 - f_1$ (Fastl and Zwicker 2007). This means that the auditory CDT is not well modeled as a regular cubic distortion. Its characteristics under varying circumstances are far more idiosyncratic than the QDT. In particular, the level's dependency on both frequency separation and frequency range is another reason why CDTs are difficult to use in a controlled way for synthesis, even though under ideal circumstances the level of the CDT is significantly higher than the QDT.

CDT Ratios

The CDT is most clearly audible when the ratio of the acoustic signals, f_2/f_1, lies between 1.1 and 1.25. Ratios within this range coincide with musical intervals between a major second and a major third. And, as we expect with musical intervals, ratios below 1.14 produce auditory roughness (or disso-

nance from the musical perspective). In addition, the CDT itself falls so close to f_1 and f_2 that what one typically perceives is a three-tone aggregate. As the example from Sibelius's First Symphony illustrates (Cambpell and Greated 1994), if the ratio, f_2 / f_1, forms a musical interval, the CDT will form another musical interval to yield a three-note chord. These relationships are summarized in Table 1 using simple integer ratios for illustration.

Synthesis Techniques

For the purposes of sound synthesis, the direct generation of quadratic and cubic difference tones from a pair of pure tones suffers from important limitations. As noted earlier, the CDT is comparatively louder than the QDT, but the close proximity of the CDT's frequency to the acoustic stimuli limits the circumstances in which the listener can easily distinguish it from the acoustic tones. In order to create QDTs at levels that the listener can recognize, the acoustic pure tones have to be presented at a level that is uncomfortable for most listeners, especially for any extended period of time.

Haworth solved the problem for QDTs in concert settings by utilizing a sinusoidal complex with constant difference frequencies, akin to the stimuli of Pressnitzer and Patterson discussed earlier. In the composition "Correlation Number One" (2010), each adjacent pair of sinusoids produces the identical QDT frequency, adding linearly to its total gain and, thereby, increasing the level of the distortion

Figure 2. Quadratic difference tone (QDT) spectrum (dashed lines) produced by pure tones (solid lines) with a constant frequency interval of F.

tone (Haworth 2011). Not only did the combination of acoustic sinusoids increase gain, but it also produced components that were harmonics of the primary QDT. Importantly, increasing the number of acoustic tones (and consequently spreading them over a wider frequency range) permits the subjective level of the acoustic tones to be reduced, thus greatly diminishing the problem of listener fatigue.

Clearly, the musical context in which auditory distortion products are used dictates to a large degree how successful the effects will be. We have noted that the listener must be able to distinguish the distortion products from the acoustic tones, and for this to happen careful attention must be paid to the frequency spectrum. Generally, researchers have focused on QDTs below 1 kHz with pure tones between 1 and 5 kHz. This gives some guidance to the most practical frequency ranges to use when there are no competing sounds. Recognition of the presence of auditory distortion products requires that they be aurally separable from acoustic sounds by pitch or by other means. This has an important impact on the choice of synthesis methods. Other high-frequency acoustic signals overlapping the frequency range of the acoustic signals stimulating the auditory distortion tones can produce unintended side effects and weaken the impact of the distortion products. Also, the presence of other acoustic signals overlapping the range of the distortion products themselves can mask and destroy their effect. It may be obvious to say, but auditory distortion products, like many other aspects of synthesis, are best adjusted and optimized by ear. Many imaginative effects can be achieved through creative use of synthesis. We summarize the most important synthesis methods here.

Direct Additive Synthesis

(Refer also to Audio Example Group 2 in Appendix 1.)

Pressnitzer and Patterson (2001) demonstrated that multiple pure tones synthesized at sequential upper harmonics of a fundamental, F, produce a harmonic QDT spectrum with the fundamental F. The gains of the individual harmonic components of that spectrum are a summation of the QDT
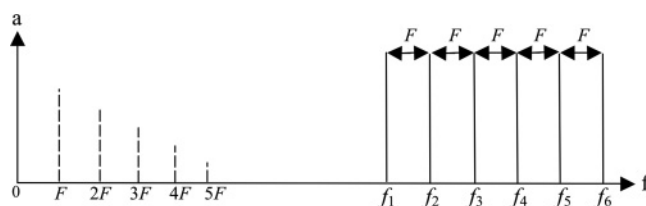
contributions produced by each pair of pure tones. For example, they demonstrated that harmonics 15 to 25 of a 100-Hz fundamental, each at 54 dB SPL, produces a harmonic QDT spectrum with a fundamental only 10–15 dB lower than the gain of the acoustic tones.

But to produce a QDT harmonic spectrum, the acoustic pure tones (f_1, f_2, f_3, etc.) do not need to be harmonics of the QDT fundamental, they only need to be separated by the constant frequency interval F ($F = f_2 - f_1 = f_3 - f_2$, etc.). This produces a QDT spectrum with a fundamental of F as shown in Figure 2. The exact quality of the resulting distortion tones experience depends on the choice of f_1 and the number of acoustic, sinusoidal components.

By itself, this technique can produce QDT spectra that are clearly audible in typical loudspeaker reproduction at moderate sound levels. And from this starting point, many classic time-domain synthesis processes can be introduced with trivial ease, for instance, amplitude modulation (AM). There are two possibilities for AM that each yield slightly different results depending on how many acoustic signals are producing the effect. Modulating all pure tones together produces a single, amplitude-modulated sound. Modulating all pure tones except for the lowest, f_1, enhances the effect of a sustained pure tone plus an amplitude-modulated distortion product. The latter case provides better subjective timbral segregation between the acoustic tones and the distortion tone, whereas in the former case the two tend to fuse. Altering the modulation rate has predictable results. A pleasant tremolo effect occurs up to approximately 15 Hz, and then "roughness" between 20 Hz and 30 Hz. Increasing the modulation rate much further introduces sidebands in the acoustic signals that may interfere with the intended distortion spectrum.

