Pitch Detection Algorithms*

Gareth Middleton

Abstract

Two algorithms to detect the fundamental frequency of a signal: one in the time domain (Autocorrelation) and one in the frequency domain (Harmonic Product Spectrum / HPS)

1 Autocorrelation Algorithm

Theory

Fundamentally, this algorithm exploits the fact that a periodic signal, even if it is not a pure sine wave, will be similar from one period to the next. This is true even if the amplitude of the signal is changing in time, provided those changes do not occur too quickly.

To detect the pitch, we take a window of the signal, with a length at least twice as long as the longest period that we might detect. In our case, this corresponded to a length of 1200 samples, given a sampling rate of 44,100 KHz.

Using this section of signal, we generate the autocorrelation function r(s) defined as the sum of the pointwise absolute difference between the two signals over some interval, perhaps 600 points.

Graphically, this corresponds to the following:

*Version 1.2: Dec 17, 2003 3:25 pm -0600

†http://creativecommons.org/licenses/by/1.0

Shifting the signal

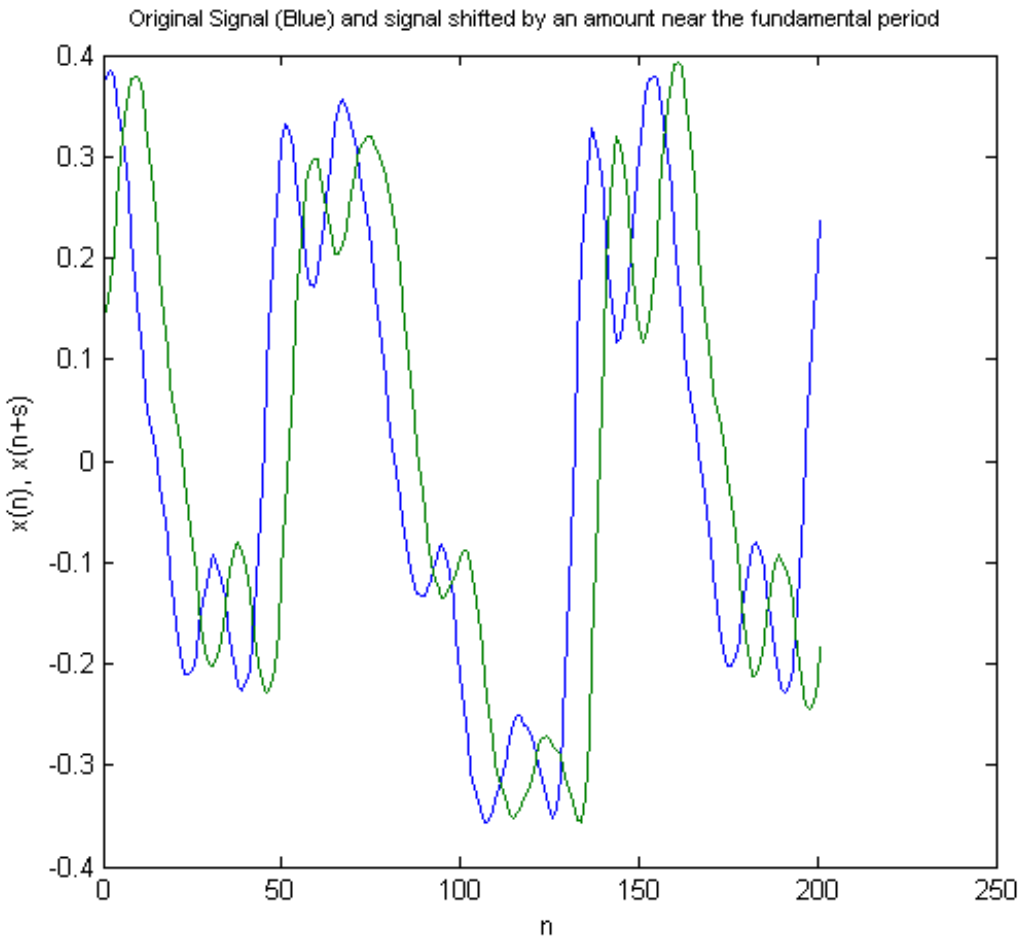

Figure 1: Here, the blue signal is the original and the green signal is a copy of the original, shifted left by an amount nearing the fundamental period. Notice how the signals begin to align with each other as the shift amount nears the fundamental period.

