A Computationally Efficient Multipitch Analysis Model

Gokulraj R (2020102042)   Sushil Kumar Yalla (2020102071)

1 Abstract

We aim to design a computationally efficient multipitch analysis model that is as close as possible to the model presented in the paper [3] assigned to us. This model essentially splits the signal ino two channels, computes the generalized autocorrelation function for both the channels, and sum them to give the summary autocorrelation function (SACF), which is then processed to obtain enhanced SACF (ESACF). We also demonstrate how this model is computationally more efficient than the unitary pitch analysis model [2] by Meddis and O'Mard.

2 Meddis and O'Mard Model

The unitary pitch analysis model by Meddis and O'Mard [2] and its predecessors by Meddis and Hewitt [1] are some of the recent models of time-domain pitch analysis. This unitary model shows good correspondence to human pitch perception but it has a problem that its algorithm is computationally expensive because the analysis is carried out using a multichannel filterbank which splits the signal into 32-120 channels depending upon the number of filters used. Finding the autocorrelation function (ACF) for the signals in each channel makes this model computationally inefficient.

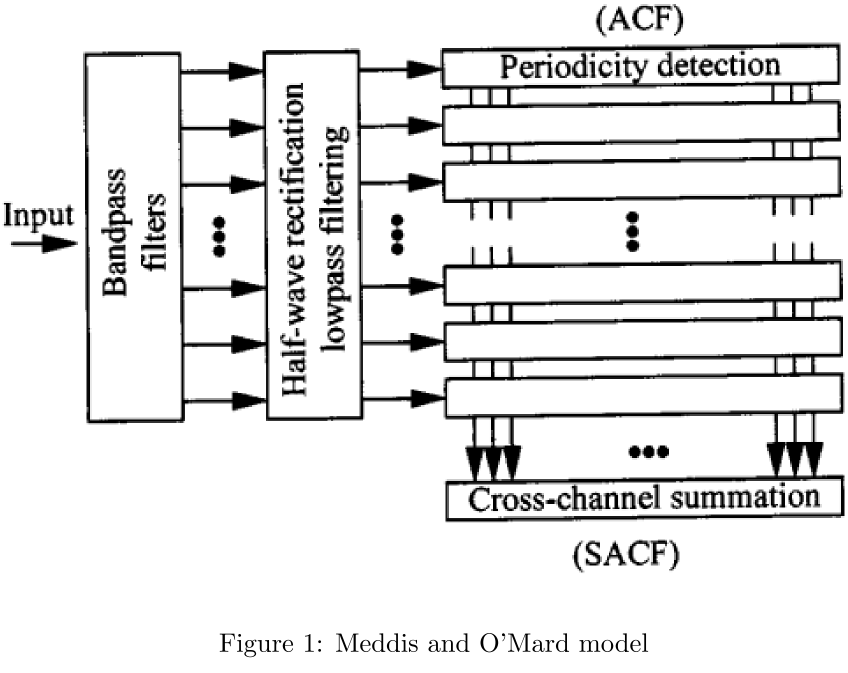


Figure 1: Meddis and O'Mard model

Fig. 1 shows the block diagram of this model. The ACF of all the channels are summed to get the Summary Auto Correlation Function (SACF) which is used in pitch analysis. There are many ways to compute the autocorrelation or a similar periodic measure. The time-domain approach is a common choice and this is used in [2], [1]. The computation of the autocorrelation in time-domain for each channel is not computationally efficient. This motivates the develpement of a simplified pitch perception model that is computationaly more efficient and still qualitatively retains the accuracy of multichannel systems.

3 Proposed Model

The model proposed in [3] splits the signal into only two channels rather than multiple channels. Also, in this model, Discrete Fourier Transform (DFT) based computation of autocorrelation is used for computational efficiency. This approach allows for processing the signal in frequency-domain. Nonlinear compression of DFT magnitude is used to enhance the performance of the analysis. Such a compression is not readily implementable in a time-domain system.