Figure 3. Dynamic sinusoidal synthesis of a QDT spectrum based on an audio signal input. The frequency of the lowest acoustic component is f_1, and N is the number of additional sinusoids synthesized.

Applying the same principles to the frequencies of the acoustic signals produces frequency modulation of the QDT spectrum. In the simplest case, the f_1 frequency best remains static, and the modulation is applied to the frequency separation between the other components ($f_2 - f_1, f_3 - f_2$, etc.). In this way, the frequency modulation of the fundamental, F, is easily audible, while the modulation of the acoustic components is less distinct because the pitch interval of their deviation is much smaller than for the acoustic tones.

The rates of both AM and FM require subtle adjustment, otherwise the roughness caused by beating of the acoustic frequencies will interfere with the segregation of the QDT spectrum. The FM rate parameter, in particular, offers a few additional possibilities. If it is set sufficiently high, even at relatively small frequency deviations the pitch sensation of the distortion tone will be lost. In itself, this gives a fairly dull, static sound, rather like narrowband noise. But if one applies a repeating sequence of short time windows to the sound stream, akin to synchronous granular synthesis, then the results become more interesting. If we choose a slow frequency modulation rate (<12 Hz) and a repeating envelope with a sharp attack and sloping decay, then, due to the closeness of the distortion tone and its unresolved pitch, one perceives a fluttery, wind-like sound that appears to bristle against the ear. The techniques described here are employed in Haworth's compositions "Correlation Number One" (2011) and "Vertizontal Hearing (Up & Down, I then II)" (2012).
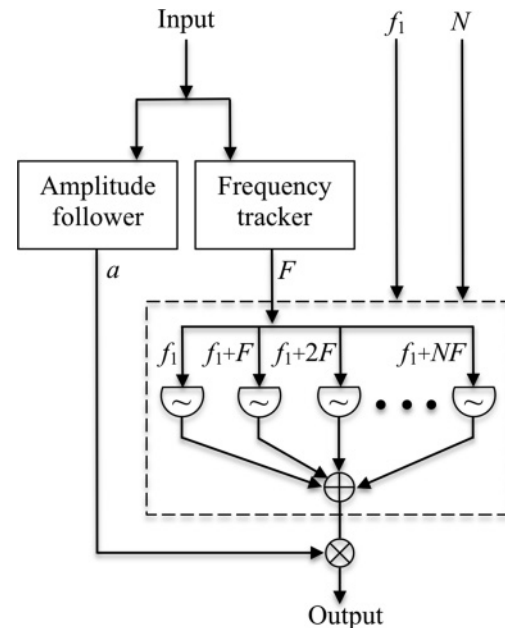
Dynamic Sinusoidal Synthesis

(Refer also to Audio Example Group 3 in Appendix 1.)

The basic processes described in this article can also be applied to situations in which the fundamental frequency and the overall amplitude are dynamically changing. Most importantly, the pitch and amplitude of the QDT spectrum can be made to follow the characteristics of a model signal, including a recorded or real-time performance. Again, in order for the distortion tone to be heard clearly by the listener, the synthesis must again rely on multiple pure tones with a constant frequency offset.
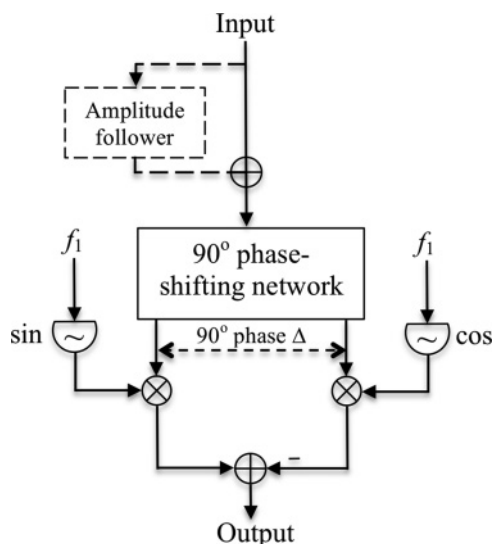
Consider the algorithm illustrated in Figure 3. Here the audio input signal is fed to a frequency tracker and an amplitude follower to dynamically extract its fundamental, F, and its amplitude, a. Then, F is fed to a sinusoidal oscillator bank along with two values set by the user: f_1, the frequency of the lowest sinusoid, and N, the number of additional sinusoids to synthesize. N will determine the strength of the QDT spectrum's fundamental and the number of its possible harmonics. Typically one oscillator with frequency f_1 remains constant, while N additional oscillators dynamically follow the value of F at integral multiple offsets from $f_1, f_1 + F, f_1 + 2F, \ldots f_1 + NF$. The sum of the sinusoids produced by the oscillator bank is multiplied by the output of the amplitude follower, a, to recreate the original envelope. The result of the synthesis will be like that shown in Figure 2, only dynamic.

In this way, the QDT spectrum can mimic the dynamic character of a live or prerecorded sound. Of course, the success depends on the nature of the sound material and the degree to which there is a fundamental frequency to extract. Then too, the same oscillator-bank technique can be used

Kendall et al.   **13**

Figure 4. Adaptation of single-sideband modulation to creating QDT spectra. The carrier frequency is represented by f_1*.*

Figure 5. The acoustic spectrum produced by the adaptation of single-sideband modulation for the synthesis of QDT spectra.

