Spectral Analysis for the Ear

Matthew J. Goupell and William M. Hartmann

Department of Physics and Astronomy
Michigan State University
hartmann@pa.msu.edu

Abstract

A central theme in elementary acoustics is that complex sounds consist of pure-tone spectral components and can be analyzed as such. To make this concept plausible it is helpful to demonstrate it in a way that has immediate appeal to the ear. The following article discusses a method for making components evident to the ear – the Cancelled Harmonics technique - wherein a harmonic of a complex tone is repeatedly removed and reinserted, causing the harmonic to stand out. This effect is known to psychoacousticians as "enhancement." Although previous demonstrations of enhancement have focused on low-order harmonics, the present research demonstrates an effect up to the 69th harmonic. It is shown that this limit is set by absolute threshold when the fundamental frequency is high. It is probably set by masking when the fundamental frequency is low.

1. Introduction

"Cancelled Harmonics" is demonstration number one on the IPO-NIU-ASA compact disc entitled *Auditory Demonstrations* [1]. In this demonstration, a harmonic of a complex tone is alternately turned off (cancelled) and turned on. As a result the manipulated harmonic is heard as an independent entity standing out from the complex-tone background. Rather surprisingly, this harmonic can often be heard for an indefinitely long time after finally being turned on. Helmholtz [2] quotes Seebeck's remark that the duration depends on the "... liveliness of our recollection of the tones heard separately."

By this technique, many harmonics of a complex tone can be made individually audible, providing dramatic perceptual confirmation that Marin Mersenne [3] was right about the presence of harmonics in complex tones. By contrast, normal listening to the complex tones of speech or music does not reveal individual harmonics. Instead, the harmonics are collectively absorbed into the global property of tone color or timbre.

The Cancelled Harmonics demonstration on the CD presents a complex tone with a fundamental frequency of 200 Hz and 20 harmonics of equal amplitude. In succession, the demonstration exposes the first ten harmonics by turning them off and on in a seven-interval sequence. Thus the maximum harmonic frequency exposed in this way is $10 \times 200 = 2000$ Hz.

The research described in this report was performed to answer the question, "How high can the enhanced harmonic effect go?" The answer is important to a fundamental understanding of the effect. Possibly the effect depends on spectral resolution. The harmonic identification task of Plomp and Mimpen [4] found that harmonics could be resolved up to the fifth. *A priori* it seemed possible that the enhanced harmonic demonstration is just a less demanding version of the same thing - capable of exposing harmonics as high as the tenth because they are, at least to some extent, resolvable. The tenth harmonic was recently identified as the limit in the resolution and pitch experiments of Bernstein and Oxenham [5]. That view would suggest that the enhanced harmonic effect might not work much beyond the tenth. Alternatively, the effect may depend upon neural synchrony, as suggested by mistuned harmonic experiments. An individual harmonic can be made audible by mistuning it slightly. Exposing harmonics in this way appears to depend on neural synchrony [6], and the effect disappears somewhere between 2 and 3 kHz.

2. Experiment 1 – Highest audible harmonic

Experiment 1 followed the format of the compact disc's Cancelled Harmonics demonstration, making a harmonic separately audible by cycling it off and on in a seven-interval sequence. The listener was required to say whether he heard a harmonic standing out.

2.1. Stimulus

As with the CD demonstration, the tone, with fundamental frequency f_0, was continuous. Only the single manipulated harmonic was cycled. The tone consisted of all harmonics, with equal amplitude, up to 20,000 Hz. The phases of the harmonics were randomized on each presentation to avoid a persistent special case.

The complex tone was computed in a Tucker-Davis array processor (AP2), converted by a 16-bit DAC (DD1), and lowpass filtered with a corner frequency of

20 kHz and a rolloff of -115 dB/octave. The duration of one of the seven intervals was about 1.3s making the seven-interval sequence duration about 9.2s.

In different experimental runs the nominal fundamental frequency was one of 17 different ISO one-third-octave frequencies: 50, 63, 80, 100, 125, 160, 200, 250, 315, 400, 500, 630, 800, 1000, 1250, 1600, and 2000 Hz.

The stimulus was presented diotically at a level of 40 dB SPL for each component through Etymotic ER2 ear insert phones. The ear-insert phones provided an acceptably flat frequency response at the high frequencies of interest in this experiment. Listeners were seated in a double-walled sound attenuating room.

