# Auto-Tuning Mother Nature: Waves in Music and Water

Hunter C. Brown
University of Delaware
Physical Ocean Science and Engineering Program
Newark, Delaware 19716
email:hcbrown@udel.edu

***Abstract*—In the late 1990s, the application of a well known mathematical transformation revolutionized the music industry. That process, known as Auto-Tune, relies on the ability of the Fourier Transform to identify dominant frequencies within a time-series record of a signal. Mega-stars like Madonna, Cher, Britney Spears, Tim McGraw, Kanye West, and many more heavily use Auto-Tune not only to ensure perfect pitch, but often simply for the unique sounds the function produces. Perhaps unknown to these singers, wave experts were using exactly the same mathematical functions to identify ocean wave characteristics forty years prior to the first mainstream music hit (Cher's "Believe") to use Auto-Tune. This paper presents an overview of the Fourier Transform and its application in both the music industry and the scientific study of water waves.**

## I. Introduction

In 1998, the musician Cher released a song titled "Believe" which relied heavily on the Auto-Tune technique to produce unique vocal sounds. After the release, musicians around the world were clamoring for the new technology that could correct a singer's voice to perfect pitch or transform a smooth pitch change into a stepped pitch output. The company behind Auto-Tune, Antares Audio Technologies, was the brainchild of former Exxon seismic engineer Harold Hildebrand. He realized that by applying the same principles used in the study of seismic waves in the Earth's crust to acoustic waves, new sound effects could be achieved.

It turns out, the mathematical foundation of the Auto-Tune method has been known to mathematicians for over 200 years. Auto-Tune owes its existence to a lecture from 1807, when Joseph Fourier described a method to approximate any signal through a combination of trigonometric functions [1]. His method, which led directly to what is now widely known as the Fourier transform, rocked the mathematics world and eventually led to wide-ranging application throughout many scientific fields.

In addition to recording artists and pop stars, ocean scientists have been using the Fourier transform to determine wave properties since the 1950s [2]–[4]. After applying the Fourier transform to a wave time-series, non-directional wave information such as the dominant frequency (and period) can be directly identified from the frequency spectrum, and the significant wave height can be recovered from the power spectral density (also computed via the Fourier transform).

This paper presents a brief overview of the Fourier transform and Discrete Fourier transform, followed by real-world applications in popular music and oceanographic wave characterization.

## II. The Fourier Transform

In the mid-1700s, Leonard Euler, a giant in the field of mathematics, succeeded in deriving one of the greatest mathematical statements in history. Namely, through correspondences with Roger Cotes (who worked directly with Isaac Newton), and building upon Johann Bernoulli's work with natural logarithms and integration factors, Euler determined that,

$$e^{-i\omega t} = cos(\omega t) + isin(\omega t) \tag{1}$$

This formula, perhaps suspicious at first glance, has had a profound influence on mathematics. Only through formal proof were many convinced of the veracity of the statement, which seems to deny an intuitive understanding. Fewer than 60 years later, Joseph Fourier used Euler's equation to solve the heat equation in a metal plate [1]. This work directly led to his realization that the amount of energy in a signal at any particular frequency, $f$, could be determined by an integral of the form,

$$\mathcal{F}[x(t)] = X(f) = \int_{-\infty}^{\infty} x(t)e^{-2\pi ift}dt \tag{2}$$

The approximation of the original function $x(t)$ is accomplished through a combination of *sine* and *cosine* functions (evident by the product of $cos(2\pi ft) + sin(2\pi ft)$ simplified as $e^{-2\pi ift}$). Given a signal as a function of time, say $x(t)$, the Fourier transform, denoted $\mathcal{F}$, converts the original signal into a function of frequency, $\mathcal{F}[x(t)] \rightarrow X(f)$. The result is known as the *frequency spectrum* of the original signal $x(t)$. This frequency spectrum describes the amount of energy at each frequency of $x(t)$.

We have so far been discussing the Fourier transform as applied in the continuous domain. While the original theory dealt with a signal that was both continuous and infinite

978-1-4673-0831-1/12/$31.00 ©2012 IEEE

in length, modern usage must contend with signals that are neither continuous nor infinite.