The presence of the carrier with frequency f_1 *depends on whether the optional amplitude follower shown in Figure 4 is implemented.*

in direct synthesis, where the composer specifies dynamic *F* and *A* directly. The important property of the acoustic input signal not captured by this algorithm is timbre. For the composer interested in such an effect, changes in the brightness of the QDT spectrum can be controlled through dynamic modulation of *N* (which affects the number of the harmonics in the resulting spectrum).
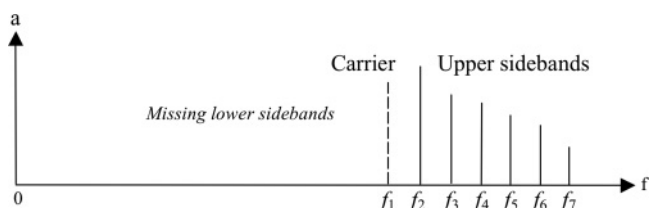
Single-sideband Modulation

(Refer also to Audio Example Group 4 in Appendix 1.)

The arrangement of the partials in Figure 2 might be somewhat suggestive of the spectra produced by sinusoidal frequency modulation in the manner described by Chowning (1973), with sidebands spaced at constant frequency intervals. But distortion product synthesis through conventional FM is only partially effective. The problem with FM is that half the lower sidebands are 180 degrees out of phase with the upper sidebands, and the gain of the resulting QDTs is diminished due to phase cancellations. Conventional amplitude modulation fares no better.

The solution is to use single-sideband amplitude modulation, which is most familiarly used as a technique for frequency shifting (Bode and Moog 1972). As in the case of dynamic sinusoidal synthesis,

the technique is easily implemented in real time. Figure 4 illustrates an adaptation of single-sideband modulation for creating QDT spectra. An optional amplitude follower (explained more thoroughly subsequently) is shown at the beginning, whose role is to add a DC component to the input signal proportional to the signal gain. The input audio signal is sent to a 90-degree phase shifting network that outputs two versions of the input, 90 degrees out of phase to each other. These two outputs are multiplied by two pure tones at f_1 that are also 90 degrees out of phase, here represented as sine and cosine waves. The multiplications each produce ring modulation (otherwise known as double-sideband suppressed-carrier amplitude modulation), but in this case the amplitude follower reintroduces the carrier frequency, f_1, in the output signal by adding the DC component to the input. The amplitude spectra produced by each ring modulation are the same, but the phase relationships are different. When the two ring-modulated signals are subtracted, the lower sidebands will be 180 degrees out of phase, while the upper sidebands are in phase. The result is that the lower sidebands are cancelled out and only the upper sidebands remain, as illustrated in Figure 5.

A particular advantage of single-sideband modulation over dynamic sinusoidal synthesis is that the number of sideband components depends on the input signal. More harmonics in the input signal produce more components in the output and, therefore, more harmonics in the QDT spectrum. The timbres of the input signal and the QDT spectrum do not match up in any precise way, but brighter inputs will create brighter distortion spectra. The use of the amplitude follower is especially recommended in cases when there are few harmonics. An additional advantage of single-sideband modulation is that noise in the input signal retains its noisy character in the output. This helps to support speech

intelligibility by reinforcing noisy consonants. And, whereas monophonic music is easily recognizable, polyphonic music suffers from the fact that acoustic signal is not limited to components that are equally spaced in frequency, as they were in Figures 2 and 5. The perceptual result is that the distortion products sound similar to ring modulation.

Analytic Solution for Matching Timbre with Three and Four Harmonics

(Refer also to Audio Example Group 5 in Appendix 1.)

A more precise control of the timbre of a QDT spectrum is possible by controlling the relative amplitudes of the multiple pure tones in the acoustic signal (A_1, A_2, . . . A_5, corresponding to f_1, f_2, . . . f_5). Five acoustic components will produce four harmonics in the QDT spectrum. In the case that the target amplitudes (T_1, T_2, T_3, T_4) are static, then the computation of the amplitudes needs only be performed once. In the more general case that the target amplitudes are dynamically changing, the computation needs to be performed for each update to the target spectrum. There are two critical restrictions here. The first is that the limited number of acoustic components reduces the overall intensity of the distortion tone (requiring that the acoustic signal be reproduced at higher levels than the other methods). The second is that the target amplitudes of all the QDT spectrum's harmonics need to be known at each update. These are best determined by sinusoidal analysis, which inevitably leads to latency in real-time situations. Nonetheless, within these restrictions the distortion tone's timbre can be directly controlled.

It is possible, under certain constrains, to specify in Equation 3 an arbitrary harmonic signal $y(t)$ with a finite number of harmonics, N, and then calculate the amplitudes of the sum of sinusoids needed at the input to produce the target spectrum. The complexity of this procedure increases quadratically as N grows. To understand this, consider that an input of N sinusoids to a quadratic function, such as the one in Equation 3, produces N regular sinusoids in the output and $(N^2 - N)/2$ intermodulation products. (Consult Sea 1968 for further discussion.)

As an example, let us assume that we want to specify $y(t)$ with a fundamental frequency, F, and a spectrum of three harmonics ($N = 3$) with amplitudes T_1, T_2, and T_3. Let us denote the amplitudes of the sinusoids of the input signal as A_1, A_2, A_3, and A_4. Then, if we extend Equations 4 and 5, the problem reduces to solving the following system of equations:

$$T_1 = A_1 A_2 + A_2 A_3 + A_3 A_4$$
$$T_2 = A_1 A_3 + A_2 A_4$$
$$T_3 = A_1 A_4$$
$$A_1 = 1 \tag{6}$$

(Please refer to Appendix 2 for a full derivation and discussion.)