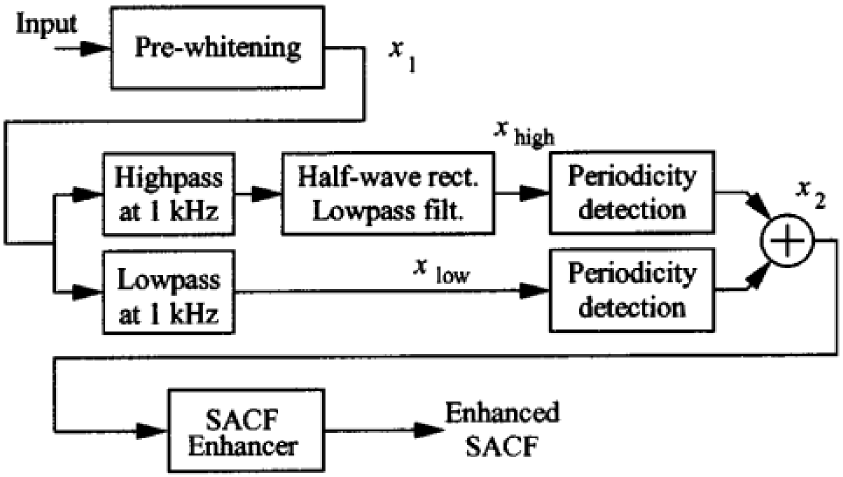


Figure 2: Proposed Model

Fig. 2 shows the block diagram of the proposed model. The signal is first pre-whitened to remove short-time correlations. Then a Hamming window of 1024 samples (window size = 46.4 ms and sampling frequency = 22050 Hz) is applied. The middle part of this model resembles the Meddis and O'Mard model, except that here the signal is split into only two channels, one with frequencies above 1kHz and the other with frequencies below 1kHz.

Both the highpass and lowpass filters used here are Butterworth type and has filter order 2. The lowpass filter also shows highpass characteristics at 70 Hz (i.e., its passband is 70 Hz to 1000 Hz). The high channel signal is half-wave rectified and lowpass filtered (using a filter similar to that used to separate the channels). The autocorrelation function (ACF) for each channel is now calculated using DFT and IDFT. The sum of ACF of the high channel and the low channel gives the Summary Auto Correlation Function (SACF), which is denoted as x_2 in Fig. 2

$$x_2 = \mathrm{IDFT}(|\mathrm{DFT}(x_{low})|^k) + \mathrm{IDFT}(|\mathrm{DFT}(x_{high})|^k) = \mathrm{IDFT}(|\mathrm{DFT}(x_{low})|^k + |\mathrm{DFT}(x_{high})|^k)$$

The value of k determines the frequency domain compression. The ACF computation using DFT allows the control of the parameter k. For normal autocorrelation, $k = 2$ but the performance of the model for different values of k has been analyzed and it is found that the model gives the best results for $k = 0.67$. FFT and IFFT algorithms are used to compute DFT and IDFT efficiently.

In the final block of Fig. 2, the SACF is processed to obtain Enhanced Summary Auto Correlation Function (ESACF). SACF contains much redundant and spurious information that makes it difficult to estimate which peaks are true peaks. A peak pruning technique is used to remove the false peaks in the SACF curve.

4 Simulations and Results

The test input signal we used is a sum of two sinusoids with fundamental frequencies 140.0 Hz and 148.3 Hz and white Gaussian noise with signal-to-noise ratio (SNR) of 2.2 dB. Fig. 3 shows the plot of the input signal. We skipped the pre-whitening part as we were unsure about its implementation.

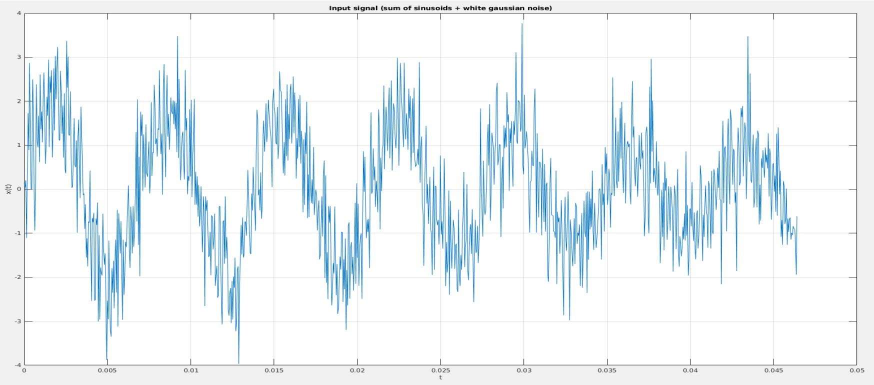


Figure 3: Input Test Signal (Sum of sinusoids with added white Gaussian noise)

4.1 Windowing

A Hamming window of 46.4 ms with a sampling frequency of 22050 Hz (i.e., $0.0464 \times 22050 = 1024$ samples) is used. Fig. 4 shows the plot of the signal after windowing.