### A. Discrete Functions

Working from a discretely sampled signal, however, adds significant complexity to any attempt at characterizing the original continuous signal. In the discrete domain, the Discrete Fourier Transform (DFT) is written as,

$$X_n = \sum_{k=0}^{N-1} x_k e^{-2\pi ikn/N} \tag{3}$$

where $x_k \equiv x(t_k)$, and $k = 0 \ldots N-1$ is the number of sample points. Note that, in addition to being composed discretely sampled points, the record is of finite duration. The DFT is an approximation of the continuous Fourier transform. This approximation suffers from a well known effect called *spectral leakage*, where, during the transform operation, energy leaks across discrete frequency bands (via aliasing) to create a spreading of energy in the resulting spectrum (Figure 2).

### B. Windowing

One method for reducing spectral leakage is to use window functions (moving weighted functions) in either the time or frequency domains [5], [6]. These windows act like narrowband, low-pass filters and minimize the transition edges of the sampled waveform by tapering, or smoothing, any boundary jump discontinuities. Typically, window functions are of some finite length shorter than the record under scrutiny; they are smooth, continuous, and zero at both ends.

One popular window function is the Hanning Window which appears visually as a zero-ended raised cosine waveform of one-half period of all positive values. The boundary effects (jump discontinuities) of the original discrete signal can be diminished by multiplying the signal by a window function before taking the Fourier transform. An exception to the zero-endpoint rule is the rectangular window function. The rectangular window has a value of 1 within the allowable domain, and is zero everywhere else. Some natural phenomenon that is measured and digitized for a short time is essentially the original waveform multiplied by a rectangular window.

Alternatively, by attempting to estimate an original waveform through the summation of an infinite series of sine and cosines (Fourier Series), the boundary effects are readily seen. This infinite series, an estimate of the actual record, displays energy outside of the original signal domain. Window functions help reduce this spurious energy.

Window functions also reduce another artifact in Fourier Series representations known as the Gibb's Phenomenon. Jump discontinuities are impossible to model with Fourier Series and, although the estimate may be arbitrarily close with ever larger finite sets of sines and cosines, will always exhibit overshoot (also called ringing). Tapering the edges around discontinuities via a window function allow the Fourier Series to converge to the modified signal without the ringing residuals.

## III. Application to Music

One particular type of signal is sound (or acoustic waves). Acoustic waves are pressure fluctuations that can be recorded by microphones which convert the displacement of a membrane, caused by local air pressure changes, to a magnitude represented by an electrical signal, such as current. This process can be visualized as the opposite of what occurs in a loud-speaker, where an electrical signal causes a membrane to vibrate and create sound waves (again in the form of air pressure changes). In the modern sound system, as the microphone membrane vibrates, the generated electrical signal is recorded at discrete time steps by an analog-to-digital converter (ADC) and stored in a digital form representing the amplitude of the vibration at a particular time (known as the displacement time-series).

These discrete time steps are determined by the *sampling frequency*, or the number of measurements taken per second. Experiments have shown that the human ear can hear sound frequencies of approximately 20 to 20,000 hertz [7]. In order to perfectly reproduce any signal in that range, we must sample at some frequency higher than the Nyquist rate, namely a minimum of twice the maximum frequency [8]. If we could be guaranteed that an incoming signal had no energy at frequencies higher than 20,000 Hz, we could safely sample at exactly twice that frequency with no degradation in the reconstructed signal. In the real-world, however, we cannot make that assumption and must sample at a higher frequency to minimize the effects of energy in frequencies above 20kHz.

The sampling rate of the ADC, therefore, has traditionally been fixed at 48kHz since this is the minimum required to match the NTSC (National Television System Committee) video frame rate of 29.97 frames per second. Compact Discs, however, have a sample rate of 44.1kHz because this rate is easily adapted to both common video formats: NTSC in North America and PAL (Phase Alternating Line) outside of North America. Both of these rates are capable of describing any sound signal within the 20-20kHz range with an acceptable degradation (called aliasing) from energy in the frequencies higher than 22kHz.