This seems to be a fairly simple set of equations that—at first sight—should not present much of a problem. But it is actually a nonlinear system, because it involves solving a second order polynomial. Indeed, the solutions for some of the coefficients are conjugate pairs, as shown in the following set of equations:

$$A_1 = 1$$
$$A_2 = \frac{T_2 \pm \frac{\sqrt{T_2^2 + 2T_2 T_3^2 + 2T_2 + T_3^4 - 2T_3^2 - 4T_1 T_3 + 1}}{2} - \frac{T_3^2}{2} + \frac{1}{2}}{T_3}$$
$$A_3 = \frac{T_2}{2} \pm \frac{\sqrt{T_2^2 + 2T_2 T_3^2 + 2T_2 + T_3^4 - 2T_3^2 - 4T_1 T_3 + 1}}{2}$$
$$+ \frac{T_3^2}{2} - \frac{1}{2}$$
$$A_4 = T_3 \tag{7}$$

If we extend the problem to allow for five sinusoidal components, we have the system:

$$T_1 = A_1 A_2 + A_2 A_3 + A_3 A_4 + A_4 A_5$$
$$T_2 = A_1 A_3 + A_2 A_4 + A_3 A_5$$
$$T_3 = A_1 A_4 + A_2 A_5$$
$$T_4 = A_1 A_5$$
$$A_1 = 1 \tag{8}$$

Kendall et al.   **15**

In this particular case, the solutions for each amplitude A_n consist of one real solution and a pair of complex conjugate solutions, as solving the system in Equation 8 involves finding the root of a third-order polynomial. The solutions are so long and intricate that they are essentially impractical. If the number of harmonics is increased to six, the network of equations is impossible to solve analytically (see Appendix 2). We have found that the method of calculating the A_n coefficients by direct solution is practical only up to three harmonics in the QDT spectrum.

Spatial Imagery

(Refer also to Audio Example Group 6 in Appendix 1.)

Synthesis with auditory distortion products often produces quite distinctive spatial imagery in loudspeaker reproduction. In particular, sonic images typically appear much closer than the loudspeakers and very close to the head. This is one of the most distinctive properties of CTs. This difference in localization assists the listener in segregating the CT, and good segregation is critical for the tones to be heard as musical material in their own right.

The localization of CTs is relatively easy to understand and to predict. Most interestingly, localization in loudspeaker reproduction is quite different from conventional audio and very similar to headphone reproduction. The reason for this is obvious with a little explanation. Distortion product tones are produced within each of the two ears. For localization purposes, it is as if the signals bypass the outer and middle ears and arrive directly in the left and right basilar membranes. The perceived spatial location is governed by interaural time differences (ITDs) and interaural intensity differences (IIDs) in a way that is quite analogous to what happens with headphone lateralization. By themselves, distortion product tones lack externalization: Their imagery is within or just outside of the head. Far left and far right images are characteristically at the sides of the head. This is to say, the tones behave exactly as they would in headphone reproduction if they were, in fact, normal acoustic signals. Stereo panning

of the high-frequency acoustic signals moves the resulting distortion product tones between the ears just as if they themselves were being panned. How closely ITD and IID lateralization of the combination tones matches that of acoustic signals has not been studied, but our informal testing suggests that the relationship is not linear; IIDs of the acoustic signals have a somewhat exaggerated effect on the combination tones. For example, the tones may appear to be at the left ear when the intensity balance of the acoustic tones is roughly 75 percent versus 25 percent. This exaggeration of the lateralization may reflect the nonlinear relationship of the CT energy to the physical acoustic energy.

The important thing to remember, then, when considering loudspeaker reproduction is that the spatial perception of combination tones is not directly influenced by head-related transfer functions. Thus, the elevation, the front or back positioning, and the distance of the loudspeakers matter only in how they affect interaural differences. That said, it is undoubtedly the case that many listeners' subjective experience will be influenced by the overall acoustic environment, and they may report images that lie in the space between the listener and the plane of the loudspeakers. In particular, the listener's perception may be influenced by the localization of the high-frequency acoustic tones: If they are elevated, then the combination tone may appear elevated also. Then again, turning one's head in loudspeaker reproduction can quickly shift the interaural balance and produce rapid, disorienting shifts in the apparent location of the distortion product image between the far left and far right.

An important consequence of this unusual localization with loudspeakers is that the normal rules of spatial hearing with loudspeakers are broken. Most importantly, the precedence effect has no influence. In normal stereo reproduction with loudspeakers, a centered phantom image is only possible in the "sweet spot" where the left and right loudspeaker signals arrive at the left and right ears roughly simultaneously. In off-center listening locations, the spatial image collapses in the direction of the first-arriving sound, that is, from the direction of the closest loudspeaker. But distortion product

images do not shift until there is a significant shift in the intensity balance when far off center—the effective sweet spot is quite large because the intensity ratio between the ears changes quite slowly with distance. In multi-channel reproduction, distortion product images will shift toward the side of the head that is receiving the most energy, and therefore these images will shift as the composite intensity balance between the ears shifts. Still, the images retain the quality of being inside or close to the head.

Because of the spatial stability of the subjective images, distortion product synthesis is very adaptable to settings in which the listeners walk about the space. An interesting side effect is that distortion tones with frequencies lower than 500 Hz (which are influenced by the phase of the signals arriving from loudspeakers) will appear to have peaks and valleys located in fixed physical locations that the listener can walk through, a bit like having waves of intensity mapped across the space.

It is also possible to create broad, diffuse spatial imagery by some very simple means. The pure-tone complexes from multiple loudspeakers can produce decorrelation of the CTs when the fundamental frequencies are offset by small amounts or when applying small amounts of random frequency jitter to the individual loudspeakers. The resulting imagery combines closeness, in terms of the localization of the distortion products, with the diffuseness one might expect to achieve from using decorrelation in more conventional signal processing. Splitting up the acoustic pure tones among the different loudspeakers produces especially complex spatial imagery, a technique that is particularly sensitive to changes in the listener's head position. More empirical work would need to be done, however, to quantify the influence of phase, ITD, and IID on the distortion product images in these settings. For the purposes of this article, the composer's ear should be the judge of the effect.