Intuitively, it should make sense that as the shift value s begins to reach the fundamental period of the signal T, the difference between the shifted signal and the original signal will begin to decrease. Indeed, we can see this in the plot below, in which the autocorrelation function rapidly approaches zero at the fundamental period.

Autocorrelation Function

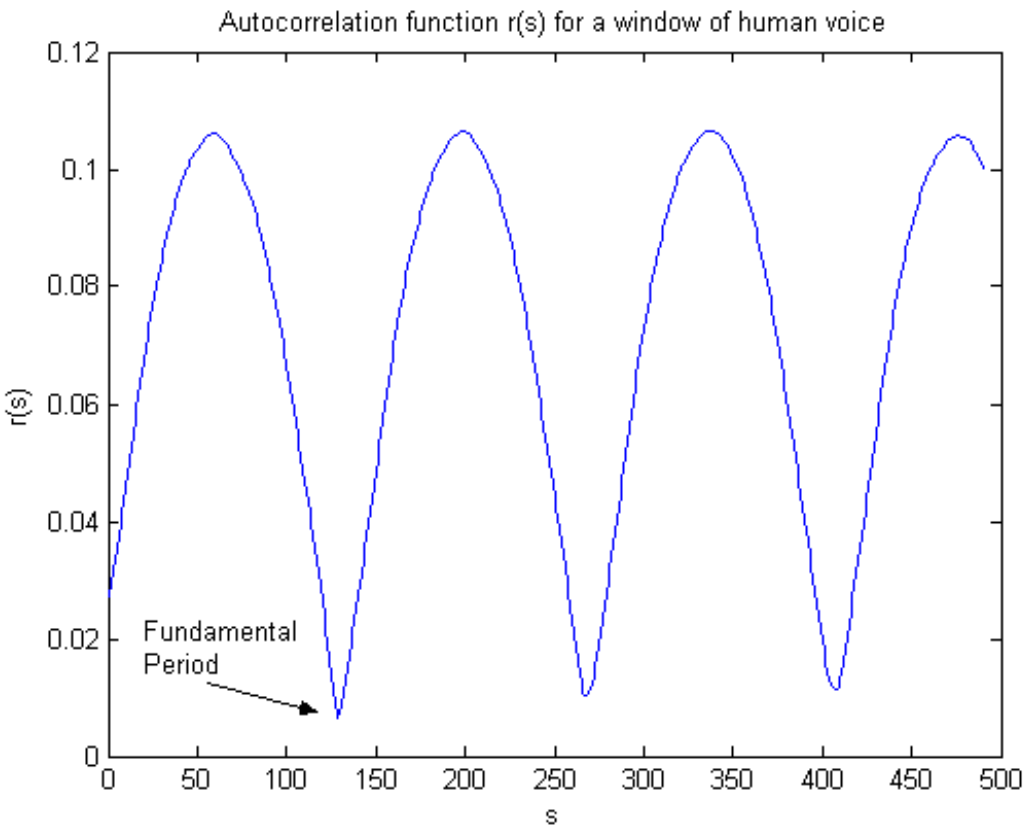

Figure 2: The fundamental period is indentified as the first minimum of the autocorrelation function. Notice that the function is periodic, as we expect. r(s) measured the total difference between the signal and its shifted copy, so the shifs approach k*T, the signals again align and the difference approaches zero.