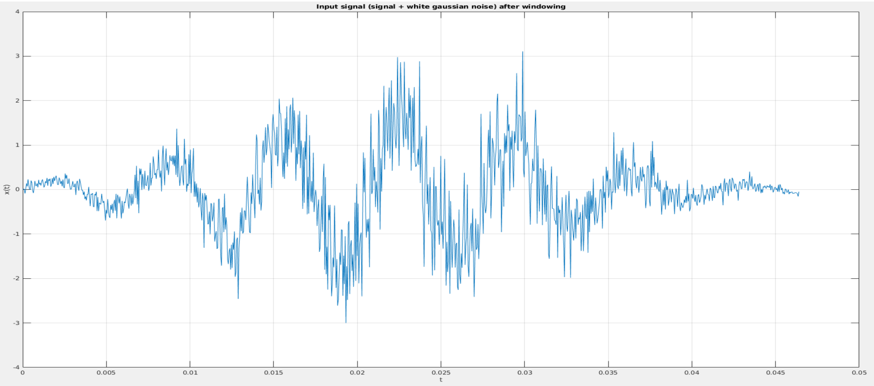
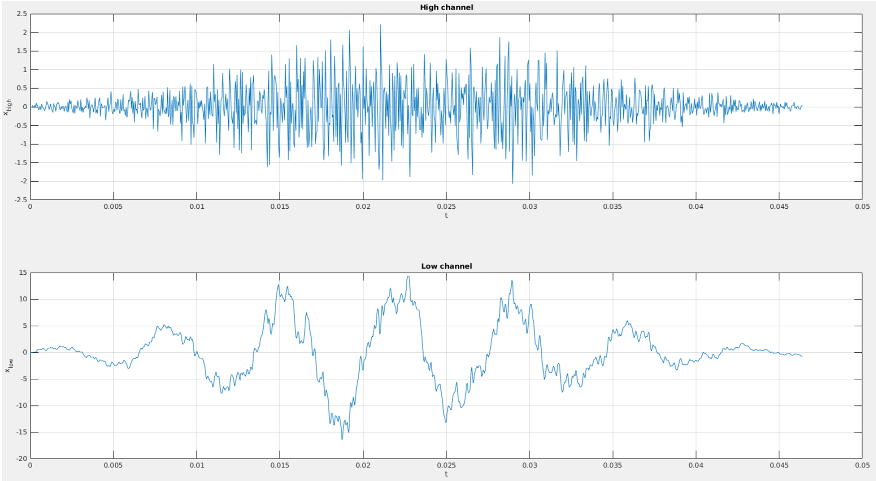


Figure 4: Signal after applying Hamming window

4.2 Highpass and Lowpass filters

The signal is then split into two channels using a highpass and a lowpass filter. The highpass filter has its passband from 1000 Hz to 10000 Hz. The lowpass filter has its passband from 70 Hz to 1000 Hz. Both these filters are of the Butterworth type. The resolution of the ESACF for different values of filter order has been analyzed in [3]. The filter order 2 for each transition band is the best compromise between the resolution of the ESACF and the number of spurious peaks in the ESACF. Fig. 5 shows the plot of the high channel and the low channel signal.

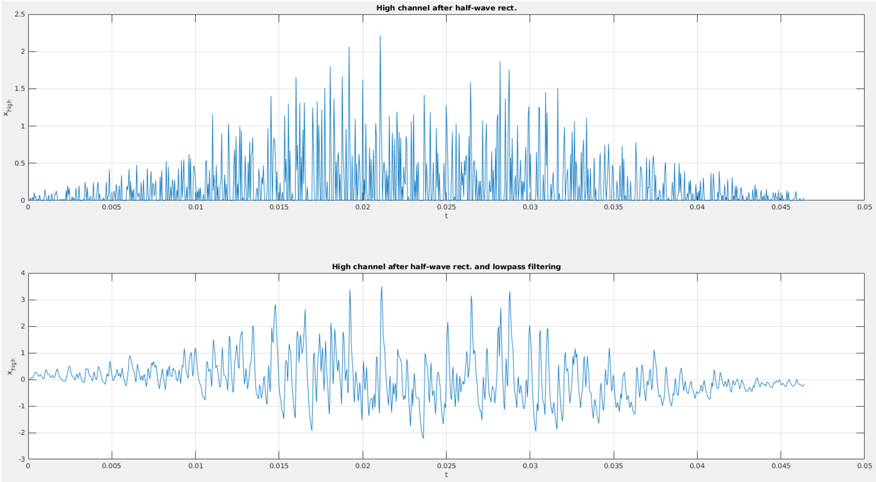

High channel

Low channel

Figure 5: The top and the bottom plots show the high channel and the low channel signal respectively

4.3 Half-wave rectification and lowpass filtering of the high channel signal

The high channel signal is now half-wave rectified and low pass filtered. The half-wave rectifier allows only the positive values of the signal and it removes the negative values. The lowpass filter used here is the same lowpass filtered that we used in the previous step. Fig 6 shows the plots of the high channel signal after half-wave rectification and after low pass filtering.


High channel after half-wave rect.