Conclusions

Auditory distortion has long been an object of musical fascination, though in practice its use has tended to favor ear-guided improvisatory contexts including drone-based music. Maryanne Amacher's sonic artworks demonstrated that auditory distortion products could be directly controlled, and with high-intensity sound could be clearly audible in concert settings. Christopher Haworth was the first to accomplish this at moderate sound levels (Haworth 2011). Moving on from their work, we have codified multiple synthesis techniques by which the auditory distortion of the quadratic difference tone can be controlled in a decisive and artistically useful manner. The diversity of synthesis techniques ranges from static tones to dynamic sounds that track the pitch and amplitude of recorded sources, creating distinctive subjective images of those sources. Because these synthesis techniques are based on distortion products produced within the inner ear, their spatial properties are particularly distinctive and the normal rules of spatial hearing with loudspeakers are broken. The range of subjective possibilities for distortion products created through precise digital control has been extended well past that typical of psychoacoustic experiments, and into a domain where distortion product tones share the dynamic properties of traditional acoustic sound, albeit within practical limitations. Sound synthesis with auditory distortion products has thus become an area for artistic exploration and further development.

Acknowledgments

References

Amacher, M. A. 1999. *Sound Characters: Making the Third Ear*. New York: Tzadik TZ 7043, compact disc.

Amacher, M. A. 2004. "Psychoacoustic Phenomena in Musical Composition: Some Features of a Perceptual Geography." *FO(A)RM* 3:16–25.

Bode, H., and R. A. Moog. 1972. "A High-Accuracy Frequency Shifter for Professional Audio Applications." *Journal of the Audio Engineering Society* 20(6):453–458.

Campbell, M., and C. Greated. 1994. *The Musician's Guide to Acoustics*. Oxford: Oxford University Press.

Cheney, E. W., and D. R. Kincaid. 2009. *Linear Algebra: Theory and Applications*. Burlington, Massachusetts: Jones and Bartlett.

Chowning, J. 1973. "The Synthesis of Complex Audio Spectra by Means of Frequency Modulation." *Journal of the Audio Engineering Society* 21(7):526–534.

Conrad, T. 2002. *Early Minimalism: Volume One*. New York: Table of the Elements, liner notes to compact disc.

Fastl, H., and E. Zwicker. 2007. *Psychoacoutics: Facts and Models*. 3rd ed. Vol. 22 of *Springer Series in Information Sciences*. Berlin: Springer.

Fletcher, H. 1924. "The Physical Criterion for Determining the Pitch of a Musical Tone." *Physical Review* 23(3):427–437.

Gold, T. 1948. "Hearing II: The Physical Basis of the Action of the Cochlea." *Proceedings of the Royal Society of London B: Biological Science* 135(881):492–498.

Goldstein, J. L. 1967. "Auditory Nonlinearity." *Journal of the Acoustical Society of America* 41(3):676–699.

Haworth, C. 2011 "Composing with Absent Sound." In *Proceedings of the International Computer Music Conference*, pp. 342–345.

von Helmholtz, H. L. F. (1885) 1954. *On the Sensations of Tone*. A. J. Ellis, trans. Unabridged and unaltered republication of the second edition of the translation with a new introduction by H. Margenau. New York: Dover.

Houtsma, A. J. M., and J. L. Goldstein. 1972. "The Central Origin of the Pitch of Complex Tones: Evidence from Musical Interval Recognition." *Journal of the Acoustical Society of America* 51(2):520–529.

Johnsen, N., and C. Elberling. 1983. "Evoked Acoustic Emissions from the Human Ear: III; Findings in Neonates." *Scandinavian Audiology* 12(1):17–24.

Kemp, D. T. 1978. "Stimulated Acoustic Emissions from within the Human Auditory System." *Journal of the Acoustical Society of America* 64(5):1386–1391.

Kendall, G., C. Haworth, and R. F. Cádiz. 2012. "Sound Synthesis with Auditory Distortion Products." In *Proceedings of the International Computer Music Conference*, pp. 94–99.

Kirk, J. 2010. "Otoacoustic Emissions as a Compositional Tool." In *Proceedings of the International Computer Music Conference*, pp. 316–318.

Mahler, G. 1990. *Symphony No. 9* and *Symphony No. 10—Adagio*. (Claudio Abbado, Vienna Philharmonic Orchestra.) Berlin: Deutsche Grammophon DG 423 564-2, compact disc.

Niblock, P. 1974. *3 to 7–196*. (Stereo mix.) New York: Archive Phill Niblock.

Oram, D. 1972. *An Individual Note: Of Music, Sound and Electronics*. London: Galliard.

Parker, E. 1978. *Monoceros*. London: Incus 27. Reissued 1999, London: Coronoscope CPE2004-2, compact disc.

Plomp, R. 1965. "Detectability Threshold for Combination Tones." *Journal of the Acoustical Society of America* 37(6):1110–1123.

Pressnitzer, D., and R. D. Patterson. 2001. "Distortion Products and the Perceived Pitch of Harmonic Complex Tones." In D. J. Breebart et al., eds. *Physiological and Psychophysical Bases of Auditory Function*. Maastricht: Shaker, pp. 97–104.

Ritchie, J. 2003. *Jean Ritchie Ballads from Her Appalachian Family Tradition*. New York: Smithsonian Folkways SFW CD40145, compact disc.

Ritsma, R. J. 1962. "Existence Region of the Tonal Residue I." *Journal of the Acoustical Society of America* 34(9A):1224–1229.

Rosloniec, S. 2008. *Fundamental Numerical Methods for Electrical Engineering*. Berlin: Springer.

Schaefer, K. L., and Abraham, O. 1904. "Zur Lehre von den sogenannten Unterbrechungstönen." *Annalen der Physik* 318(5):996–1009.

Schouten, J. F. 1940. "The Perception of Pitch." *Philips Technical Review* 5(10):286–294.