We can detect this value by differentiating the autocorrelation function and then looking for a change of sign, which yields critical points. We then look at the direction of the sign change across points (positive difference to negative), to take only the minima. We then search for the first minimum below some threshold, i.e. the minimum corresponding to the smallest s. The location of this minimum gives us the fundamental period of the windowed portion of signal, from which we can easily determine the frequency using

FAST-Autocorrelation

Clearly, this algorithm requires a great deal of computation. First, we generate the autocorrelation function r(s) for some positive range of s. For each value of s, we need to compute the total difference between the shifted signals. Next, we need to differentiate this signal and search for the minimum, finally determining the correct minimum. We must do this for each window.

In generating the r(s) function, we define a domain for s of 0 to 599. This allows for fundamental frequencies between about 50 and 22000 Hz, which works nicely for human voice. However, this does require calculating r(s) 600 times for each window.

In effort to improve the efficiency of this algorithm, we created an alternative called FAST Autocorrelation, which has yielded speed improvements in excess of 70%.

We exploit the nature of the signal, specifically the fact that if the signal was generated using a high sampling rate and if the windows are narrow enough, we can assume that the pitch will not vary drastically from window to window. Thus, we can begin calculating the r(s) function using values of s that correspond to areas near the previous minimum. This means that, if the previous window had a fundamental period of 156 samples, we begin calculating r(s) for s = 136. If we fail to find the minimal s in this area, we calculate further and further from the previous s until we find a minimum.

Also, we note that the first minimum (valued below the threshold) is always going to correspond to the fundamental frequency. Thus, we can calculate the difference equation dr(s)/ds as we generate r(s). Then, when we find the first minimum below threshold, we can stop calculating altogether and move on to the next window.

If we use only the second improvement, we usually cut down the range of s from 600 points to around 200. If we then couple in the first improvement, we wind up calculating r(s) for only about 20 values of s, which is a savings of (580) * (1200) = 700000 calculations per window. When the signal may consist of hundreds of windows, this improvement is substantial indeed.

Limitations of Autocorrelation

The autocorrelation algorithm is relatively impervious to noise, but is sensitive to sampling rate. Because it calculates fundamental frequency directly from a shift in samples, it follows that if we have a lower sampling rate, we have lower resolution in pitch.

As stated earlier, autocorrelation is also extremely expensive computationally. However, using the adaptive techniques described above, computation can be expedited and run in near-real time.

2 Harmonic Product Spectrum

Theory

If the input signal is a musical note, then its spectrum should consist of a series of peaks, corresponding to fundamental frequency with harmonic components at integer multiples of the fundamental frequency. Hence when we compress the spectrum a number of times (downsampling), and compare it with the original spectrum, we can see that the strongest harmonic peaks line up. The first peak in the original spectrum coincides with the second peak in the spectrum compressed by a factor of two, which coincides with the third peak in the spectrum compressed by a factor of three. Hence, when the various spectrums are multiplied together, the result will form clear peak at the fundamental frequency.