In order to visualize the use of the Fourier transform, consider the following example. Imagine we strike an A-440Hz (above middle C) tuning fork near a microphone. In a perfect world, the microphone would register a perfect 440Hz time-series sinusoidal oscillation in the membrane and convert the displacement into electrical current. In our example, we measure the electrical current at 1000Hz. This is well above the Nyquist rate of 880Hz required to capture a 440Hz signal and should be high enough to reduce much of the distortion caused by energy aliasing near the upper frequency limit of 1kHz. Figure 1 shows what the electrical measurement looks like over time. In our theoretical experiment, the amplitude of the sinusoid is exactly one, to simplify calculations. In the real world, the amplitude of the signal corresponds to the displacement of the microphone membrane (volume). In effect, for our example, we are envisioning that the tuning fork maintains a constant volume during the sampling period.

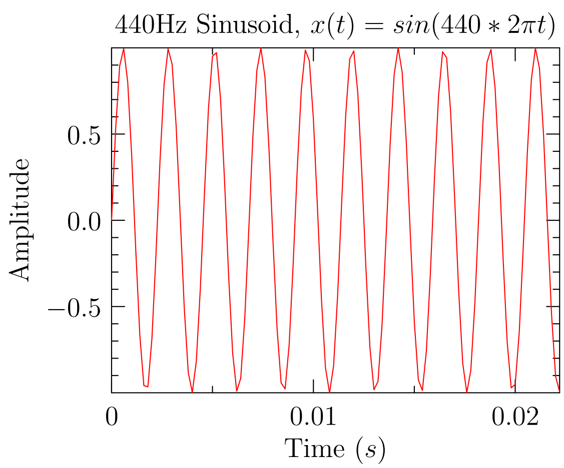


Fig. 1. The sampled 440Hz signal. This is a pure 440Hz sine wave sampled at 1000Hz.

If we process this time series, $x(t) = sin(2\pi f t)$, where frequency $f = 440Hz$, using the Fourier transform, we see the result pictured in Figure 2. The graph shows the relative amount of energy at each frequency in the signal. It is clear that the dominant component of the signal is a 440Hz tone. For a perfectly sampled 440Hz tone we would expect an impulse function at 440Hz. In other words, if we record an A-440 tone and apply the Fourier transform to the digital sample, we will expect to see a large spike (of energy) at the 440Hz frequency. In this case, we see smooth ramps in the graph on either side of 440Hz.

These phantom energy observations arise from spectral leakage introduced during the Fourier transform of the discrete data. Since we are approximating a continuous signal with a discrete signal, the error in the approximation leads to a loss and spreading of energy.

In a sound studio, each instrument and singer is often recorded on an independent audio track that can be manipulated later without influencing other sounds. In this way, it is possible, for example, to completely remove a singer or the drums from a song without influencing any other voices. The post-processing manipulation typically includes removing noise, extraneous sounds, timing mistakes, and sometimes Auto-Tuning a voice. Using the Fourier transform,

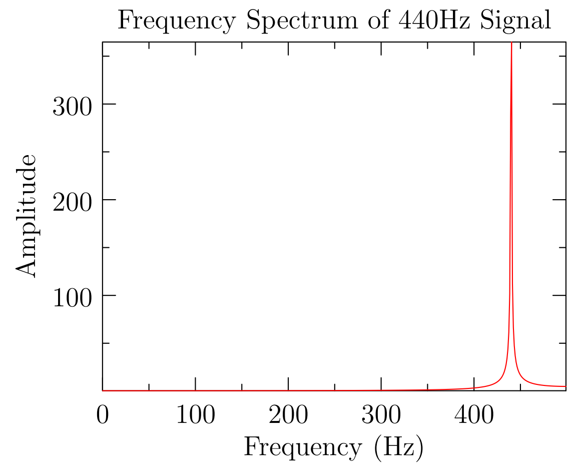


Fig. 2. The peak at 440Hz is quite evident, but note the spreading of energy to nearby frequencies. All discretely sampled signals will experience this effect, known as *spectral leakage*; however, there are techniques available to help reduce the spreading.

we learned that we can identify sound frequencies in an audio sample recorded from a microphone. This is the heart of Auto-Tune.