Sea, R. G. 1968. "An Algebraic Formula for Amplitude of Intermodulation Products Involving an Arbitrary Number of Frequencies." *Proceedings of the IEEE* 56(8):1388–1389.

Smoorenburg, G. F. 1970. "Pitch Perception of Two-Frequency Stimuli." *Journal of the Acoustical Society of America* 48(4B):924–942.

Straebel, V. 2008. "Technological Implications of Phill Niblock's Drone Music, Derived from Analytical Observations of Selected Works for Cello and String Quartet on Tape." *Organised Sound* 13(3):225–235.

Zwicker, E. 1955. "Der Ungewöhnliche Amplitudengang der Nichtlinearen Verzerrungen des Ohres." *Acustica* 5(Supplement 1):67–74.

Appendix 1: Audio Examples

The audio examples that accompany this article can be listened to with either stereo loudspeakers or headphones, although Examples 6a–b are best listened to over headphones.

[Editor's note: Sound examples to accompany this article are present on the *Journal*'s Web site at www.mitpressjournals.org/doi/suppl/10.1162/ COMJ_x_00265.]

Example Group 1: Missing Fundamental versus Combination Tones

Here a fundamental difference between the phenomena of the missing fundamental (or residue pitch) and combination tones is demonstrated, a difference in what happens under conditions of masking. In each example there is a pitch gliding from middle C to the octave below.

Examples 1a–b: Missing Fundamental

In Example 1a, a pure-tone complex with harmonics 5–8 produces a glide in pitch that is easily heard. In Example 1b, the same pure-tone complex as 1a is initially combined with low-pass noise with a 1,500-Hz cutoff. The pitch of the pure-tone complex is heard even when the noise is present.

Examples 1c–e: Combination Tones Producing Pitch

In Example 1c a summation of 15 sinusoids with its lowest frequency, f_1, at 2,800 Hz produces the same slide in perceived pitch as Examples 1a–b. In Example 1d the same signal as Example 1c is initially combined with low-pass noise with a 1,500-Hz cutoff, but, although the sinusoids can always be heard, the sliding pitch is only heard when the noise fades away. The noise has effectively masked the combination tone. In Example 1e the same signal as Example 1c is initially combined with a pure tone at F# below middle C. The sliding pitch is only heard when the tone fades away. This time, the pure tone has effectively masked the combination tone.

Example Group 2: Direct Additive Synthesis

Synthesis of auditory distortion products by the direct additive synthesis with pure tones.

Examples 2a–b: Frequency of the Distortion Product's Fundamental

Audio Example 2a demonstrates the case in which the frequencies of the pure tones are identical with high harmonics of the distortion product's fundamental. In this case, the pure tones represent harmonics 10–35 of a fundamental, F, of middle C (261.6 Hz). On the face of it, this case is again similar to the phenomenon of the missing fundamental (residue pitch). But Example 2b produces a virtually identical result while the frequencies of the pure-tone complex are inharmonic partials of the frequency of the fundamental. The lowest pure tone, f_1, has a frequency of 2,800 Hz, and the next 24 pure tones have a constant separation of F. This demonstrates that the generation of auditory distortion products is not dependent on a harmonic relationship between F and f_1.

Examples 2c–d: Oblique and Contrary Motion

The same method of synthesis as Examples 2a and 2b is used, except that the frequencies of the pure-tone complex are dynamically changing. In Example 2c, the frequency of the perceived fundamental, F, is fixed while f_1 and all of the pure tones rise in frequency. This creates oblique motion. In Example 2d, F is decreasing while f_1 is increasing, an example of contrary motion.

Example 2e: Amplitude Modulation

Again, F is middle C and f_1 is 2,800 Hz with 25 pure tones in the complex. The pure-tone complex is amplitude modulated with a modulation frequency that increases from 1 to 200 Hz.

Example 2f: Frequency Modulation

This is the same as Example 2e, except that the pure-tone complex is frequency modulated with

Kendall et al. **19**

a modulation frequency that increases from 1 to 200 Hz.

Example Group 3: Dynamic Sinusoidal Synthesis

The fundamental, F, and the amplitude of the pure-tone complex are determined by pitch tracking and amplitude following a recorded audio excerpt.

Example 3a: Original Audio Excerpt

This excerpt is performed by the Appalachian folk singer Jean Ritchie. The song is "The Unquiet Grave" taken from the CD *Jean Ritchie: Ballads from Her Appalachian Family Tradition* (2003).

Example 3b: Dynamic Synthesis with Harmonic Pure Tones

Pitch tracking of the Jean Ritchie excerpt determines the fundamental, F, which determines the frequencies of a pure-tone complex of harmonics 20 through 45. The amplitude of the pure-tone complex is determined by amplitude following the audio excerpt.

Example 3c: Dynamic Synthesis with Inharmonic Pure Tones

Again, pitch tracking of the Jean Ritchie excerpt determines the fundamental, F, which this time determines the frequencies of a pure-tone complex with f_1 fixed at 2,800 Hz. The amplitude of the pure-tone complex is again determined by amplitude following the audio excerpt.

Example Group 4: Single-Sideband Modulation

Combination tones are produced by single-sideband modulation applied to source signals with a carrier frequency, f_1, set at 2,800 Hz. Example 4a is the same Jean Ritchie recording used in the Example Group 3. Example 4b is an original source recording of a short recitation of "Kubla Khan" by Samuel Coleridge. Example 4c is the Coleridge excerpt with single-sideband modulation applied. Example 4d is single-sideband modulation applied to the opening of Gustav Mahler's Tenth Symphony (Mahler 1990).

