INFLUENCE OF FUNDAMENTAL WAVEFORM ON PERCEIVED VOLUME AND PITCH IN AUDIBLE GRAPHS

Dr. Robert Shelton

NASA Learning Technologies Regional Center Project Manager
Education Office
NASA Johnson Space Center
2101 NASA Parkway
Houston, TX. 77058
robert.o.shelton@nasa.gov

ABSTRACT

This paper will describe an issue that arose in the implementation of graph sonification in an accessible mathematics product developed by the NASA Learning Technologies team at the Johnson Space Center in Houston Texas. The software provides text descriptions and audible displays of 2-dimensional graphs of equations, raw data, or simulations, and these audible displays make use of pitch to indicate the height component(s) of the curve. It is therefore important for us to provide the clearest possible pitch cue over a wide (4 octave) spectrum. One significant barrier to use of a relatively wide frequency range is the fact that pure tones are attenuated significantly in the lower frequencies by inexpensive sound systems common on personal computers. In this paper, we describe several approaches to the problem of choosing waveforms which clearly communicate pitch while remaining audible at low frequencies. We end by describing our compromise solution and by offering researchers an applet-based tool for further experimentation.

1. INTRODUCTION

The anatomy of the audible display for many applications which use sound to represent visual information [1], [2] and *MathTrax* [3], is based on the conventional "tone graph" wherein height (or heights) are indicated by pitch and left-right position is indicated by time and stereo panning. In the case of *MathTrax*, the visible display is swept left to right over time. A visible vertical bar tracks across the screen as the reference point moves from left to right. Where the bar touches the graph (always at most a single touch for functions, multiple touches for graphs with multiple branches), one or more tones is generated, and the pitch of the tone indicates the distance of the touch above or below the center of the screen. We provide a review capability to allow the user to interactively explore the graph by moving the reference bar left or right with keyboard commands. There is a window that displays numerical values which can be reviewed at any time, thus the interactive mode allows the user to find an area of interest by exploring with ears and fingers, and then to read off the particulars from the data window.

A design consideration in any such scheme should be to provide a clear pitch clue over a range of frequencies necessary to convey the dynamic range of the data in question. The frequency range of human hearing was documented and quantified as the classical Fletcher-Munson loudness curves, [4] which are relatively flat over a usefully wide range; however, the response of inexpensive speakers or headphones of the type likely to be used is in general, non-linear, and unpredictable, thus attempts to compensate by simply adjusting power as a function of frequency are not likely to succeed. We explore the influence of the shape of the fundamental periodic waveform on perceived volume and pitch, through harmonic analysis and experimentation.

Theoretically, we know that deviation from a pure sine wave at the fundamental frequency necessarily introduces harmonic multiples of the fundamental. As we will see, it is often much easier for speakers to radiate low frequencies when the fundamental waveform contains some sharp edges; however, the question is how to introduce roughness into the fundamental without disturbing the listeners' perception of the fundamental pitch with extraneous harmonics. In the end, this is a complicated question which suggests an experimental approach, and to this end, we provide an applet [5]

1.1. Pitch and Frequency

The ear registers frequency in a complicated way. The sensation of pitch can register tiny changes in frequency, however the ability to determine absolute frequency – perfect pitch – is rare. We register changes in frequency on a logarithmic scale, thus we associate an exponential mapping of pitch to frequency given by the equation

$$f = f_0 2^p \tag{1}$$

where f is the frequency in Hertz, and p is the number of octaves away from a reference pitch f_0.

There is a tendency to identify tones which are a whole number of octaves apart, and this partial wrapping effect is the basis of the famous Shepard tone acoustical illusion as well as many others. For further discussion, of acoustical illusions see [6]

1.2. The Problem

In our first naive attempt to produce an audible graph, we used a simple sine wave as the tone used to indicate the height of the graph above a reference line. We chose an interval from 110 Hz to 1760 Hz, musically, the four octave range centered at the A above middle C. In this range of frequencies, the ear has excellent sensitivity to pitch; however, the optimum region for volume sensitivity is more in the 3 KHz range. Even more pronounced than the ear's typical loss of sensitivity and perceived drop in volume for signals below 200Hz is the extreme attenuation of the acoustical power radiated at low frequencies by speakers typically found in personal computers. We offer a brief "back of the envelope" analysis to describe this effect.