As a singer vocalizes a particular note, the pitch may vary slightly up and down from the desired sound as the vocal chords attempt to maintain a steady note. With the Fourier transform, it is possible to identify these divergences and subsequently shift the pitch back to the desired tone. Following our 440Hz example, imagine our tuning fork has a dent causing the pitch to ring at 430Hz. After recording the sound of the fork, the Fourier Transform of the recording would show an energy peak at 430Hz. We could then post-process the recording by shifting the frequency up 10Hz to get back to the desired 440Hz note. Now, when we play our modified recording, we would hear a 440Hz tone.

In Figure 3, we show an Auto-Tune scheme to correct pitch to the nearest 100Hz. Therefore, an input pitch of 120Hz would be corrected down to a pitch of 100Hz. Similarly, an input tone of 160Hz would be corrected up to a pitch of 200Hz. This results in a step function with perfect jumps in between notes, contrasting with the smooth ramp generated by a singer smoothly transitioning from a low note up to a high note.

The sound software used in the music industry also includes a speed parameter which dictates how quickly the original pitch is corrected to the desired note. An instant speed setting indicates that the original tone should be corrected instantaneously to the target pitch. In Figure 3, the speed is set to instantaneous (jumps in the blue lines). Otherwise, the output frequency will more slowly approach the intended tone, closely resembling how the human voice attains a certain pitch. This would be evident by ramp functions at the jump discontinuities between notes (Figure 4 for example).

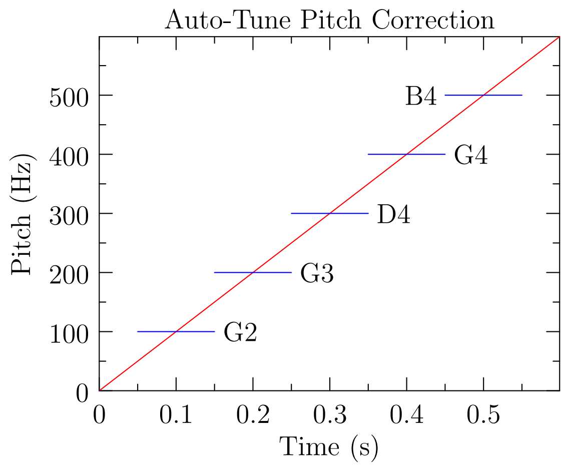


Fig. 3. This figure shows the relation between input pitch (red) and corrected pitch (blue) for a 5-step Auto-Tune. In this case, we corrected the pitch to the nearest 100Hz frequency. The approximate note name and octave (piano) are designated next to the corrected pitch.

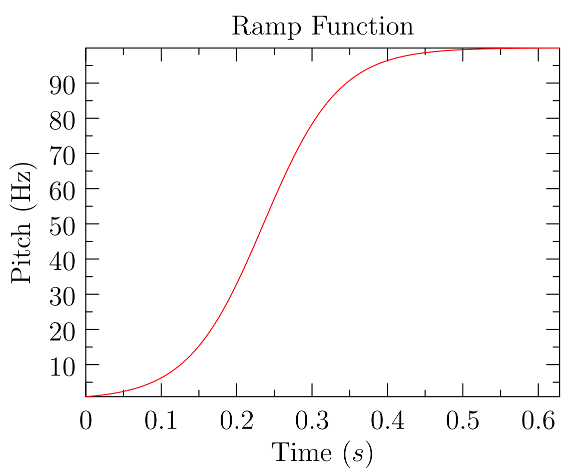


Fig. 4. This figure shows an example ramp function which smoothly transitions from one pitch to another 100Hz higher. This sigmoid shape is defined by the equation $P(t) = a/(1 + e^{-[t*b-(c*\pi)]})$. Where $a$ indicates how far the pitch should change (in Hz), $b$ is a spreading factor that expands or contracts the sigmoid shape over time, and $c$ indicates the horizontal offset. In this case $a = 100$, $b = 20$, and $c = 1.5$

## IV. Application to Ocean Waves

When the sea-surface displacement at a point in the ocean is measured over time and viewed as a time-series, the result is very similar in appearance to the time-series of sound waves. Both sound and water waves may be approximated through combinations of trigonometric functions (often *sines* and *cosines*) and therefore can take advantage of the Fourier transform. Instead of microphones measuring air pressure fluctuations, monitoring buoys are used to measure the ocean surface. These buoys have highly sensitive sensors that measure the buoy's inertial motions. From these measurements, a time-series of the sea-surface may be recovered [2], [4], [9].