Example Group 5: Analytic Solution with Three Harmonics

A sinusoidal analysis of the source signal is used to synthesize combination tones that match the recorded data through an analytic solution. Example 5a is a recreation of the Jean Ritchie excerpt with only the first three harmonics. Example 5b is the analytic solution for exactly synthesizing those three harmonic as combination tones. f_1 is fixed at 2,800 Hz. Example 5c is the same solution implemented at three different frequency levels. f_1 is set at 2,800, 5,400, and 8,055 Hz.

Example Group 6: Spatiality

These examples are most clearly heard in headphone reproduction, and the clarity of the results with loudspeaker reproduction will depend on the physical setting. In both Examples 6a and 6b the QDT has a pitch gliding from middle C to the octave, and in both cases the position of the QDT is heard to move back and forth between the left and right sides. In Example 6a this effect is created by a single tone that is panned back and forth between the left and right channels. This tone is produced by direct additive synthesis of 15 pure tones with f_1 set at 2,800 Hz. The position of the QDT roughly follows the acoustic tone. In 6b there are different acoustic tones in the left and right channels, each with 15 pure tones, but one with f_1 at 2,500 Hz and the other with f_1 at 3,267 Hz. Shifting the gain in the left and right channels in a manner that is analogous to panning, the QDT appears to move back and forth between the left and right channels again even though the left and right ear signals are completely different (dichotic listening). This illustrates that the QDT tone is itself a phenomenon of the peripheral auditory system, whereas lateralization is a phenomenon of the central auditory system.

Example 7: Christopher Haworth, Correlation Number One

"Correlation Number One" is a fixed media work that focuses on the quadratic difference tone as its primary musical material. It is originally an eight-channel piece and is provided here in a stereo mix. "Correlation Number One" uses direct additive synthesis with AM and FM extensions as described in the Synthesis Techniques section. Values for f_1 generally lie between 2.5 kHz and 3 kHz, and the fundamental, F, as well as amplitude envelopes for the pure tones, are supplied via global variables. In the original eight-channel version, "Correlation Number One" sends just one sinusoid to each output channel, rather like resynthesizing a signal after fast Fourier transform (FFT) analysis and sending each frequency component to a different speaker. The number of pure tones, N, is always a multiple of eight, the number of output channels. In some ways, the spatial effect is quite similar to FFT-based spectral spatialization, because what is perceived is not the individual components surrounding the listener in space, but a fused sound with a strong fundamental frequency, in this case, the QDT spectrum. The localization in this case, however, is quite different; as described in the section on Spatial Imagery, the distortion product spectrum localizes in or very close to the head, rather like headphone listening. Another important difference is that decorrelation applied to the pure tones has little effect on stream segregation. Whereas modest decorrelation in spectral spatialization would cause the apparent fusion of the source to break down into individual components, with auditory distortion synthesis this affects only the timbre of the QDT.

Appendix 2: Calculating Quadratic Difference Tones

Let us model a sound signal $x(t)$ as the sum of N partials of frequencies f_1, $f_1+ F$, f_1+2F, ..., $f_1 + (N-1)F$ with amplitudes $A_1, A_2, A_3, \ldots, A_N$. Without loss of generality, we consider all partials with zero phase.

We therefore use the following expression for $x(t)$:

$$x(t) = \sum_{k=0}^{N-1} A_{K+1} \cos((f_1 + kF)t) \qquad (9)$$

We want to study the effects of a quadratic transformation to $x(t)$, such as:

$$y(t) = \alpha \left(\sum_{k=0}^{N-1} A_{K+1} \cos((f_1 + kF)t) \right)^2 \qquad (10)$$

We are particularly interested in the difference tones that appear as a result of the quadratic process. These tones have frequencies $F, 2F, \ldots, nF$. If $x(t)$ contains N harmonics, we know that $y(t)$ will contain $N-1$ terms at frequency F, $N-2$ at frequency $2F$, $N-3$ at frequency $3F$, and so on. We would like to calculate the N amplitude coefficients A_k that are needed to generate a predetermined $s(t)$ signal, consisting only on the difference tones and their harmonics present in $y(t)$. We know that $s(t)$ will contain $N-1$ harmonics, with fundamental frequency F, and amplitudes $T_1, T_2 \ldots T_{N-1}$. In order to do this, we need to find equations that relate the N amplitude coefficients A_k of $x(t)$ with the $N-1$ amplitude coefficients T_k of $s(t)$.

As a first approximation, let assume $N = 2$, and $\alpha = 1$. Then, we obtain the following equation:

$$y(t) = (A_1 \cos(f_1 t) + A_2 \cos(f_1 + F))^2 \qquad (11)$$

Expanding and rearranging, we get:

$$y(t) = \frac{A_1^2}{2} + \frac{A_2^2}{2} + \frac{1}{2} A_1^2 \cos(2 f_1 t) + \frac{1}{2} A_2^2 \cos(2(f_1 + F)t)$$
$$+ A_1 A_2 \cos(F t) + A_1 A_2 \cos((2 f_1 + F)t) \qquad (12)$$

From this last equation, it is clear that for $N = 2$ we obtain only one difference tone of frequency F, with amplitude

$$T_1 = A_1 A_2 \qquad (13)$$

In the case $N = 3$, we have:

$$y(t) = \frac{A_1^2}{2} + \frac{A_2^2}{2} + \frac{A_3^2}{2} + \frac{1}{2} A_1^2 \cos(2 f_1 t)$$
$$+ \frac{1}{2} A_2^2 \cos(2(f_1 + F)t) + \frac{1}{2} A_3^2 \cos(2(f_1 + 2F)t)$$

Kendall et al. **21**

$$+ (A_1 A_2 + A_1 A_3) \cos(Ft) + A_1 A_3 \cos(2Ft)$$
$$+ A_1 A_2 \cos((2f_1 + F)t) + A_1 A_3 \cos((2f_1 + F)t)$$
$$+ A_2 A_3 \cos((2f_1 + 3F)t) \tag{14}$$