This experiment was performed on five male listeners: A, M, N, W, and X. All but Listener W were between the ages of 21 and 25 and had normal hearing; Listener W was 63 and suffered from mild bilateral hearing loss. Listeners M and W were the authors.

2.2. Experimental procedure

An experimental run was dedicated to finding the highest audible enhanced harmonic for a single fixed fundamental frequency. The task was Yes-No. For each value of the manipulated harmonic number the listener responded that a harmonic was or was not audible above the background complex tone. Fixing the fundamental frequency for an entire run helped establish a good sense of the background. Thus 17 runs were required for a data set that included all the ISO fundamental frequencies. Three data sets were collected for each listener.

During the course of a run the experimenter chose the harmonic numbers of the manipulated harmonics with the goal of finding the highest audible harmonic accurately and efficiently. The experimenter concentrated stimuli in the region of the listener's uncertainty while also exploring a wide range looking for surprises. The highest audible harmonic was found when the listener responded “No” for several consecutive values of manipulated harmonic number above the highest value for which a “Yes” response had been obtained.

The protocol also required that not fewer than 10% of the trials should be “catch” trials in which no harmonic was manipulated. In fact, 14.6% of the trials were actually catch trials, and in the end there were only four false alarms among those 390 trials, 3 for listener W and 1 for listener N. The small number of false alarms gives one confidence in this experiment.

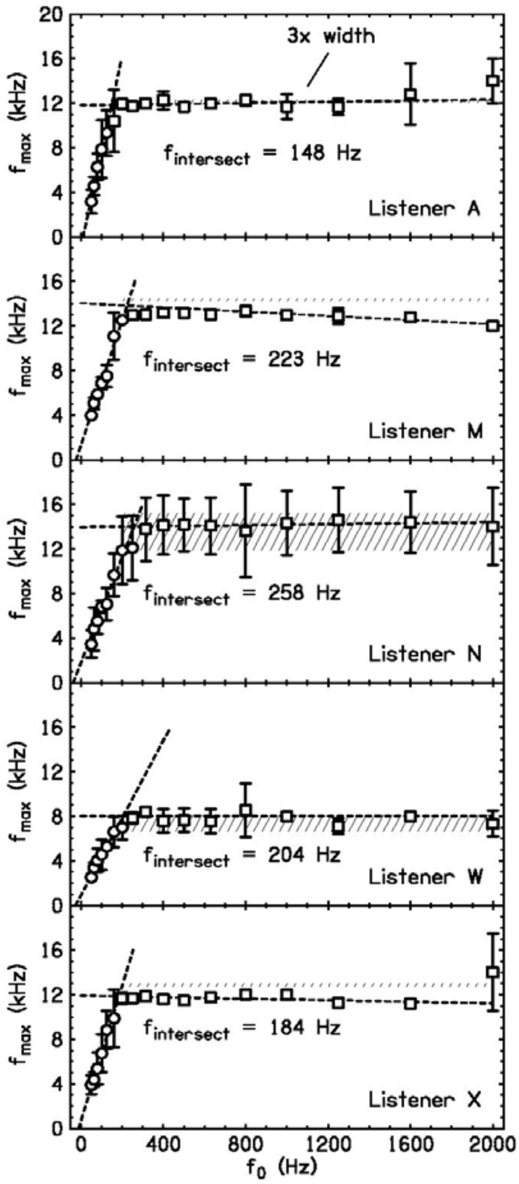

Figure 1: The frequency of the highest audible harmonic, f_{max}, is given as a function of the fundamental frequency, f_0, for five listeners by open symbols (Experiment 1). Error bars are two standard deviations in overall length. Circles and squares are used to indicate data points that were assigned to different straight lines by the fitting algorithm. These straight lines are shown dashed. Hatched regions are centered on the highest frequency for a barely audible sine tone with a level of 40-dB SPL, as determined by a Békésy tracking program (Experiment 2). The width of the hatched regions is the frequency range for a level variation of ± 2 dB about 40 dB. The region for listener A is so narrow that its width has been multiplied by 3 to make it visible in this figure.

2.3. Results

Figure 1 shows the highest audible harmonics for all five listeners as a function of fundamental frequency f_0. The open symbols show the mean value of the three data sets, and the error bars have an overall length of two unbiased ($N-1 = 2$) standard deviations. For listeners A, M, and X, the error bars were larger for $f_0 \leq$ 200 Hz than for $f_0 >$ 200 Hz.