Imagine a buoy that is constrained to move only in the z-axis with the waves (i.e. it never pitches or rolls). The buoy has an on-board accelerometer, which measures the vertical axis accelerations (including gravity) experienced by the buoy. It is possible, with calculus, to recover an estimate of the sea-surface displacement from these measurements. Introductory calculus explains that velocity is the derivative of displacement with respect to time. Similarly, acceleration is the time derivative of velocity.

$$D(x) = x = asin(\omega t) \tag{4}$$

$$V(x) = \dot{x} = \frac{dx}{dt} = a\omega cos(\omega t) \tag{5}$$

$$A(x) = \ddot{x} = \frac{d^2x}{dt^2} = \frac{dV(x)}{dt} = -a\omega^2 sin(\omega t) \tag{6}$$

Now going in the other direction, we may recover velocity from the integration of acceleration and displacement from the integration of velocity (all with respect to time).

$$V(x) = \dot{x} = \int A(x)dt \tag{7}$$

$$D(x) = x = \int V(x)dt \tag{8}$$

If we convert the signal into the frequency domain before performing any derivatives (by use of the Fourier transform), we are able to take advantage of the unique properties of the Fourier transform. Notice the coefficients before the sine term in Equation 6. If we let $a = 1$, then, in the frequency domain, we must only divide by -$\omega^2$ to convert between acceleration and displacement. In the real world, negative amplitudes and frequencies have no meaning. Some representations of signals, such as phasors, can lead to apparent negative frequencies, but the negative sign is simply an artifact of the representation necessary to maintain mathematical consistency.

The process of multiple integrations seems relatively easy at first glance. Unfortunately, as $\omega$ gets very small (close to zero), the result of division by $\omega^2$ goes to infinity. In real terms, we get apparent infinite energy (displacement) at frequencies near zero. Obviously this makes no sense physically. The classical method of dealing with this is simply to ignore low frequencies! We now know the lowest frequency limit for real significant wave energy is approximately 0.035 Hz [10], [11]. Tucker explains that occasionally 0.04 Hz is used as a lower limit, but severe storms on the open ocean produce waves that are below this frequency (perhaps one to two storms per year in the North Atlantic [12]).

### A. Dominant Period

After recovering the time-series sea-surface displacement, we can apply the Fourier transform exactly like we did in our tuning fork example. This frequency spectrum, however, will be much different from the frequency spectrum generated in the previous example. The dominant period can be identified by noting the frequency with the most energy in the frequency spectrum of $x(t)$. In this case, the dominant frequency is close to 0.125Hz. This corresponds to a dominant period of around 8 seconds per wave. There also is a lot of energy in the 0.1Hz region, which corresponds to waves with 10 second periods.

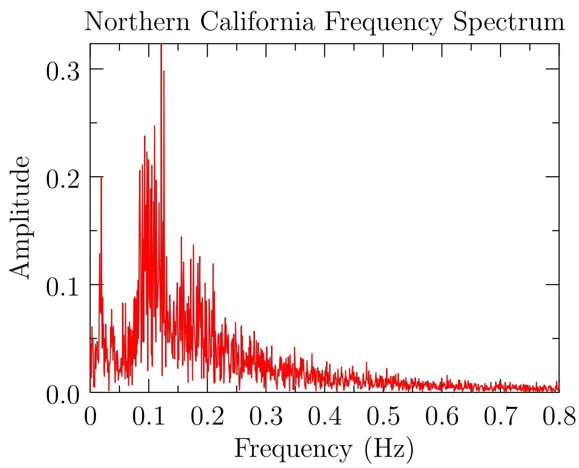


Fig. 5. Frequency spectrum from waves measured at 50Hz off the northern California shore. The significant wave heights, here, were in excess of 3.7m (12ft).