Analyzing this expression and considering the amplitudes for the difference tones, we can form the following system of equations:

$$T_1 = A_1 A_2 + A_2 A_3$$
$$T_2 = A_1 A_3 \tag{15}$$

For $N = 4$ we can proceed on a similar way and we will obtain the following system:

$$T_1 = A_1 A_2 + A_2 A_3 + A_3 A_4$$
$$T_2 = A_1 A_3 + A_2 A_4 \tag{16}$$
$$T_3 = A_1 A_4$$

Generalizing, if we consider N components in $x(t)$, $s(t)$ will contain N–1 harmonics. The amplitudes T_k of the harmonics of $s(t)$ can be calculated as follows:

$$T_1 = A_1 A_2 + A_2 A_3 + A_3 A_4 + \cdots + A_{N-3} A_{N-2}$$
$$+ A_{N-2} A_{N-1} + A_{N-1} A_N \tag{17}$$
$$T_2 = A_1 A_2 + A_2 A_3 + A_3 A_4 + \cdots + A_{N-3} A_{N-1}$$
$$+ A_{N-2} A_N$$
$$T_3 = A_1 A_2 + A_2 A_3 + A_3 A_4 + \cdots + A_{N-3} A_N$$
$$\vdots$$
$$T_{N-3} = A_1 A_{N-2} + A_{N-2} A_{N-1} + A_3 A_N$$
$$T_{N-2} = A_1 A_{N-1} + A_2 A_N$$
$$T_{N-1} = A_1 A_N$$

As there are N–1 equations for N variables, the system as such is under determined. We can then fix one of the coefficients in order to solve the problem, consequently we arbitrarily make $A_1 = 1$. The system then becomes:

$$T_1 = A_2 + A_2 A_3 + A_3 A_4 + \cdots + A_{N-3} A_{N-2}$$
$$+ A_{N-2} A_{N-1} + A_{N-1} A_N \tag{18}$$

$$T_2 = A_2 + A_2 A_3 + A_3 A_4 + \cdots + A_{N-3} A_{N-1}$$
$$+ A_{N-2} A_N$$
$$T_3 = A_2 + A_2 A_3 + A_3 A_4 + \cdots + A_{N-3} A_N$$
$$\vdots$$
$$T_{N-3} = A_{N-2} + A_{N-2} A_{N-1} + A_3 A_N$$
$$T_{N-2} = A_{N-1} + A_2 A_N$$
$$T_{N-1} = A_N$$

Now we have N–1 equations and N–1 variables, which means that, in theory, we should be able to determine all A_k coefficients in terms of the T_k coefficients alone. The solution to this system is not as simple as it may seem. Indeed, we will show that it cannot be solved algebraically for every N. The simplest approach for solving this system is by expressing in each equation a coefficient A_k in terms of A_2 and T_k coefficients, and then replacing the obtained expression in the upper equation.

The procedure as follows: We begin by taking the last equation that relates the last coefficients of $s(t)$ and $x(t)$, substituting A_N for T_{N-1} and then proceeding upwards. Each time we go up, we increase the order of the problem. The first substitution yields a linear equation in A_2, but the second substitution results in a quadratic equation for A_2. Once we travel the system all the way up, we reach the first equation of the system, which will be a N–2 polynomial in A_2, because N–2 substitutions were made.

Let us now illustrate this procedure for the case $N = 4$, previously expressed in the system of equations (16). Substituting the last equation into the second we obtain:

$$T_2 = A_3 + A_2 T_3 \tag{19}$$

this yields:

$$A_3 = A_2 T_3 - T_2 \tag{20}$$

Substituting this equation into the first equation of the system results

$$T_1 = A_2 + A_2(A_2 T_3 - T_2) + (A_2 T_3 - T_2)T_3 \tag{21}$$

Expressing this equation as a polynomial in A_2 we obtain:

$$A_2^2 + \frac{A_2(1 - T_2 + T_3^2) - (T_2 T_3 - T_1)}{T3} = 0 \qquad (22)$$

which is a second-order polynomial equation in A_2, which is what we expected.

We can now calculate the complete solution to system (17), which is:

$$A_1 = 1$$

$$A_2 = \frac{\frac{T_2}{2} \pm \frac{\sqrt{T_2^2 + 2T_2 T_3^2 + 2T_2 + T_3^4 - 2T_3^2 - 4T_1 T_3 + 1} - T_3^2 + 1}{2}}{T_3}$$

$$A_3 = \frac{T_2}{2} \pm$$

$$\frac{\sqrt{T_2^2 + 2T_2 T_3^2 + 2T_2 + T_3^4 - 2T_3^2 - 4T_1 T_3 + 1} + T_3^2 - 1}{2}$$

$$A_4 = T_3 \qquad (23)$$

Solving the general system of equations (18) is highly dependent on N, as it involves solving a polynomial of order $N-2$. For a large N, this is not only computationally challenging, it is indeed unsolvable if $N > 6$. Abel's impossibility theorem states that, in general, polynomial equations higher than fourth degree are incapable of algebraic solutions in terms of a finite number of additions, subtractions, multiplications, divisions, and root extractions operating on the coefficients (Cheney and Kincaid 2009, pp. 705). This does not mean that high-degree polynomials are not solvable, because the fundamental theory of algebra guarantees that at least one complex solution exists. What this really means is that the solutions cannot be always expressed in radicals.

Therefore, as seeking an algebraic expression for any N is impractical; if we want to specify $s(t)$ with more than four harmonics by calculating the coefficients of $x(t)$, the only way of doing that is by numerical methods such as the Newton-Rhapson, Laguerre, or the Lin-Bairstrow algorithm (Rosloniec 2008, pp. 29–47). This is the main reason why a numerical rather than an algebraic solution is the correct approach for this problem when a high number of harmonics in the target signal is desired.