Imagine that we wish to play a periodic signal $s(t)$ given by

$$s(t) = aG(\omega t) \tag{2}$$

where a is the amplitude, G is a periodic function with period equal to 2π, and ω is the radian frequency. The simplest case is that of a pure sine wave, i.e. $G(t) = \sin(t)$. The frequency N in Hertz is given by

$$N = \frac{\omega}{2\pi} \tag{3}$$

It is simple to calculate the power P_s in the signal by averaging the square of s over a suitably large interval of time.

$$\begin{gathered} P_s = \lim_{T\to\infty} \frac{\int_0^T s^2(\tau)d\tau}{T} = \lim_{T\to\infty} \frac{\int_0^T a^2G^2(\omega\tau)d\tau}{T} = \\ \lim_{T\to\infty} a^2 \frac{T\frac{\int_0^{2\pi} G^2(\tau)d\tau}{2\pi} + \rho(T)}{T} = \\ a^2 \frac{\int_0^{2\pi} G^2(\tau)d\tau}{2\pi} = a^2\epsilon_0 \end{gathered} \tag{4}$$

where ρ is some bounded periodic function and ϵ_0 is the average of the square of G over a single period – the energy of the fundamental waveform. Note that the power of the signal is independent of the frequency ω.

1.3. Behavior of a Typical Speaker

We can use some simple physics to estimate the work done on the air by the speaker, and thus derive the radiated power. For a flat object (such as a speaker cone) moving through air, the drag force on the cone is proportional to the square of its velocity through the air. The work done through an infinitesimal distance dx by a force F is given by

$$dW = Fdx \tag{5}$$

and thus the power P_r radiated by the vibrating cone is given by

$$P_r = \lim_{T\to\infty} \frac{\int_0^T |F(\tau)v(\tau)|d\tau}{T} = \lim_{T\to\infty} \frac{\int_0^T |\kappa v^3(\tau)|d\tau}{T} \tag{6}$$

where κ is the constant of proportionality relating drag force to the square of the velocity. Note that we insert absolute values because we dissipate energy independent of whether the cone is moving forward or backward.

If we make the assumption that the displacement of the cone $d(t)$ at a given frequency ω is proportional to the incoming signal, then we have

$$d(t) = \alpha(\omega)G(\omega t) \tag{7}$$

where $\alpha(\omega)$ is the amplitude of the mechanical vibration of the speaker cone at frequency ω and G is the basic waveform from 2. We can differentiate 7 to obtain the velocity $v(t)$ as

$$v(t) = \frac{d}{dt}\alpha(\omega)G(\omega t) = \omega\alpha(\omega)G'(\omega t) \tag{8}$$

and combining with 6 we obtain

$$\begin{gathered} P_r = \lim_{T\to\infty} \frac{\int_0^T \kappa\alpha^3(\omega)\omega^3|G\prime(\omega\tau)|^3 d\tau}{T} = \\ \kappa\alpha^3(\omega)\omega^2 \frac{\int_0^{2\pi} |G\prime(\tau)|^3 d\tau}{2\pi} = \\ \alpha^3(\omega)\omega^2\epsilon_1 \end{gathered} \tag{9}$$

Where ϵ_1 is the frequency-independent constant $\kappa\frac{\int_0^{2\pi} |G\prime(\tau)|^3 d\tau}{2\pi}$.

Designers of audio equipment usually design speakers to exhibit a "flat" response curve in order to faithfully reproduce a sound corresponding to their input signal. If we compare P_s and P_r as described by 4 and 9, then we can see how the amplitude of cone displacement α must vary with frequency ω in order to maintain a flat response curve. If we set $P_r = K_0P_s$, then we get

$$\alpha^3(\omega)\omega^2\epsilon_1 = K_0\epsilon_0 a^2 \tag{10}$$

therefore

$$\alpha(\omega) = K_1\omega^{-\frac{2}{3}} \tag{11}$$

where we absorb all of the frequency-independent factors into a single constant K_1. Keeping the frequency response flat in a single speaker would force the allowable amplitude of the cone vibration to expand without bound for low frequencies, and in fact, this effect is well known to anyone who has ruined a loud speaker by driving it with bass boosted beyond the design limits of the speaker. Typical PC speakers have a very small maximum displacement and thus effectively cut off frequencies which are still well within the usable audio spectrum. We can estimate how this works by revisiting 10, replacing $\alpha(\omega)$ with α_0 and K_0 with an attenuation function $K(\omega)$ and write

$$K(\omega) = \frac{\epsilon_1\alpha_0^3\omega^2}{a^2\epsilon_0} = K_2\omega^2 \tag{12}$$