### B. Significant Wave Height

In 1947, Sverdrup and Munk created a new statistical parameter by which to measure waves [13]. The new parameter, called the significant wave height, was defined as the average of the highest one-third of the waves in a given sea, $H_s$ or $\overline{H_{1/3}}$. According to their work, this was the height that trained wave observers were most likely to record when taking visual wave observations. Statistical derivations show the following relationship between the spectral density function and significant wave height [14],

$$H_s \approx 4.01\sqrt{m_0} \tag{9}$$

where $m_0$ is the area under the spectral density function (more properly called the one-sided auto-spectral density, or power spectral density, since t starts at zero), $S(f) = \lim_{t\to\infty} \frac{1}{t} E[|X(f)|^2]$. Where the $E$ signifies the expected value [15] and operates on the magnitude of the frequency spectrum squared.

## V. Conclusion

This paper presented a brief overview of the Fourier transform, and its application to music and the study of ocean waves. It showed that the core functionality behind both the Auto-Tune method and the recovery of ocean wave statistics is essentially the same; both rely heavily on Fourier Series and the Fourier transform. Often, as scientists, mathematicians, and engineers, we associate a technique with a particular domain and rarely consider significance outside its natural habitat. As Hildebrand would certainly agree, cross-discipline collaborations can often lead to scientific and commercial revolutions.

## Acknowledgment

The author would like to express his appreciation to Fabrice Veron for providing the ocean-wave time-series data from Northern California.

## References

[1] J. Fourier, "Memoire sur la propagation de la chaleur dans les corps solides," Nouveau Bulletin des Sciences par la Societe philomathique de Paris, vol. 1, no. 6, pp. 112–116, March 1808, Bernard, Paris.
[2] M. J. Tucker, "A shipborne wave recorder," Transactions of the Institute of Naval Architects, vol. 98, pp. 236–250, 1956.
[3] ——, "The accuracy of wave measurements made with vertical accelerometers," Deep Sea Research, vol. 5, pp. 185–192, 1959.
[4] M. Longuet-Higgins, D. Cartwright, and N. Smith, "Observations of the directional spectrum of sea waves using the motions of a floating buoy," in Ocean Wave Spectra. Prentice-Hall, New Jersey, 1963, pp. 111–132.
[5] F. J. Harris, "On the use of windows for harmonic analysis with the discrete fourier transform," in Proceedings of the IEEE, vol. 66, no. 1, 1978.
[6] S. S. Abeysekera and K. P. Padhi, "An investigation of window effects on the frequency estimation using the phase vocoder," IEEE Transactions on Audio, Speech, and Language Processing, vol. 14, no. 4, pp. 1432–1439, 2006.
[7] J. D. Cutnell and K. W. Johnson, Physics, 4th ed. Wiley, 1997.
[8] C. E. Shannon, "Communication in the presence of noise," in Proceedings of the Institute of Radio Engineers, vol. 37, no. 1, January 1949, pp. 10–21.
[9] H. C. Brown and G. A. Meadows, "Wave Sensing in the Upper-Great Lakes Observing System," in UNO Ocean Waves Workshop, New Orleans, LA, 2011, pp. 74–78.
[10] M. Tucker, Waves in Ocean Engineering: Measurement, Analysis, Interpretation. Ellis Horwood, 1991.
[11] M. Earle and K. Bush, "Strapped-down accelerometer effects on ndbo wave measurements," in OCEANS 1984, vol. 14, Sep 1982, pp. 838–848.
[12] M. J. Tucker, "Interpreting directional data from large pitch-roll-heave buoys," Ocean Engineering, vol. 16, no. 2, pp. 173–192, 1989, cited By (since 1996): 15. [Online]. Available: www.scopus.com
[13] H. U. Sverdrup and W. H. Munk, Wind, Sea and Swell: Theory of Relations for Forecasting. U.S. Hydrographic Office, 1947.
[14] M. K. Ochi, Ocean Waves: The Stochastic Approach. Cambridge University Press, 1998.
[15] J. S. Bendat and A. G. Piersol, Random Data: Analysis and Measurement Procedures, 4th ed. Wiley, 2010, no. 978-0-470-24877-5.