High channel after half-wave rect. and lowpass filtering

Figure 6: The top and the bottom plots show the high channel signal after half-wave rectification and lowpass filtering respectively

4.4 Periodicity Detection

The autocorrelation function for each channel is calculated in the Periodicity detection block. As mentioned earlier, we use DFT and IDFT to compute autocorrelation. Using DFT and IDFT, the autocorrelation of a signal x is defined as $\mathrm{IDFT}(|\mathrm{DFT}(x)|^k)$. Here, k represents the frequency domain compression. The summary autocorrelation function (SACF) is the sum of autocorrelation functions of the high channel and the low channel signals. So,

$$\mathrm{SACF} = \mathrm{IDFT}(|\mathrm{DFT}(x_{low})|^k + |\mathrm{DFT}(x_{high})|^k)$$

The SACF peaks get broader as k increases. However, the performance with low values of k is compromised by sensitivity to noise. $k \approx 0.67$ is a good compromise between lag-domain resolution and sensitivity to noise. Fig. 7 shows the plot of SACF for $k = 0.67$

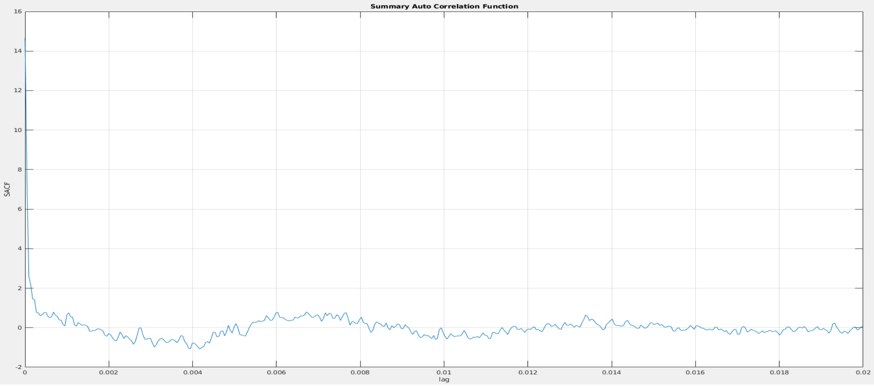


Figure 7: Summary Autocorrelation Function for $k = 0.67$

4.5 Enhancing SACF

The SACF most often contains repetitive peaks of pitches, making it hard to identify the true peaks. So, we process the SACF to remove repetitive peaks. We first clip the SACF curve to positive values and then we time-scale it by a factor of 2. Now, we subtract the time-scaled curve from the original clipped SACF curve, and again the result is clipped to have positive values only. The final result of prcessing the SACF is the Enhanced Summary Auto Correlation Function (ESACF). Fig. 8 shows the plot of ESACF.

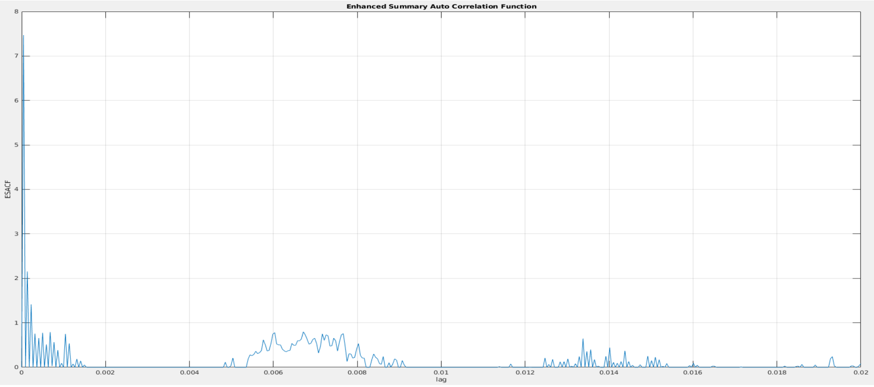


Figure 8: Enhanced Summary Autocorrelation Function

In the ESACF curve, we can clearly identify that there are two peaks (more like disturbances) corresponding to the two fundamental frequencies present in the original signal.

5 Conclusion

Although the auditory analogy of the model (Fig. 2) is not very strong, it shows some features like pre-whitening and channel separation that make it easier to interpret analysis from the point of view of human pitch perception. The main motive of this model is to improve the computational efficiency. This model proves to be better in efficiency than the Meddis and O'Mard model, while sacrificing a little bit of accuracy and auditory relevance.

This model may be used in complex audio signal processing algorithms, such as sound source separation, computational auditory scene analysis, structural representation of audio signals, etc.

References

[1] Ray Meddis and Michael J Hewitt. “Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: Pitch identification”. In: *The Journal of the Acoustical Society of America* 89.6 (1991), pp. 2866–2882.

[2] Ray Meddis and Lowel O’Mard. “A unitary model of pitch perception”. In: *The Journal of the Acoustical Society of America* 102.3 (1997), pp. 1811–1820.

[3] Tero Tolonen and Matti Karjalainen. “A computationally efficient multipitch analysis model”. In: *IEEE transactions on speech and audio processing* 8.6 (2000), pp. 708–716.