If we identify the frequency ω_0 at which the speaker achieves its maximum displacement α_0, and the corresponding pitch p_0, then we observe that the decibel attenuation imposed by the speaker for frequencies below ω_0 is given by

$$\begin{gathered} \delta(p) = 10\log_{10}\left(\frac{K(\omega)}{K(\omega_0)}\right) = \\ 10\log_{10}\left(\frac{\omega^2}{\omega_0^2}\right) = 20\log_{10}(2^p) = \\ 20\log_{10}(2)p \end{gathered} \tag{13}$$

where $\delta(p)$ is the number of decibels of attenuation relative to the frequency ω_0 and p is the relative pitch in octaves below the pitch p_0 corresponding to ω_0. Note that 13 indicates that the perceived drop in volume will be a linear function of pitch, and the constant of proportionality $20log_{10}(2)$ is approximately equal to 6.0206. This means that you lose a little over 6 db's of volume for every octave below ω_0. This "back-of-the-envelope" analysis ignores some important factors which actually cause the drop-off to be even steeper. One assumption was that the motion of the speaker cone will remain a linear multiple of the original signal for frequencies below ω_0 when in reality, the tops and bottoms of the waveforms will be flattened which results in less transmission of energy to the surrounding air. It is also clear that simple power compensation will fail because the attenuation depends on the physical characteristics of the speaker, thus pumping more power in at low frequencies will only produce a distorted waveform and will be dissipated as heat rather than radiated as sound.

Despite this well known analysis, this rapid attenuation at low frequencies is somewhat counterintuitive. We are used to hearing sounds such as male voices or a low note from a trombone or tuba rendered rather faithfully through small speakers. One reason for this is that the foregoing analysis does not model power transmitted by discontinuities in the signal. We close this section with a quick look at the special case of a square wave. Define the function $H(t)$ as follows:

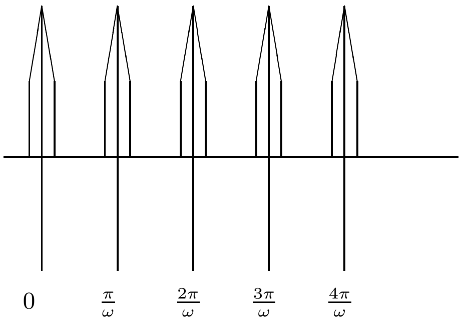

Figure 1: Power Dissipated by a Square Wave

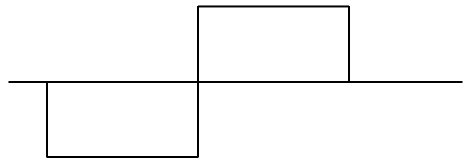

Figure 2: Square Wave

$$H(t) = 1 \text{ if } \left[\frac{t}{\pi}\right] \text{ is even and 0 otherwise} \tag{14}$$

where $[x]$ indicates the greatest integer less than or equal to x. The expression for radiated power derived in the linear case in 9 indicates a dependence on the cube of the derivative of the underlying function, in this case H. However, H is discontinuous, and thus technically, has no derivative at integral multiples of π, and it's derivative is 0 elsewhere. What will actually happen is that the signal will not really change instantaneously, nor will the cone position, but these electromechanical systems will react as quickly as they can, moving from the negative extreme $-\alpha_0$ to the positive α_0. The magnitude of the velocity of the cone $|v|$ will therefore describe a sequence of pulses as shown in 1, and since the height and width of these pulses are determined by the square wave response of the electronics and speakers, they do not depend on the frequency. The only dependence on frequency will be the spacing of the pulses which will determine the average radiated power. Note that we are merely assuming loudness to be correlated with average power. If we consider peak power, then the presence of $|v|^3$ in 6 suggests that peak power decays with the cube of ω in the smooth case; it is independent of ω in the discontinuous case.