HPS Overview

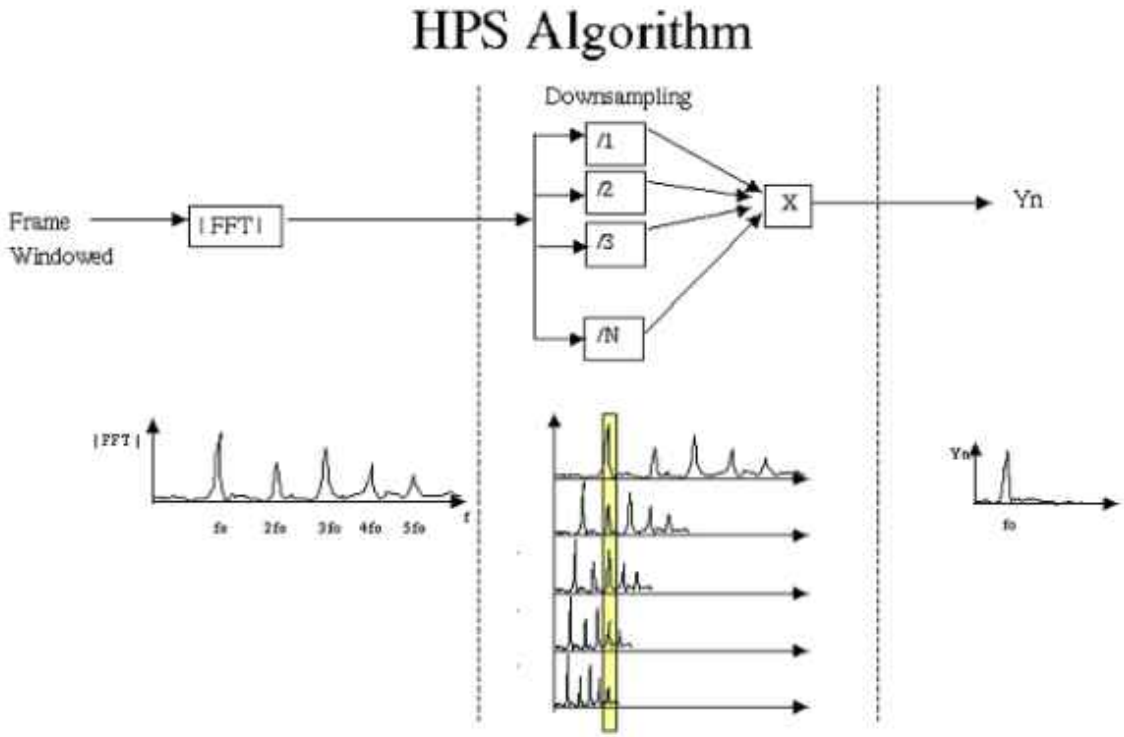

Figure 3: First, the windowed frame is taken into the frequency domain and the magnitude of the spectrum is calculated (left). Next, the spectrum is downsampled to create more compressed versions of itself (center). Notice how the higher harmonics of the fundamental frequency align with each other in the downsampled spectra. Last, a multiplication of these spectra is performed and the maximum is found (right). This corresponds to the fundamental frequency.

Method

First, we divide the input signal into segments by applying a Hanning window, where the window size and hop size are given as an input. For each window, we utilize the Short-Time Fourier Transform to convert the input signal from the time domain to the frequency domain. Once the input is in the frequency domain, we apply the Harmonic Product Spectrum technique to each window.

The HPS involves two steps: downsampling and multiplication. To downsample, we compressed the spectrum twice in each window by resampling: the first time, we compress the original spectrum by two and the second time, by three. Once this is completed, we multiply the three spectra together and find the frequency that corresponds to the peak (maximum value). This particular frequency represents the fundamental frequency of that particular window.

Limitations of the HPS method

Some nice features of this method include: it is computationally inexpensive, reasonably resistant to additive and multiplicative noise, and adjustable to different kind of inputs. For instance, we could change the number of compressed spectra to use, and we could replace the spectral multiplication with a spectral addition. However, since human pitch perception is basically logarithmic, this means that low pitches may be tracked less accurately than high pitches.

Another severe shortfall of the HPS method is that it its resolution is only as good as the length of the

FFT used to calculate the spectrum. If we perform a short and fast FFT, we are limited in the number of discrete frequencies we can consider. In order to gain a higher resolution in our output (and therefore see less graininess in our pitch output), we need to take a longer FFT which requires more time.

References

[1] Paul Boersma. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *Institute of Phonetic Sciences, University of Amsterdam, Proceedings 17*, 1993.

[2] Craig Sapp Patricio de la Cuadra, Aaron Master. Efficient pitch detection techniques for interactive music. Center for Computer Research in Music and Acoustics, Stanford University.

[3] Curtis Roads. *Computer Music Tutorial*, chapter 10. MIT Press, 1996.