For each listener, the data in Fig. 1 seemed to fall on two straight lines. Therefore, a two-line least-squares fit was performed on the data, as shown by the dashed lines in Fig. 1. The choice of the number of points fitted with each line minimized the sum of the two squared errors. The process weighted a data point with the reciprocal of its variance because different points had different variance. This fitting procedure accounted for 96% of the overall variance (for the different listeners between 94 and 98%).

The actual slopes of the high-frequency fitted lines and the actual intercepts of the low-frequency lines were all so close to zero that it proved worthwhile to run the fitting procedure again with the constraints that the low-frequency line goes through the origin and the high-frequency line has zero-slope. That fit accounted for 95% of the variance, which, compared to 96%, means that little was lost by sacrificing two of the four free parameters. The slope of the low-frequency line corresponds to an average highest harmonic number of 64, or 69 if only normal-hearing listeners are included. The value of $f_{intersect}$, where the two lines intersect, is rather similar for all listeners, ranging from a fundamental frequency of 148 Hz to 258 Hz with a mean of 204 Hz.

3. Experiment 2 – High-frequency threshold

Fig. 1 shows that the highest audible enhanced harmonic could be well represented as a constant frequency, f_{max}, for $f_0 > f_{intersect}$. It seemed likely that this constant frequency was actually the absolute threshold. To check this idea a Békésy tracking experiment was done to compare f_{max} to the listeners' absolute threshold. Of interest was the threshold frequency for a 40-dB sine tone, the level of a single component manipulated in Experiment 1.

3.1. Experimental procedure

A Békésy audiometer presented a sequence of pulsing sine tones – 400 ms on and 400 ms off, with 30-ms onset and offset ramps. The tone level changed at a rate of ±1 dB per pulse, and the frequency increased exponentially with time. The frequency range was chosen for each listener to include the frequency where threshold was 40 dB, the level of the manipulated harmonics in Experiment 1. The same ER2 phones and environment were used as in Experiment 1. The total duration of the track was six minutes.

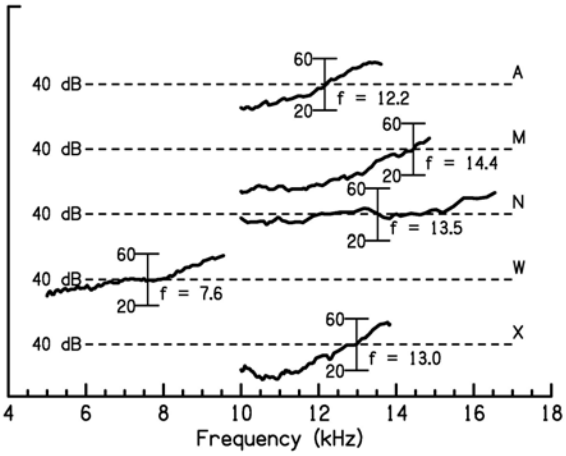

Figure 2: *Békésy threshold tracks for a sine tone in quiet. The 40-dB crossing point is shown by the dashed horizontal line and the solid vertical line.*

3.2. Results and discussion

The results of Experiment 2 are given in Fig. 2, which shows the midpoint of successive turnarounds in the entire track for each listener. The frequency of the 40-dB crossing is also shown on the graph. Except for listener W, the 40-dB crossing frequency was between 12 and 15 kHz. Listener W's lower frequency of 8 kHz was indicative of his hearing loss. The frequency range for thresholds between 38 and 42 dB SPL (40 dB ± 2 dB) was also found. This 40 dB ± 2 dB range of frequencies is shown by the hatched regions in Fig. 1.

If our conjecture is correct, the highest value of f_{max} from Experiment 1 ought to agree with the absolute threshold frequencies for the level range of 40 ± 2 dB, as measured in Experiment 2. For listeners A, N and W, f_{max} agrees with the absolute threshold frequency range. For listeners M and X the frequency of the highest f_{max} is about 10% less than the threshold frequency. The discrepancy is not disturbing because the detection tasks in Experiments 1 and 2 were rather different. Listeners M and X had no false alarms in Experiment 1. Their criterion, β, was immeasurably low. Therefore, one expects a more conservative value for the highest audible frequency (Experiment 1) than obtained from the midpoint of a Békésy track (Experiment 2).