2. WAVEFORMS

We now consider several simple waveforms and analyze their behavior as pitch cues. The easiest place to start is the square wave shown in 2. Experimentation with this waveform through tiny laptop speakers lends credence to the theory that perceived loudness is well correlated to peak as opposed to average power. The volume of the tone holds up exceptionally well at low frequencies, probably because the spectrum being registered by the ear has little to do with the fundamental frequency in the far low range, The perception is more like a series of clicks than an actual tone. Unfortunately, this frequency corruption pervades the whole usable range. The effect is like trying to sense the pitch of a basketball halftime buzzer. To better understand, consider the harmonic series for H. Since H is an odd function, i.e. $H(t) = -H(-t)$, we can write H as a sine series as follows:

$$H(t) = \sum_{n=1}^{\infty} b_n \sin(nt) \tag{15}$$

where

$$b_n = \frac{1}{\pi}\int_0^{2\pi} H(t)\sin(nt)dt = \frac{4}{n\pi} \text{ for odd } n \text{ and 0 otherwise} \tag{16}$$

or

$$H(t) = \frac{4}{\pi}\sum_{n=0}^{\infty} \frac{1}{2n+1}\sin(2n+1)t \tag{17}$$

The most noticeable feature of this expansion is that there is exactly one harmonic that is an octave of the fundamental frequency, namely, the fundamental itself. Also, the harmonics decay quite slowly as a function of n, and thus 3, 5 and 7 are clearly audible as "ringy" overtones. We will later use an applet to illustrate this as well as a surprising octave artifact which suggests that sensation of pitch may have more to do with counting than it does with harmonic analysis.

Now consider the sawtooth waveform W shown in 3 given by

$$W(t) = 2\left(\frac{t-\pi}{2\pi} - \left[\frac{t-\pi}{2\pi}\right]\right) - 1 \tag{18}$$

where $[x]$ indicates the largest whole number less than or equal to x.

It can be shown that W can be represented as the following sine series

$$W(t) = \frac{2}{\pi}\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n}\sin(nt) \tag{19}$$

Clearly, W contains more octave harmonics of the fundamental than does H; however, it is a remarkable fact that the total power of all octave harmonics is the same fraction of total power in both series. To show this, we can easily determine the fundamental octave power in H because there is only a single harmonic, namely the fundamental, which is an octave of the fundamental, consequently the fundamental octave power for H is found by integrating one half the square of the coefficient of $\sin t$ from 0 to 2π or

$$P_{oH} = 2\pi\frac{1}{2}\frac{16}{\pi^2} = \frac{16}{\pi} \tag{20}$$

The total power in H can be found by simply integrating the square of H (or 1) over the interval from 0 to 2π, thus

$$P_H = 2\pi \tag{21}$$

and the fraction $F_{oH} = \frac{P_{oH}}{P_H}$ is given by

$$F_{oH} = \frac{8}{\pi^2} \tag{22}$$

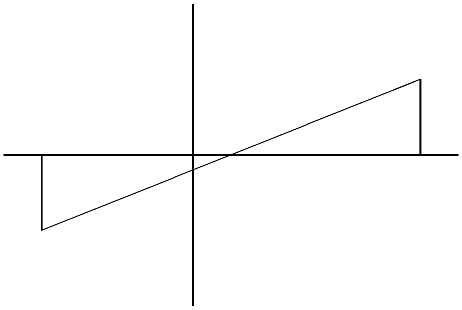

Figure 3: Sawtooth wave

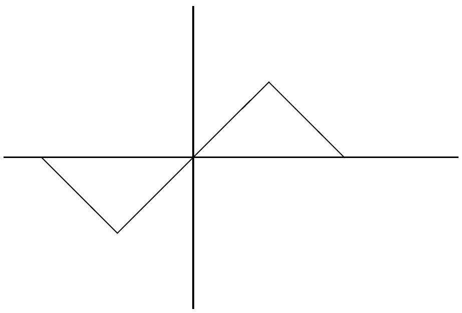

Figure 4: Triangular Waveform

Since W has infinitely many harmonics which are octaves of the fundamental, to find the fundamental octave power of W P_{oW}, we must sum a geometric series consisting of the squares of the coefficients for n a power of 2 as follows.

$$P_{oW} = 2\pi \frac{1}{2} \frac{4}{\pi^2} \sum_{j=0}^{\infty} 2^{-2j} = \frac{16}{3\pi} \tag{23}$$

As before with H, we determine the total power in W by integrating its square from 0 to 2π

$$P_W = \int_0^{2\pi} W^2(t)dt = \frac{2}{3}\pi \tag{24}$$

thus the fraction $F_{oW} = \frac{P_{oW}}{P_w}$ is given by

$$F_{oW} = \frac{\frac{16}{3\pi}}{\frac{2}{3}\pi} = \frac{8}{\pi^2} \tag{25}$$

which is identical to 22.