The uncertainties in Experiments 1 and 2 also correlate well. Figure 2 shows that the audiogram tracks for listeners N and W have a small slope at 40 dB leading to a wide hatched range in Fig. 1. As might be expected, the error bars in Experiment 1 are large for these listeners. Above $f_{intersect}$ the error bars average 2955 Hz for N and 822 Hz for W. By contrast, the audiogram tracks for A, M, and X are steep near 40 dB, the hatched regions are narrow, and it also turns out that the error bars in Experiment 1 are small, averaging 568 (sd=371) Hz for these three listeners.

Together, the comparison of Experiments 1 and 2 gives good evidence that for $f > f_{intersect}$ ($f > 200$ Hz) the limiting factor for the highest audible harmonic is simply the threshold in quiet. This result is consistent with critical band theory because for high values of f_0 there are few components in any critical band and little masking of high harmonics.

4. Discussion

The rapid decrease in highest harmonic frequency for $f_0 < 200$ Hz suggests that when harmonics become dense they interfere with the detection of the manipulated harmonic. The constant highest harmonic *number* result obtained for $f_0 < f_{intersect}$ suggests that the interfering (masking) harmonics enter through a constant Q filter – the higher the target frequency, the more widely the harmonics need to be spaced to allow detection. The auditory filter model indicates an effective signal to noise ratio by the following argument.

Suppose that an auditory filter bandwidth Δf is one-third of an octave. Then, the bandwidth containing the signal of interest is

$$\Delta f = 0.232 f \qquad (1)$$

where f is the frequency of the manipulated harmonic. It follows that the number of harmonics of f_0 in an one-third-octave band around the maximum audible harmonic is $n_c = 0.232 n_{max}$. In words, the number of components in a one-third-octave auditory filter is about 23 percent of the maximum audible enhanced harmonic number.

For $n_{max} = 69$, as found for normal listeners in Experiment 1, $n_c = 16$. Of these 16 harmonics in the band, one is the target and the others are neighbors that serve as masking noise. Thus the signal to noise ratio is S/N = 10 log (1/15) = –11.8 dB.

5. Summary

The enhanced harmonic effect, whereby an individual harmonic of a complex tone is made audible by turning it off and on, has been extended to high frequencies. Two complementary experiments, using tones with harmonics of equal amplitude, show that if the fundamental frequency is greater than 200 Hz the frequency of the highest audible harmonic corresponds to absolute threshold - higher than 10 kHz for normal hearing listeners.

If the fundamental frequency is less than 200 Hz the harmonic number of the highest audible harmonic is constant, as predicted by critical band theory with constant-Q auditory filters. The highest audible harmonic number is near 69. Although experiments like those of Plomp and Mimpen [4], Bernstein and Oxenham [5] and Hartmann et al. [6] expose harmonics below the tenth or below 2000 Hz, the enhanced harmonic effect is capable of exposing harmonics that are much higher, with frequencies as high as the highest audible frequency in quiet.

6. Acknowledgements

We are grateful to Peter Xinya Zhang for serving as experimenter while the first author was a listener. This work was supported by the NIDCD of the NIH under grant DC 00181.

7. References

[1] Houtsma, A.J.M., Rossing, T.D., and Wagenaars, W.M. *Auditory Demonstrations*, compact disc, Acoustical Society of America, Philips Eindhoven, The Netherlands, 1987.

[2] Helmholtz, H.L.F *On the Sensations of Tone*, fourth edition, tr. A.J Ellis (1885), Dover edition of 1954 p. 61, 1877.

[3] Mersenne, M. *Traité des instrumens*, Sebastian Cramoisy, Paris, book IV, Prop. 11, 208-9, 1636.

[4] Plomp, R. and Mimpen, A.M. “The ear as a frequency analyzer II,” J. Acoust. Soc. Am. **43**, 764-767, 1968.

[5] Bernstein, J.G. and Oxenham, A.J. “Pitch discrimination of diotic and dichotic tone complexes: Harmonic resolvability or harmonic number?” J. Acoust. Soc. Am. **113**, 3323-3334, 2003.

[6] Hartmann, W.M., McAdams, S., and Smith, B.K. “Matching the pitch of a mistuned harmonic in an otherwise periodic complex tone,” J. Acoust. Soc. Am. **88**, 1712-1724, 1990.