For the sake of completeness and comparison, we provide the Fourier expansion of the triangular form given by

$$T(t) = H(t) \left(1 - 2| \frac{t}{\pi} - \left[\frac{t}{\pi} \right] - \frac{1}{2} | \right) \tag{26}$$

The Fourier coefficients of T are given by

$$T(t) = \frac{8}{\pi^2} \sum_{n=0}^{\infty} \frac{(-1)^n \sin(2(n+1)t)}{(2n+1)^2} \tag{27}$$

Due to the continuity of T, the coefficients decay as $\frac{1}{n^2}$, and the tone has a correspondingly smoother sound than $H(t)$ or $W(t)$.

Although it does not survive at low frequencies as well as the square or saw tooth waves, the triangular waveform remains clearly audible (on laptop speakers) below 200Hz where the pure sine essentially disappears.

3. EXPERIMENTS

The foregoing analysis suggests the complexity of the problem and the need for a way to experiment with different waveform combinations, and to this end, we have provided an applet as a demonstration and tool. The URL for the applet is http://prime.jsc.nasa.gov/Applets/waveform.html. In addition to the visible controls, there are also some useful keyboard shortcuts. Complete instructions can be found on the web page for the applet.

One interesting clue to the way we process pitch is readily apparent by experimenting with the applet. When you start the applet, press the "Play" button. This starts a pure sinusoidal tone at 440Hz. Try pressing the "Square" button to replace the sinusoid with a square wave. You will notice that the pitch sounds higher as a square wave than with the corresponding sinusoid. In fact, you might even think that the square wave contains a strong dose of a pitch an octave above the fundamental; however, from the Fourier analysis, we know that there is no even harmonic of any sort in the square wave. Now consider the plot of the power dissipation of the square wave shown in 1. Note that the pulses occur at multiples of $\frac{\pi}{\omega}$, not $\frac{2\pi}{\omega}$, thus the discontinuities occur twice as frequently for a given pitch as they do in the case of the saw tooth wave. This provides a strong clue that the ear is actually counting rather than (or perhaps in addition to) resolving sounds into pure harmonics.

For our own *MathTrax* application, we found that using a dynamic mixture of sinusoids, saw tooth and triangular waveforms provided excellent pitch cues over a wide range of frequencies. The experimental waveform applet provides the capability to experiment with such schemes by allowing you to enter coefficients of each basis function, and these coefficients can include dependence on pitch or frequency. We find that using so-called sigmoid functions 28 to fade various waveforms in or out is an effective means to provide discontinuities at low frequencies and smoother, more aesthetically appealing curves at higher frequencies. Sigmoidal functions are described generically by the equation

$$a(p) = \frac{1}{1 + exp(c_1(p - p_0))} \tag{28}$$

where a is the amplitude coefficient, p is the pitch, c_1 determines the steepness and direction of the fade-in/out, and p_0 is the center of the transition region in the pitch interval. Note that we can use equivalent expressions in f such as

$$\begin{aligned} \alpha(f) &= \tfrac{f^\delta}{\gamma + f^\delta} \text{ for a fade-in or} \\ \alpha(f) &= \tfrac{1}{1+\gamma f^\delta} \text{ for a fade-out} \end{aligned} \tag{29}$$

where γ and δ are positive. The example on the web page uses the alternate formulation shown in 29.

4. REFERENCES

[1] Brown,L. Brewster, S. Ramloll, R. Burton, M. and Riedel, B. Design Guidelines for Audio Presentation of Graphs and Tables. In Proceedings of the 2003 International Conference on Auditory Display, Boston, MA, USA, July 6-9, 2003. Downloaded 6/10-2005

[2] Meijer, P. Hearing a Printed Graph. Downloaded from http://www.seeingwithsound.com/prgraph.htm 6-10-2005.

[3] Dexter, D. Hodgson, T. Shelton, R. Smith, S. and Truong, D. NASA MathTrax Homepage Downloaded from http://prime.jsc.nasa.gov/ 6-10-2005

[4] Fletcher, H., and Munson, W. A. Loudness, its Definition, Measurement, and Calculation. Journal of the Acoustical Society of America, 5, (2), 82-108, 1933

[5] Shelton, R. Waveform Experimental Applet Available at http://prime.jsc.nasa.gov/Applets/waveform.html

[6] Deutsch, D. Processing of pitch combinations. In D. Deutsch (Ed.) The psychology of music, 2nd Edition, Academic Press, 1999, pp. 349-412.